Structure and Mechanism of Transcriptional Regulation by the I-A 2'dG-sensing Riboswitch

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von Christina Helmling

Frankfurt am Main

2017

(D30)

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

Dekan:	Prof. Dr. Michael Karas
Erster Gutachter:	Prof. Dr. Harald Schwalbe
Zweiter Gutachter:	Prof. Dr. Alexander Heckel
Datum der Disputation:	

Meiner Familie

This thesis was prepared under the supervision of Prof. Dr. Harald Schwalbe between October 2013 and January 2017 at the Institute for Organic Chemistry and Chemical Biology at the Johann Wolfgang Goethe-University in Frankfurt am Main, Germany.

Table of Contents

List of Abbreviations

AdoCbl	Adenosylcobalamin
2-AP	2-aminopurine
ATP	Adenosine triphosphate
AT	Antiterminator
BEST	Band selective short transient
B.subtilis	*Bacillus subtilis*
2'dAP	2-aminopurin 2’-deoxyriboside
DD	dipole-dipole
2'dG	2’-deoxyguanosine
DNA	deoxyribonucleic acid
dsRNA	double stranded RNA
DTT	Dithiothreitol
CD	Circular dichroism
COSY	Correlation spectroscopy
CSA	Chemical shift anisotropy
CTP	Cytosine triphosphate
DEAE	Diethyl-aminoethyl
DMSO	Dimethyl sulfoxide
EC	Elongation complex
EDTA	Ethylenediaminetetraacetic acid
FMN	Flavin mononucleotide
GARP	Globally optimized alternating phase rectangular pulse
GlmS	Glucosamine-6-phosphate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HMQC	Heteronuclear multiple quantum coherence
HPLC	High pressure liquid chromatography
HPSF	High purity salt free
HSQC	Heteronuclear single quantum coherence
IMP	Inosine monophosphate
INEPT	Insensitive nuclei enhanced by polarization transfer
ITC	Isothermal titration calorimetry

LB	Lysogeny Broth
lncRNA	Long noncoding RNA
mfl/*M.florum*	*Mesoplasma florum*
miRNA	micro RNA
mRNA	messenger RNA
MSM	Markov state model
MWCO	Molecular weight cutoff
ncRNA	noncoding RNA
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
nt	Nucleotide
NTP	Nucleoside triphosphate
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pre-mRNA	Precursor messenger RNA
preQ1	7-Aminomethyl-7-deazaguanine
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNR	Ribonucleotide reductase
RP	Reversed phase
rRNA	Ribosomal RNA
SAM	S-Adenosyl methionine
SD	Shine-Dalgarno
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNAs	Small nuclear RNA
SOFAST	Band-selective optimized flip-angle short-transient
TEMED	Tetramethylethylenediamine
THF	Tetrahydrofurane
TLC	Thin layer chromatography
TPP	Thiamine pyrophosphate
TROSY	Transverse relaxation optimized spectroscopy
tRNA	Transfer RNA

UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultraviolet
V.vulnificus	*Vibrio vulnificus*
XMP	Xanthine monophosphate
ZMP	5‘-aminoimidazole-4-carboxamide ribonucleotide

Summary and Overview

Riboswitches are an important class of regulatory RNA elements that respond to cellular metabolite concentrations to regulate gene expression in a highly selective manner. 2'-deoxyguanosine-sensing (2'dG) riboswitches represent a unique riboswitch subclass only found in the bacterium *Mesoplasma florum* and are closely related to adenine- and guanine-sensing riboswitches.[1] The I-A type 2'dG-sensing riboswitch represses the expression of ribonucleotide reductase genes at high cellular concentrations of 2'dG as a result of premature transcription termination.

Increasing evidence within the last decade suggests that transcriptional regulation by riboswitches is controlled kinetically and emphasizes the importance of co-transcriptional folding.[2–4] Addition of single nucleotides to nascent transcripts causes a continuous shift in structural equilibrium, where refolding rates are competing with the rate of transcription.[5,6] For transcriptional riboswitches, both ligand binding and structural rearrangements within the expression platform are precisely coordinated in time with the rate of transcription. The current thesis investigates the mechanistic details of transcriptional riboswitch regulation using the I-A 2'dG-sensing riboswitch as an example for a riboswitch that acts under kinetic control.

Chapter I describes the relevant background of riboswitches with the focus on transcriptional regulation and kinetic control. Basic principles of riboswitch-mediated gene regulation are outlined in 1.2.1, available evidence for kinetic control by transcriptional riboswitches is summarized in 1.2.2, and a detailed background on guanine-, adenine- and 2'dG-sensing riboswitches can be found in sections 1.2.3-1.2.4.

Chapter II describes methodological approaches applied to investigate riboswitch structure and function and is subdivided into biochemical and biophysical approaches. Biochemical approaches include PCR, *in vitro* transcription and enzymatic splinted ligation for RNA preparation, gel-electrophoresis and HPLC for RNA analysis and purification, and transcription assays for functional studies. Biophysical approaches include UV-spectroscopy for RNA analysis, ITC to determine ligand-binding affinities, stopped-flow spectroscopy to determine ligand-binding kinetics, and NMR-spectroscopy including two

different time-resolved approaches to study structure, dynamics and folding kinetics of RNA.

Chapter III describes the structural characterization of both regulatory ON- and OFF-states of the I-A type 2'dG riboswitch by NMR. NMR characterization of the full-length riboswitch in presence and absence of ligand shows that the riboswitch adopts the regulatory OFF-state at thermodynamic equilibrium in both the ligand-bound and ligand-free form. This result reveals that the riboswitch operates under kinetic control, where the regulatory ON-state represents a transient, metastable state that folds during transcription. Once formed, this ON-state is kinetically stabilized until the point of regulatory decision. The regulatory ON-state was first identified in a 122 nt long transcriptional intermediate. NMR analysis of this transcriptional intermediate revealed two conformations that exchange on the seconds time scale. In addition, a discrepancy in the allosteric modulation between ON- and OFF-states in comparison to guanine and adenine riboswitches was identified. Therefore, allosteric states in the I-A 2'dG riboswitch were further compared to *mfl* I-B 2'dG- and III-B guanine-sensing riboswitches. The ligand-dependent structural rearrangement in all three *mfl* riboswitches was found to be identical. This result suggests that the difference in the ligand-dependent allosteric modulation is not related to 2'dG riboswitches but appears to be a bacterium specific feature. In addition, the preliminary NMR results on the I-B 2'dG-sensing riboswitch suggests that this riboswitch also exhibits structural heterogeneity in the ON-state.

Following structural characterization of the 2'dG-sensing riboswitch, a methodological approach to rapidly synthetize a large number of transcriptional intermediates for NMR analysis was developed. This approach can be applied to monitor how structures evolve during transcription as single nucleotides are added to the mRNA chain. The approach addresses two issues that impede rapid preparation of RNA for NMR-spectroscopic analysis by *in vitro* transcription. First, run-off transcriptions with T7-polymerase typically yield inhomogeneous 3'-ends with non-templated n+1 and n+2 nucleotide additions. Self-cleaving 3'-ribozymes are commonly applied to resolve this issue, which further decelerates the process of RNA preparation. Second, standard purification by HPLC or PAGE is time-consuming and cannot be parallelized. The 3'-end homogeneity for run-off transcription from PCR templates could be significantly improved by using DMSO as a co-solvent in transcription reactions. In combination with 2'-methoxy modifications implemented at the two 5'-nucleotides in DNA templates,[7] random nucleotide addition to

the 3'-end of the RNA by T7 polymerase could be completely inhibited. Transcription products produced under these conditions proved to be exceedingly pure and homogeneous, such that simple buffer exchange of the transcription mixture in centrifugal concentrators yielded NMR spectra comparable to samples purified by PAGE or HPLC. The whole procedure of RNA preparation can be performed within two days and is divided into three steps. First, DNA template preparation is performed simultaneously within two hours for a multitude of transcriptional intermediates by varying reverse primers accordingly. Second, transcription reactions are performed overnight. Third, the transcription mixture is washed with water and exchanged into buffer within 10 hours.[8]

In **Chapter IV,** the mechanistic details of transcriptional riboswitch regulation is investigated by combining biophysical analysis of transcriptional intermediates with light-induced NMR experiments to monitor ON-state folding in real time. A total of 29 transcriptional intermediates were synthesized to investigate structure, ligand binding affinity and ligand binding kinetics in the context of transcription.

NMR structural profiling in the absence of ligand reveals that the metastable ON-state is stable during the transcription of 24 nucleotides between transcript lengths 113 and 137 nt. This result implies that the rate for ON-state folding must occur on the same time scale as required by the polymerase to synthesize these 24 nucleotides, otherwise transcription would always be terminated. The regulatory OFF-state regains thermodynamic stability only five nucleotides prior to the genetic decision. Therefore, the time frame during which the ON-state must be kinetically stabilized is relatively small.

Ligand binding affinity analysis by NMR shows that the ligand-bound state is only populated >90% in the individual aptamer domain over a sequential frame of 10-13 nucleotides from transcript lengths 80 to 93 nt. Elongation of the mRNA chain leads to a continuous loss in the ligand-bound population to 70% during transcription of the subsequent 20 nucleotides to a transcript length of 113 nt. Even in the full-length riboswitch, the ligand-bound state is only populated by 70%. The decrease in ligand-binding affinity during mRNA chain elongation is supported by thermodynamic analysis of transcriptional intermediates using ITC. Ligand binding kinetics derived from stopped-flow fluorescence spectroscopy of transcriptional intermediates show that kinetics of ligand binding do not change as the RNA is elongated by single nucleotides. Consequently, ligand binding kinetics are solely guided by fluctuations in the cellular 2'dG concentration, where a decrease in the ligand-bound population with increasing RNA length directly reflects a

response mechanism to the cellular 2'dG concentration. At high concentrations of 2'dG, ligand association rates are high and the ligand-bound state will be populated to a larger degree. Decreasing 2'dG concentrations decelerate the process of ligand association, directly resulting in a lower population in the ligand-bound state. According to these results, the rate of ligand binding associated with the critical concentration of 2'dG required to trigger riboswitch function resembles the rate of transcription required to synthesize 30 nucleotides. This rate is in the range of 0.3 - 3 s^{-1} according to reported transcription rates for bacterial polymerases.[9–11]

Finally, ON-state folding was monitored by real-time NMR by means of the 122 nt, ON-state stabilized, transcriptional intermediate. This construct was trapped in the aptamer conformation by photolabile protection groups implemented in the expression platform. The construct was prepared by enzymatic ligation of an isotope labeled 85mer, prepared by *in vitro* transcription, to a photocaged 36mer, prepared by solid-phase synthesis. Light-induced cleavage of the photolabile protection groups initiates refolding to the ON-state and can be monitored directly. The kinetic experiments show that the ligand-free aptamer domain folds to the ON-state at a rate of 1.3 s^{-1} and the ligand-bound aptamer domain at a rate of 0.08 s^{-1}. While folding of the ligand-free aptamer domain occurs on the same time scale as the transcription of the determined 24 nucleotide window of opportunity for ON-state folding (0.4 - 3.7 s^{-1}), ligand binding decelerates ON-state folding slightly below these transcription rates. Further, co-transcriptional folding was simulated during the synthesis of these 24 nucleotides both in presence and absence of ligand. The simulations suggest that the I-A 2'dG-sensing riboswitch mediates the regulatory outcome only in a dynamic range of ~40-50% and requires transcription rates of 10-50 nt/s to be regulatory active. In comparison, reported rates of transcription for bacterial polymerases are between 10-90 nt/s.[9–11]

Results presented within this chapter strongly suggest that transcriptional riboswitches do not operate as binary ON/OFF switches. In contrast, population ratios are continuously fine-modulated as the RNA is extended by single nucleotides and directed by small changes in kinetic rates for both ligand-binding and ON-state folding. In this kinetically coordinated system, the ligand-dependent population difference between ON and OFF states at the regulatory decision point does not exceed 50% and strongly depends on bacterium specific transcription rates.

Chapter I

General Introduction

1.1. Role and function of RNA in cellular processes

The classic central dogma of molecular biology describes the flow of genetic information from DNA and its replication, to the transcription of messenger RNA (mRNA) and finally the translation of proteins.[12,13] In this pathway leading to protein synthesis, RNA was silently declared to serve simply as a messenger of genetic information for protein translation. According to this former paradigm, proteins were believed to constitute the cellular machinery on their own by combining catalytic activity with the ability to regulate gene expression through transcription factors.[14] RNA was classified as the primordial molecule of life, both capable of catalytic activity and storage of genetic information, but appeared to have been replaced by the more stable and easy to replicate DNA as genetic carrier and the chemically more versatile proteins.[15] The discovery of a vast amount of non-protein coding RNAs within the last decades have caused a significant paradigm shift regarding the role of RNA in cellular processes.

In 1977, introns were discovered.[16,17] At this point in time, these non-coding RNA segments were simply dismissed as degrading transcriptional noise. However, it became increasingly evident that the large amount of intronic RNA contains many functional RNAs.[18] The first structural RNA elements found to exert enzymatic activity were discovered in 1982 in terms of RNA-mediated self-splicing and were consequently termed ribozymes.[19,20] The identification of tRNAs and rRNAs was followed by small nuclear RNAs in 1986 (snRNAs), more recently termed spliceosomal RNAs.[21] snRNAs are involved in diverse RNA-RNA and RNA-protein complexes and constitute an essential co-factor in RNA splicing.[22] In 1995, the class of small nucleolar RNAs (snoRNAs) was found in the nucleolus. snoRNAs guide proteins to cause methylation (subclass box C/D) and pseudouridylation (subclass box H/ACA) of rRNAs, tRNAs and snRNAs, essential for ribosomal (tRNA) and cellular (mRNA maturation and pre-mRNA splicing) function.[23–25]

RNA was first evidenced to play a role in gene regulation by the discovery of small regulatory RNAs (miRNA) in 1993 and later in 1998 by small-interfering RNAs (siRNA). These RNA classes constitute the today well-established system called RNA interference pathway (RNAi).[26,27] siRNAs derive from longer dsRNA precursor RNAs and once processed, bind to the AGO protein component of the RNA-induced silencing complex (RISC). The RNA interference pathway is based on sense-antisense RNA base pairs guiding RISC to specific locations within the mRNA to induce gene silencing by either translation inhibition or RNA degradation, thereby controlling many cellular processes.[28,29]

siRNAs are very gene specific by perfect Watson-Crick complementary to the mRNA, whereas miRNA are less precise in targeting mRNA and are therefore capable of regulating many different genes. Small regulatory RNAs have been found also in bacteria. Their function involves simple antisense mechanisms to expose or protect cis-acting sites of the mRNA by inducing changes in the secondary structure.[30,31]

In 2002, cis-acting regulatory RNA sequences, termed riboswitches, were discovered.[32,33] Riboswitches bind cellular metabolites with extraordinary selectivity to regulate the transcription and translations of proteins directly related to the catabolism or metabolism of these metabolites. Today, riboswitches are recognized as important gene control elements in diverse bacterial species, and are also found in plant and fungi.[34–36]

The large amount of research invested into deciphering the diverse functions of ncRNA is constantly revealing new classes of functional RNA. For example, long non-coding RNA (lncRNAs), roughly defined as ncRNAs larger than ~200 nt, were recently evidenced to exhibit very cell type specific expression patterns. Some lncRNAs have been found to mediate epigenetic changes by guiding chromatin remodeling protein complexes to specific genomic loci. lncRNAs are today believed to play an important role in embryonic stem cell differentiation, although their study is still in its infancy and it remains unclear if all of these RNA sequences contain structural features and are biologically relevant.[37–40] Very recently, bacterial and archaeal genomes were found to contain loci assembled from regularly spaced repeats as part of their immune system. The interspace is composed of virus-derived DNA sequences, termed clustered regularly interspaced short palindromic repeats (CRISPR).[41,42] Transcription and processing of the RNA produces small RNAs that serve as guide for their effector complex to destroy viral DNA or RNA.

Findings within the last two decades have uncovered a vast amount of ncRNAs, which exert diverse cellular function and demonstrate the importance of RNA in all forms of life today. Despite its simplicity in primary structure assembled by its four ribonucleotide units, RNA appears to be universally involved in all kind of cellular processes exhibiting extraordinarily complex molecular functions. Contrary to former beliefs, the structural complexity and selectivity of RNA rivals protein receptors and antibodies despite their structural advantage of 20 different amino acid building units. In particular, the strict requirement of RNA riboswitches to sense cellular metabolite concentrations with very high selectivity and to couple this sensor with an allosteric cellular response signal demands the formation of extraordinarily complex structural architectures. The study of these unique

RNA elements can yield fundamental insights into the plasticity of RNA and RNA folding dynamics. This thesis aims to contribute in elucidating how ligand recognition mechanisms and allosteric folding rearrangements are evolutionary fine-tuned to control gene expression.

1.2. RNA riboswitches

1.2.1 Gene regulation by riboswitches

Riboswitches are gene control elements that are typically found in the 5‘-untranslated regions of bacterial mRNAs and regulate downstream gene expression.[36] Regulated genes are directly involved in transport or metabolism of the sensed ligand that dictates the regulatory outcome. In bacteria, gene expression is commonly controlled on the level of transcription or translation. Riboswitches have also been found to operate in eukaryotes (plants and fungi), although less common, where they control gene expression on the level of mRNA processing.[43]

Regulatory mechanism

Riboswitches are composed of two allosteric domains: the metabolite sensor domain (aptamer) and the expression platform. The regulatory function of riboswitches is based on the ability of the aptamer domain to selectively recognize particular metabolites within the cell. Metabolite binding causes structural rearrangements within the aptamer domain. The adjacent expression platform directly evaluates the status of ligand binding and will interact with the aptamer domain if the ligand is not bound.[44]

Among the large variety of riboswitch classes, ligand binding to the aptamer domain can either repress (OFF-switch) or allow (ON-switch) gene expression. In case of riboswitches that regulate gene expression on the translational level, the Shine-Dalgarno (SD) sequence is directly located on the expression platform. The allosteric ligand-dependent modulation of the riboswitch controls access of the 30*S* ribosomal subunit to the SD sequence by either sequestering the SD sequence in base pairing or leaving it accessible to allow ribosome docking and translation initiation (Figure 1A). In case of transcriptional regulation, ligand binding either induces the formation of a rho-independent terminator hairpin, a mechanism commonly used by bacteria to abort transcription,[45,46] or the formation of a competing antiterminator helix, which allows transcription to proceed. In eukarya, riboswitches are

located on introns, where they control gene expression by alternative splicing. Similar to the mechanism of translational control, ligand binding can sequester or release a splice site to impede or allow splicing by the spliceosome. Among these three different levels of regulation, transcription termination represents the most prevalent mechanism of gene regulation by riboswitches. Regulation on the level of transcription is the most direct form of genetic control, since the mRNA is not unnecessarily synthesized in the first place.[47]

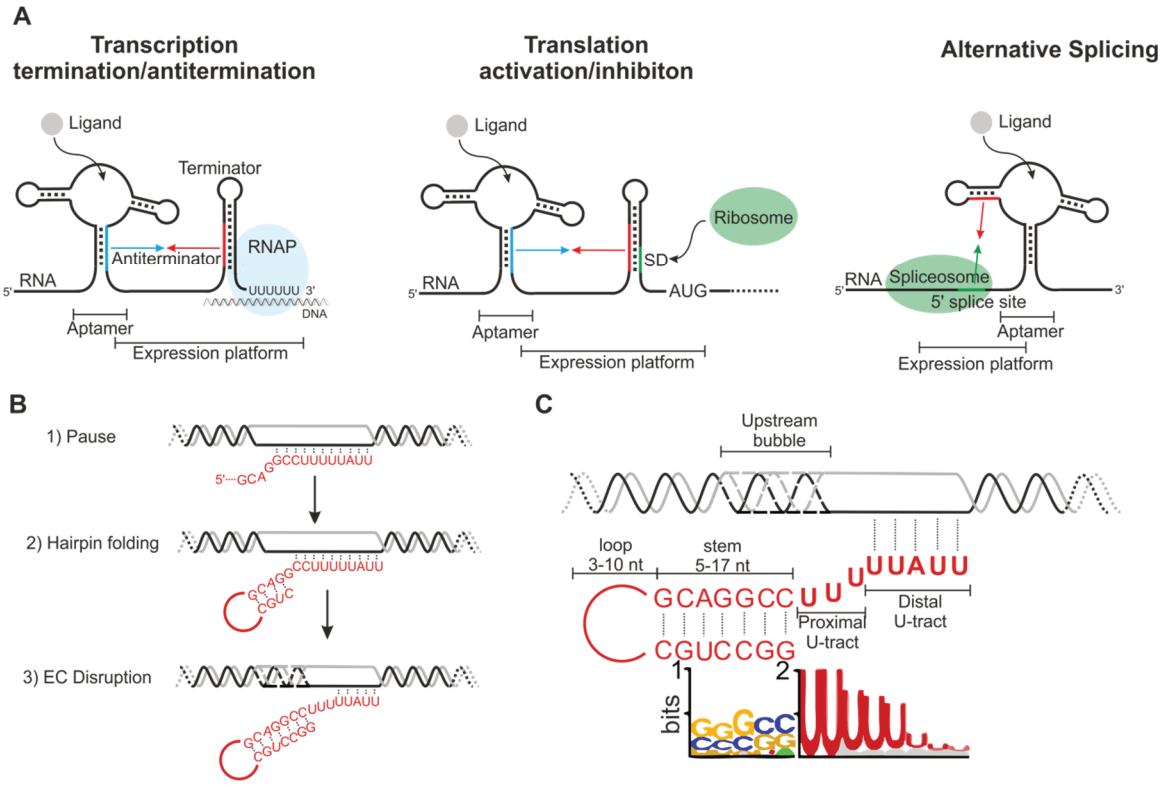


Figure 1: A) Most common established mechanisms of riboswitch-based gene regulation: Transcription termination/antitermination, Translation activation/inhibition and splicing control. B) Three step mechanism of rho-independent transcription termination: 1) Transcriptional pausing caused by the poly-U sequence followed by 2) the formation of the terminator hairpin and finally by 3) the disruption of the elongation complex. C) Sequence conservation of terminator hairpins in *E.coli*.[48,49] The extent of conserved nucleotides in the RNA stem and U-track are shown below the RNA according to Peters *et al*.[49]

Regulation of transcription is achieved by the formation of a rho-independent terminator, a mechanism of prokaryotes to achieve transcription termination without the assistance by an RNA helicase termed Rho.[50] Although the basic principle of the termination process is well-known, the structure of a stalled elongation complex has not been determined to date and the mechanistic details still remain uncertain, even after decades of studies.[51] Sequential and mechanistic features of transcriptional terminators have been primarily studied in *E.coli* with little insight into diversities between bacterial species. Available mechanistic studies on transcription termination can be summarized to the following: The

elongation complex of transcription (EC) contains a 8-10 bp long RNA:DNA duplex within a 12-14 nt transcription bubble and about 5 nucleotides of single stranded RNA in the exit channel.[52,53] Intrinsic RNA termination elements are typically characterized by a G-C rich hairpin followed by a 7 to 8 nt U-track, but may exhibit differences in their exact sequences.[45,46] The process of intrinsic termination has been suggested to occur in at least three steps involving an initial pause, followed by the formation of a termination intermediate and finally the dissociation of the EC (Figure 1B).[49] The U-track located directly adjacent to the terminator hairpin acts as a pause site causing the polymerase to temporally stall, and generates a time window for terminator hairpin folding. The terminator hairpin extends in the exit channel to the paused U-track and causes melting of ~3 bp within the DNA:RNA duplex,[54,55] destabilizing the elongation complex to the point of RNA dissociation. Figure 1C shows sequential features of intrinsic terminators. In *E.coli*, terminators vary from 5 to 17 bp in length with a strong preference of GC base pairs at the five nucleotides prior to the U-track. Terminator loops vary from 3 to 10 nt, with 4 nt being the most common loop size. The strong preference for tetra-loops in terminator hairpins is based on these loop architectures, particularly the UNCG and GNRA family, to significantly contribute to the formation of very stable hairpins.[56,57] The stem-loop structure is followed by at least two Us, with a high conservation for U in the following 3 nucleotides (positions +3 to +5). Nucleotides at position +6 to +8 exhibit much more sequential variety. It has been suggested that imperfect U-tracks at positions +4 to +8 (distal U-tract) can be compensated by a high % of A/U residues further downstream (positions +10 to +12). This distinction in conservation likely reflects two classes of terminators with differences in their detailed mechanism of action.[49]

Riboswitch classes

Riboswitches have been found in a variety of organisms including well-studied bacteria such as *B.subtilis* and *E.coli*, where they control fundamental metabolic pathways by directly sensing fluctuations in related metabolite concentrations. In *B.subtilis*, the number of known riboswitch classes is greater than the number of validated protein factors known to bind metabolites.[58] According to the current state of research, 24 different riboswitch classes have been validated (Figure 2). Riboswitches are assigned to specific classes based on their sequence conservation, which is directly related to the metabolite they sense.

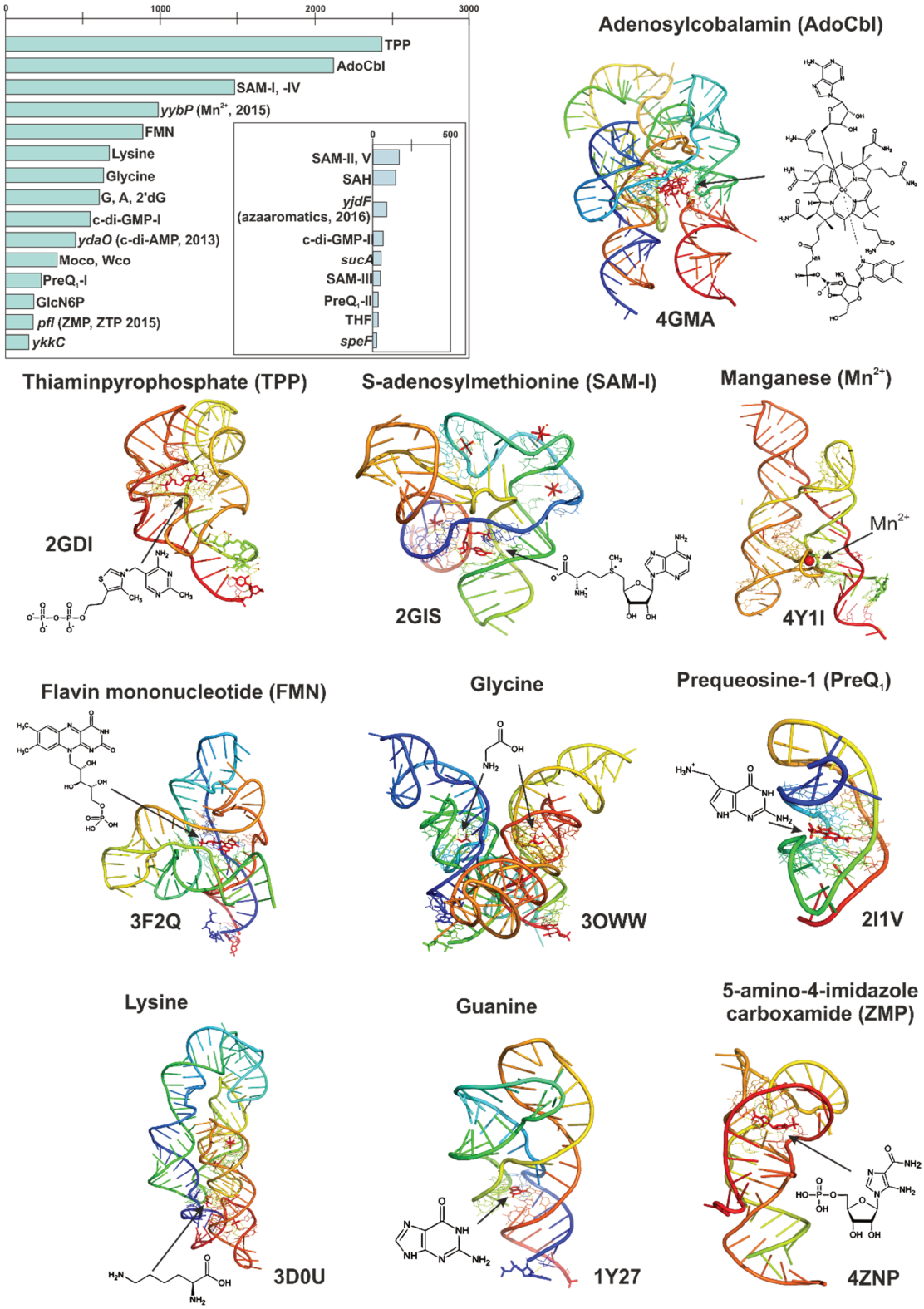


Figure 2: Riboswitch class representatives of validated riboswitches relative to their occurrence in genomes of ~700 bacterial species according to Breaker *et al.*, 2012[58] and recently discovered riboswitches binding c-di-GMP-I,[59] ZMP,[60] Mn,[61] and possible azaaromatic compounds.[62] Reported crystal structures of the most abundant riboswitches highlighting the global fold and ligand binding architecture of the *btuB* AdoCbl riboswitch from *E.coli*,[63] the *thiM* TPP riboswitch from *E.coli*,[64] the SAM-I riboswitch from *T. tengcongensis*,[65] the Mn^{2+} riboswitch from *Lactococcus lactis*,[66] the FMN riboswitch from *Fusobacterium nucleatum*,[67] the glycine riboswitch from *V. cholera*,[68] the PreQ1 riboswitch from *B.subtilis*,[69] the *asd* lysine riboswitch from Thermotoga maritima,[70] the *xpt* guanine ribowitch from *B.subtilis*,[71] the *pfl* ZMP riboswitch from *Thermosinus carboxydivorans*.[72]

The most abundant riboswitch class binds thiamin pyrophosphate (TPP),[32,34] the only validated riboswitch class found also in plant and fungi to control alternative splicing of precursor mRNAs.[73] Apart from the TPP riboswitch, the ten most common riboswitch aptamer classes sense adenosylcobalamin (AdoCbl),[74] S-adenosyl methionine (SAM),[75] manganese,[61] flavin mononucleotide (FMN),[32,33] lysine,[76] glycine,[77] guanine/adenine/2'-dG,[1,35,78] the bacterial second messengers c-di-GMP[79] and c-di-AMP.[59]

Validated riboswitch classes are assumed to only represent a small proportion of existing riboswitch classes and contain some motifs, such as *sucA* and *speF*, which have not been assigned to a specific ligand yet. New riboswitch classes are continuously uncovered along with the identification of new types of ligands bound by aptamers. One of the most abundant riboswitch classes, the *yybP-ykoY* motif was only very recently identified to respond to Mn^{2+}, and is assumed to regulate the expression of a membrane protein involved in metal homeostasis.[61,66] The extremely common *ydaO* riboswitch motif, in 2012 hypothesized to bind ATP,[80] was recently reported to bind c-di-AMP with higher affinity.[59] Further, the *pfl* motif was found to respond to 5-aminoimidazole-4-carbocamide ribonucleotide (ZMP), a precursor for the putative alarmone 10f-THF, only last year.[60] For the *yjdF* motif, the precise natural ligand still remains unknown but the motif has been reported to respond to azaaromatics, possibly regulating a response system to toxic compounds with chemical structures similar to azaaromatics.[62] To summarize, riboswitches bind a collection of ligands including coenzymes/derivatives, metal ions, amino acids, purine nucleobases, an amino sugar, and regulate the biosynthesis and transport of these essential metabolites.

Among the discovered riboswitch classes, few utilize very unique mechanisms to control gene expression. For example, a riboswitch binding glucosamine-6-phophsate (GlcN6p) acts more like a ribozyme-riboswitch, regulating gene expression by metabolite responsive self-cleavage that leads to selective degradation of the mRNA coding region.[81,82] Further, the dynamic range of gene regulation can be significantly improved or fine-tuned by stacking two riboswitches in tandem. Two consecutive riboswitches responding to the same ligand lower the ligand concentration required to achieve the same effective gene control by one half.[83] In the glycine riboswitch, two aptamer domains are stacked in tandem. Two contradicting mechanisms of action have been proposed for this riboswitch. The first model proposes cooperative ligand binding, where glycine binding to one aptamer domain increases the ligand affinity of the second aptamer domain.[68,84,85] In contrast, a more recent

study suggests that dimerization of the two domains lowers the energy barrier for ligand binding of the regulatory relevant second aptamer domain.[86,87] In the 5'-UTR of the *Bacillus clausii metE* mRNA two riboswitches with different ligand selectivities are stacked in a tandem arrangement to achieve gene repression by either coenzymes AdoCbl or SAM.[85] In addition, SAM-II and SAM-IV riboswitches have been found to contain both transcription and translational control elements within one riboswitch.[88]

There are also exceedingly rare riboswitch classes, such as the 2'-deoxyguanosine (2'dG) sensing riboswitch (dGsw). The architecture and binding mode of this riboswitch is highly related to guanine riboswitches, which are among the most common riboswitch classes. Despite their similarity, currently only two 2'dG selective riboswitches have been discovered in only one bacterial species *Mesoplasma florum*. 2'dG represents an essential and universal monomer for DNA synthesis in all living organisms, raising the question of why this riboswitch class is found only in one bacterial species.[89]

Structurally, riboswitches are assembled of recurring motifs including tertiary interactions. Secondary structure elements and nucleotides relevant for ligand recognition, particularly in the core of the binding pocket, are typically strikingly well conserved in aptamer domains. The high sequence conservation appears to be imperative to form highly selective binding pockets for specific metabolite interactions, with only four types of nucleotides available as building units. In contrast, the expression platform varies largely in sequences due to the different processes that are controlled by riboswitches.[73,74,90,91] Some riboswitches are extremely conserved in their sequence and architecture, such as the TPP riboswitch.[35] In contrast, the precursor of quenosine 1, prequenosine (PreQ1), is bound by two classes with different ligand-binding folds,[92,93] and the second messenger c-di-GMP is sensed by both c-di-GMP-I and c-di-GMP-II, which have a very diverse global architecture.[79,94]

Most typical global structural features of riboswitches are kink-turns, ribose zippers, and E-loops, which cause bending to allow long-range tertiary interactions (Figure 2). Some riboswitches, such as the SAM-II, $PreQ_1$ and ZMP riboswitch, adopt pseudoknots when bound to their ligand.[69,95–97] These motifs stabilize a global arrangement, typically composed of parallel aligned helical bundles with coaxially stacked base pairs. As illustrated in Figure 2, the TPP, SAM-I, manganese and guanine-sensing riboswitches exhibit typical coaxial stacking of two helices, while in the lysine riboswitch, three helices are stacked coaxially.[98] The binding pocket is composed of a multi-way helical junction

and fully encapsulates the ligand. The ligand is stabilized by a network of hydrogen bonds supported by stacking and/or electrostatic interactions.[99]

Riboswitches are classified into two types: Type I RNAs feature a structurally ordered ligand-free state, where an initial encounter complex with the ligand is followed by only small local changes in conformation upon tight ligand binding.[100–102] Purine-sensing riboswitches (adenine and guanine) are model representatives of this class. Type II RNAs contain unfolded aptamer domains in the absence of ligand,[65] and ligand binding leads to global and distal conformational changes, such as the GNRA like tetraloop docking of L5 to helix P3 in the TPP riboswitch.[34,64,103]

1.2.2 Thermodynamic and kinetic components of riboswitch-based gene regulation

Dissociation constants K_D determined for riboswitch aptamers vary from 10 pM for the c-di-GMP aptamer from *Vibrio cholera*[104] to ~200 µM for a GlmS riboswitch.[81] These K_D values manifest a discrepancy between much higher cellular metabolite concentrations and the concentrations actually needed by the riboswitch to allow ligand binding according to equilibrium K_D values. For transcriptional riboswitches, single round *in vitro* transcription assays have shown that much larger ligand concentrations are required to achieve a half maximal regulatory response (T_{50}) in comparison to apparent K_D values.[76,105,106]

Experiments published in 2005 on the FMN riboswitch from *B.subtilis*, which is located prior to the ribDEAHT operon, encoding genes responsible for riboflavin biosynthesis, provided the first evidence for the importance of kinetics in transcriptionally modulated gene regulation by riboswitches.[2] Ligand binding to the FMN riboswitch induces premature termination of transcription (OFF switch). In this study, the FMN concentration required to induce a half maximal modulation of transcription (T_{50}) was determined at various parameters influencing the speed of transcription, such as the nature of the polymerase, NusA, a protein transcription factor that has been shown to extend the time RNA polymerase spends on pause sites during transcription,[107] and the NTP concentration. In addition, pause sites were mapped in single round transcription assays and mutated to show that in the absence of transcriptional pausing, higher FMN concentrations are required to trigger riboswitch function. As a general finding, reducing the speed of transcription by RNA polymerase leads to increased transcription termination and vice versa. This observation can be explained by an extension in the time window for ligand binding at slower transcription rates. Further, pause sites appear to be intentionally positioned adjacent

to the aptamer domain to generate time for ligand binding. Within the time frame of transcription, the riboswitch cannot equilibrate with its cognate ligand. In this riboswitch, a 10-fold higher ligand concentration than suggested by the K_D is necessary to trigger riboswitch function.

Another study by Block *et al.* in 2012 on the *pbuE* adenine-sensing riboswitch from *B.subtilis*, a riboswitch that terminates transcription at low adenine concentrations (ON-switch), used single molecule force spectroscopy to monitor co-transcriptional folding by fully extending the RNA under high force, measuring folding states through transcript extension.[3] Their experiments suggest that ligand binding to the aptamer domain can only occur once during transcription or never due to conformational lock down by the more stable terminator conformation, while ligand-binding stabilizes the aptamer domain fold sufficiently to prevent terminator formation. However, experiments performed under these conditions do not directly observe conformational rearrangements during transcription and therefore only yield a vague insight into co-transcriptional folding. Further, co-transcriptional folding has been modeled for the SAM-I and *pbuE* adenine-sensing riboswitch by Lutz *et al.*[108] In their theoretical approach, they have implemented strict spatial restraints arising from RNA-polymerase interactions as the nascent RNA chain leaves through the exit channel. Similar to the study of Block *et al.*, they suggest that substructures fold as they are transcribed and may be disrupted by newly synthesized RNA unless previously synthesized RNA is stabilized by ligand binding.

Regulation on the level of transcription already implies a kinetic parameter because transcribing past the 5'-UTR is an irreversible decision by the polymerase. In contrast, in translational riboswitches the full mRNA may first be synthesized, and ribosome access can be controlled in a later step. However, in bacteria transcription and translation are coupled, raising the question if translationally controlled riboswitches are, at least in part, also kinetically controlled. Lafontaine and coworkers have directly compared the *pbuE* adenine-sensing transcriptional riboswitch from *B.subtilis* with the translational *add* adenine-sensing riboswitch from *V.vulnificus*.[4] Both systems represent ON-switches that allow gene expression at high adenine concentrations. By means chemical probing and 2-AP fluorescence quenching, they find that adenine binds to the isolated aptamer domains with similar dissociation constants. However, adenine binding to the full-length riboswitch to induce an allosteric conformational switch can only be observed for the *add* riboswitch.

Similar to the FMN riboswitch, NTP concentrations were found to be proportional to the adenine concentration required to achieve a half maximal modulation of transcription termination for the *pbuE* riboswitch. In addition, the authors suggest that NusA promotes ligand binding by extending the time RNAP pauses after transcription of the aptamer domain. Regarding translational control, *in vitro* reporter gene assays performed with the *add* riboswitch by using either mRNA or DNA as a template for gene expression suggest that the ligand-dependent modulation of gene expression remains the same regardless of whether transcription or translation are coupled. At least in this model system, the results imply that the riboswitch has sufficient time to equilibrate in a coupled transcription and translation framework.

The reversibility of ligand binding in translational riboswitches can be derived from the small energy difference between the two allosteric states, including usually only few base pairs difference. For the *add* adenine riboswitch, a three-state mechanism has recently been proposed, which compensates for increased ligand binding at lower temperatures to keep the switching efficiency constant over a broad temperature range.[109] In this system, energy differences between allosteric states are strictly fine-tuned to ligand-binding and temperature variations thermodynamically (Figure 3A). In contrast, high thermodynamic stability is a critical feature of intrinsic terminators to execute their function (see Figure 1C), and ligand binding to the aptamer domain, which accounts for about -8 kcal/mol, can usually not compensate this energy difference. While from a thermodynamic perspective, ligand binding is more efficient at lower temperatures, the speed of ligand association (k_{on}) rather than the K_D represents the decisive functional parameter in kinetically controlled riboswitches, and these rates increase with higher temperatures. Increased ligand association rates may, at least in part, already be compensated by an increase in the polymerase speed at higher temperatures.[110,111]

Combined, these data support a coupling of decision making and co-transcriptional folding in transcriptional riboswitches. The term “switch” may be misleading in transcriptional riboswitches, as it implies an interconversion between different folds responding to fluctuations in the metabolite concentration. In contrast, the RNA is more accurately biased towards one of the two folding pathways that ultimately dictates the genetic decision (Figure 3B).[2] The aptamer domain is always transcribed first and the time for ligand binding is limited. Ligand binding is suggested to lock the aptamer conformation to prevent an interaction with the expression platform, and therefore must occur before the expression

platform is synthesized. However, detailed mechanistic studies on this co-transcriptional event are not available. Further, it appears that refolding to form an interaction between aptamer and expression platform (ON switch: terminator, OFF switch: antiterminator) in the absence of ligand must also be limited in time, but this folding rearrangement has not yet been investigated either. A very large number of studies have been performed on isolated aptamer domains. However, the precise programming in time regarding ligand binding, pausing and folding rearrangements by transcriptional riboswitches demands the study of mechanistic details in the context of transcription.

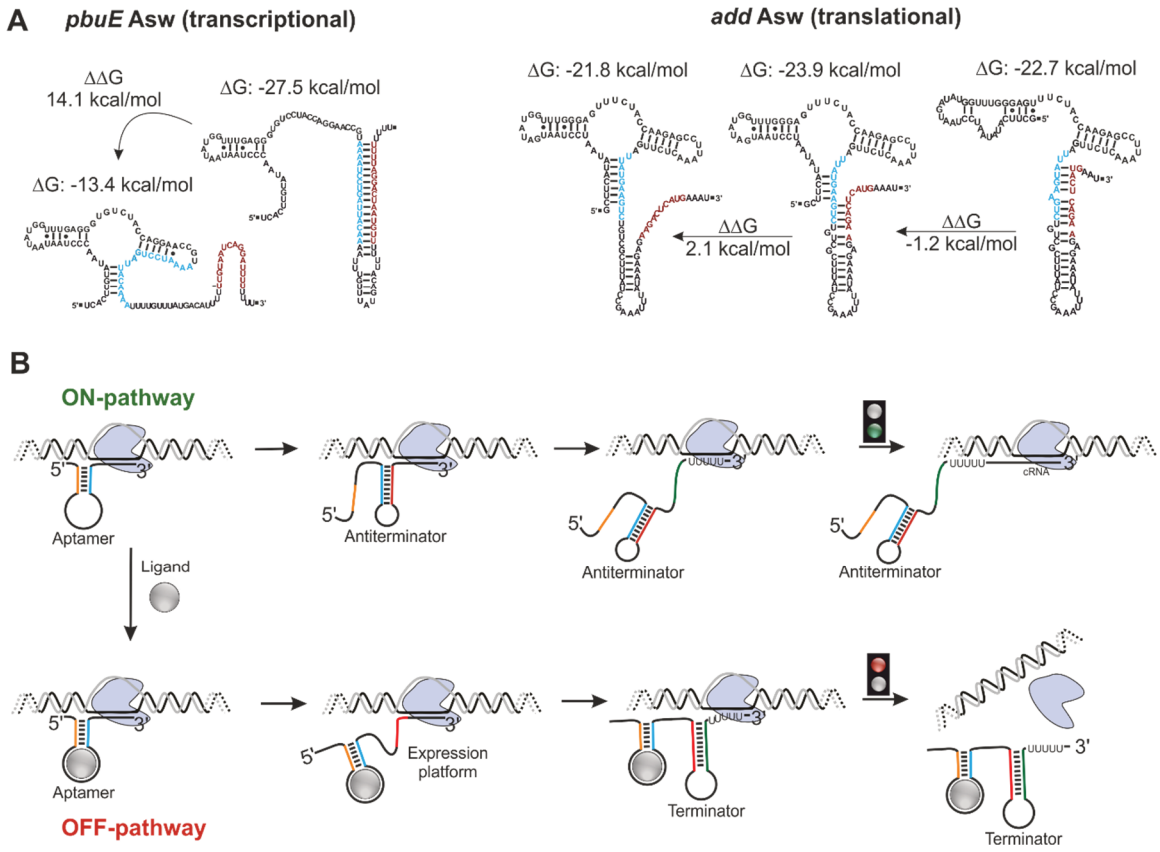


Figure 3: A) Secondary structure of the regulatory decisive conformations in the *pbuE* adenine-sensing riboswitch and *add* adenine-sensing riboswitch.[71] Approximate free energies according to mfold[112] for the particular states and energy differences between the states are indicated accordingly. B) Exemplary model for kinetic regulation for OFF-switches. During transcription, the aptamer domain is always transcribed first and metabolite binding needs to take place directly thereafter to trap the aptamer domain conformation. Otherwise, the antiterminator conformation is transcribed, which is believed to inhibit ligand binding and lead to the opposite regulatory outcome.

The mechanism of kinetic regulation is contrary to most ncRNAs (as outlined in Chapter 1.1) and proteins, which typically adopt a distinct functional fold to respond to the cellular environment. The complex mechanism of coupling synthesis and folding to cause a default biological outcome may be an advantage of kinetically controlled riboswitches compared to protein genetic factors because it adds an additional functionality for fine-tuning responses to different metabolite concentrations.[88] Concentrations required to trigger riboswitch function may be further adjusted by changing ligand association rates through aptamer mutations, by altering the amount of nucleotides incorporated in the linker connecting aptamer domain and expression platform to change the speed of riboswitch synthesis, or by designing pause sites at distinct locations to generate time for ligand binding or refolding.

1.2.3 Purine-sensing riboswitches

As outlined in section 1.2.1, many riboswitches recognize metabolites that contain a purine or related moiety. Among riboswitches binding purine related metabolites, the class binding guanine, adenine and 2'dG represent the most well-studied riboswitch class. While adopting intriguingly similar architectures, guanine and adenine riboswitches are distinct in their ability to discriminate between many chemically closely related metabolites available in high concentrations in the cell. Due to their small size, they serve as model system to decipher regulatory function by riboswitches.[35,78] Purine riboswitches control fundamental metabolic pathways in purine transport, synthesis and salvage.[35] Figure 4 shows the metabolic pathway of purine *de novo* synthesis in *B.subtilis*,[113] while highlighting the pathways controlled by the most well-studied transcriptional riboswitch model systems in *B.subtilis*: the *xpt* guanine-sensing riboswitch (*xpt* Gsw),[100] and the *pbuE* adenine-sensing riboswitch (*pbuE* Asw).[78]

The *xpt* Gsw represents an OFF-switch, terminating transcription at high guanine concentrations, while the *pbuE* Asw represents an ON switch that terminates transcription at low adenine concentrations. The *xpt* Gsw binds guanine, hypoxanthine and xanthine, therefore straightforwardly repressing the expression of xanthine phosphoribosyltransferase, which catalyzes the conversion of xanthine to XMP, provided the concentrations of these metabolites are high within the cell to avoid excessive synthesis of XMP. Similarly, the *pbuE* Asw is assumed to repress the expression of a putative

hypoxanthine exporter at low adenine concentrations, such that hypoxanthine is maintained within the cell for ATP synthesis.

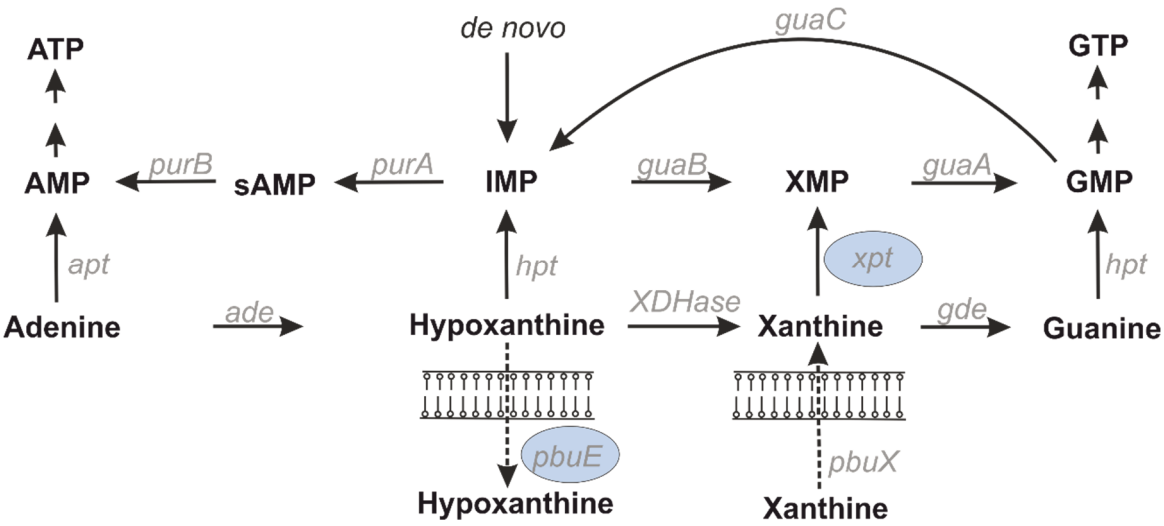


Figure 4: Purine interconversion and catabolic pathway in *B.subtilis*. Enzymes involved in metabolic interconversion and transport are indicated according to their genes: *apt*, adenine phosphoribosyl-transferase; *hpt*, hypoxanthine-guanine phosphoribosyltransferase; *xpt*, xanthine phosphoribosyltransferase; *ade*, adenine deaminase; *guaB*, IMP dehydrogenase; *guaA*, GMP synthetase; *purA*, adenylosuccinate synthetase; *purB*, adenylosuccinate lyase; *guaC*, GMP reductase; *pbuE*, putative hypoxanthine exporter; *pbuX*, xanthine permease.[113]

The secondary structure of conformations leading to gene expression (ON-state) and terminating gene expression (OFF state) for the *xpt* Gsw and *pbuE* Asw are depicted in Figure 5A,B.[71,102] In analogy to the general allosteric modulation performed by riboswitches (Figure 1), the sequence shared between aptamer and expression platform (in blue), termed switching sequence, represents the crucial element for regulatory signaling. The switching sequence is typically located at the boundary of the two domains, and is either involved in aptamer domain formation or acts as part of the expression platform to base pair with the strand highlighted in red. In the *xpt* Gsw, it forms the antiterminator helix (blue-red) in the absence of ligand, while in the *pbuE* Asw, it forms the terminator helix (blue-red) in the absence of ligand.

Guanine and adenine riboswitch aptamers have a very high sequence conservation (nucleotides highlighted in red within the aptamer domain in Figure 5) and adopt the same secondary structure composed of a three-way helical junction connecting three helices P1, P2, and P3.[35,78] Base paired helical segments exhibit little sequence conservation, while sequences of loops L2, L3 and particularly the joining region between helices (J1/2, J2/3, and J3/1) are highly conserved. Purine-sensing riboswitches are representatives of class I riboswitches that do not undergo major structural changes upon ligand binding and in

consequence many conserved nucleotides have been found be relevant to keep an open conformation.[114,115] The loops L2 and L3 associate through tertiary interactions with the two G-C base pairs G38-C60 and G37-G61.[114,116] Loop-loop interactions align helices P2 and P3 parallel to each other, arranging the global architecture into a typical three-way junction stabilized by distant tertiary interactions.[117] The three-way junction represents the structural motif directly involved in ligand recognition and binding.

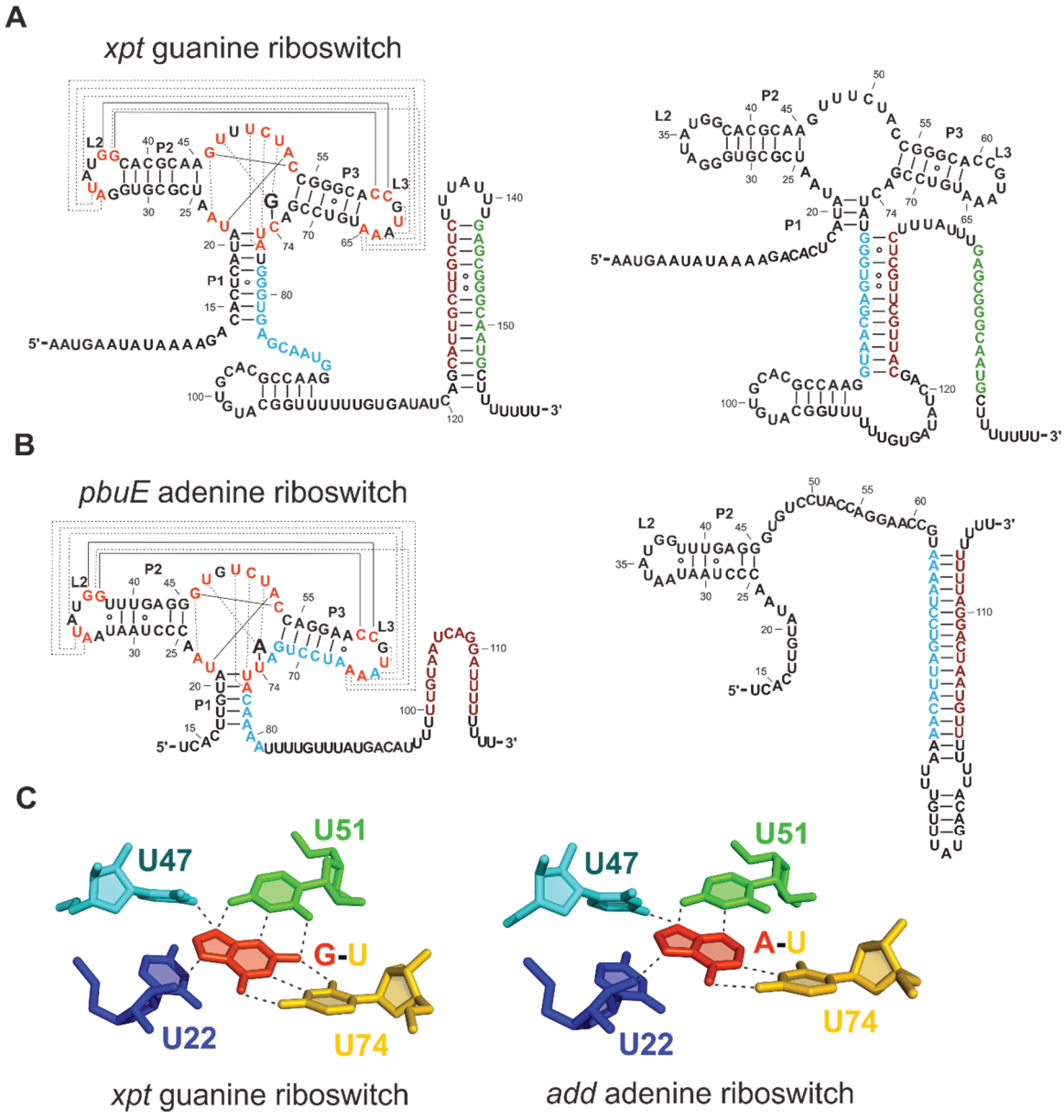


Figure 5:Secondary structure of the ligand-bound and ligand-free state of the *xpt* guanine-sensing (A) riboswitch and the *pbuE* adenine-sensing riboswitch (B).[71] Tertiary Watson Crick base pairs in the binding pocked and within loops are connected with solid lines. Tertiary hydrogen bond interactions are indicated by dashed lines. C) Hydrogen bond interactions with the ligand in the bound state for the *add* adenine riboswitch (1Y26) and *xpt* guanine riboswitch (1Y27).[71]

It has been shown that loops L2 and L3 do not alter their conformation as guanine, hypoxanthine or xanthine are bound to the *xpt* Gsw. However, the region joining helices becomes more structured upon ligand binding. The binding pocket is tightened by coaxial stacking of helices P1 and P3 and the formation of several base-triples formed by conserved nucleotides. The critical nucleotide for ligand recognition is nucleotide 74, which forms a Watson-Crick base pairing with the ligand and has been confirmed to universally be a uridine in adenine-sensing riboswitches or a cytosine in guanine-sensing riboswitches by both X-ray crystallography and NMR (Figure 5C).[71,102] A single point mutation at this particular nucleotide alters the ligand affinity from guanine to adenine and vice versa.[118] The ligand is further recognized by U51, which together form a base-triple with the ligand including additional hydrogen bond stabilization of N7 by 2'-OH of U22. Within helix P1, two additional base-triples, A21-U75-C50 and U20-A76-U49 further stack with the ligand base-triple. These tertiary interactions stabilize helix P1, and sequester the 3'-P1 switching segment to prevent an interaction with the expression platform in the ligand-bound state.[119] In the ligand-bound state, the ligand is closely enveloped by RNA and is 98% solvent inaccessible. While ligand binding does not lead to major global structural changes, Mg^{2+} has been shown to induce a large compaction in structure, and is essential for stabilizing tertiary loop-loop interactions.[101,120,121]

1.2.4 Riboswitches in the bacterium *Mesoplasma florum*

The genome of the strain *Mesoplasma florum L1* was sequenced completely in 2004 by Knight *et al.* [RefSeq NC_006055.1]. *M. florum* were isolated from the flower of a lemon tree and are assumed to be associated with plant insect vectors. The bacterial species belongs to the class of Mollicutes, which are characteristic for their lack in a cell wall and low G+C content.[122] Mollicutes have evolved from gram positive bacteria by gradually reducing the genome size to <1 Mb. They represent one of the smallest living organisms with limited biosynthetic abilities.[123] Due to the reduction in genome size, they lack many genes involved in metabolic pathways and in transcription regulation. This deficit causes a dependency on their host organism to supply them with essential nutrients.[124] In comparison to *E.coli* and *B.subtilits*, they lack many DNA repair proteins, which allows quick incorporation of mutations to rapidly evolve and adapt to different environmental conditions. The genome of *M.florum* is circular with a size of about 8 kb and a G+C content of 27.02%.

Eight aptamer sequences were discovered in 2007 in the *M.florum* genome by bioinformatic search algorithms based on their closely related consensus sequence to known purine-sensing aptamer domains (Figure 6).[1] Two of these riboswitches were found to exhibit a strong selectivity for 2'-deoxyguanosine over guanine or guanosine (200 fold), being the first organism found to regulate synthesis of DNA building units by a riboswitch. In-line probing analysis of a variety of metabolites show that the I-A and I-B aptamers selectively bind 2'dG with an apparent K_D of about 80 nM.[1] The I-A aptamer sequence exhibits 39 nucleotide changes compared to the *xpt* guanine riboswitch, among which 10 nt are typically highly conserved in purine aptamers. Particularly nucleotides U22, U47 and U51 that directly interact with the ligand are mutated to C22, A47 and C51. Stems P2 and P3 are swapped in length and loop L3 is shortened by three nucleotides, typically conserved and involved in critical loop-loop interactions. The aptamer domain of the type I-A 2'dG riboswitch has been extensively studied by in-line probing, X-ray crystallography and NMR spectroscopy.[125–127] All studies reveal the U51C mutation to be crucial to achieve a discrimination between 2'dG and guanine. C51 is shifted towards C74 to generate space for the extra ribose moiety in 2'dG in comparison to guanine riboswitches (Figure 6B).

In addition, the U47A mutation has been shown to moderately affect ligand affinity, and by NMR proposed to remain flexible in solution contrary to the U47-N9 adenine/guanine interaction. X-ray data revealed C49 to be rotated outwards, thereby preventing the base triple with U20-A76 in P1, universally present in guanine and adenine riboswitches.[128] This structural feature in guanine and adenine riboswitches typically fixes the J2/3 lid to P1 to tightly close the binding pocket, which appears to be more flexible in case of 2'dG-sensing riboswitches. In fact, by means of NMR no interaction between J2/3 and P1 could be detected.[126]

Further, the shortened loop L3 causes critical changes in tertiary loop-loop arrangements. G-C base pairs G37-C61 and G38-C60 remain conserved, but the 2'dG riboswitch lacks three canonical base pairs including stacking interactions to stabilize loop-loop interactions. X-ray data propose a "key-and-lock" element to stabilize parallel helix alignment, where A65 represents the key to lock into L2, stacking on A34. P3 is further stabilized by additional stacking interaction and P2 by a G24-G25-G46 base-triple, also confirmed by NMR to be in close proximity, which cannot be observed in adenine and guanine riboswitches.[126,127]

Intriguingly, the 2'dG riboswitch discriminates against guanosine by two orders of magnitude, which differs only by a single hydroxyl group. It has been proposed that this extra 2'OH moiety can only be accommodated in the binding pocket in the unfavorable C3'-endo conformation (pdb: 3SKZ). This conformation is incompatible with the C49-3'OH interaction causing a decrease in guanosine affinity.[127]

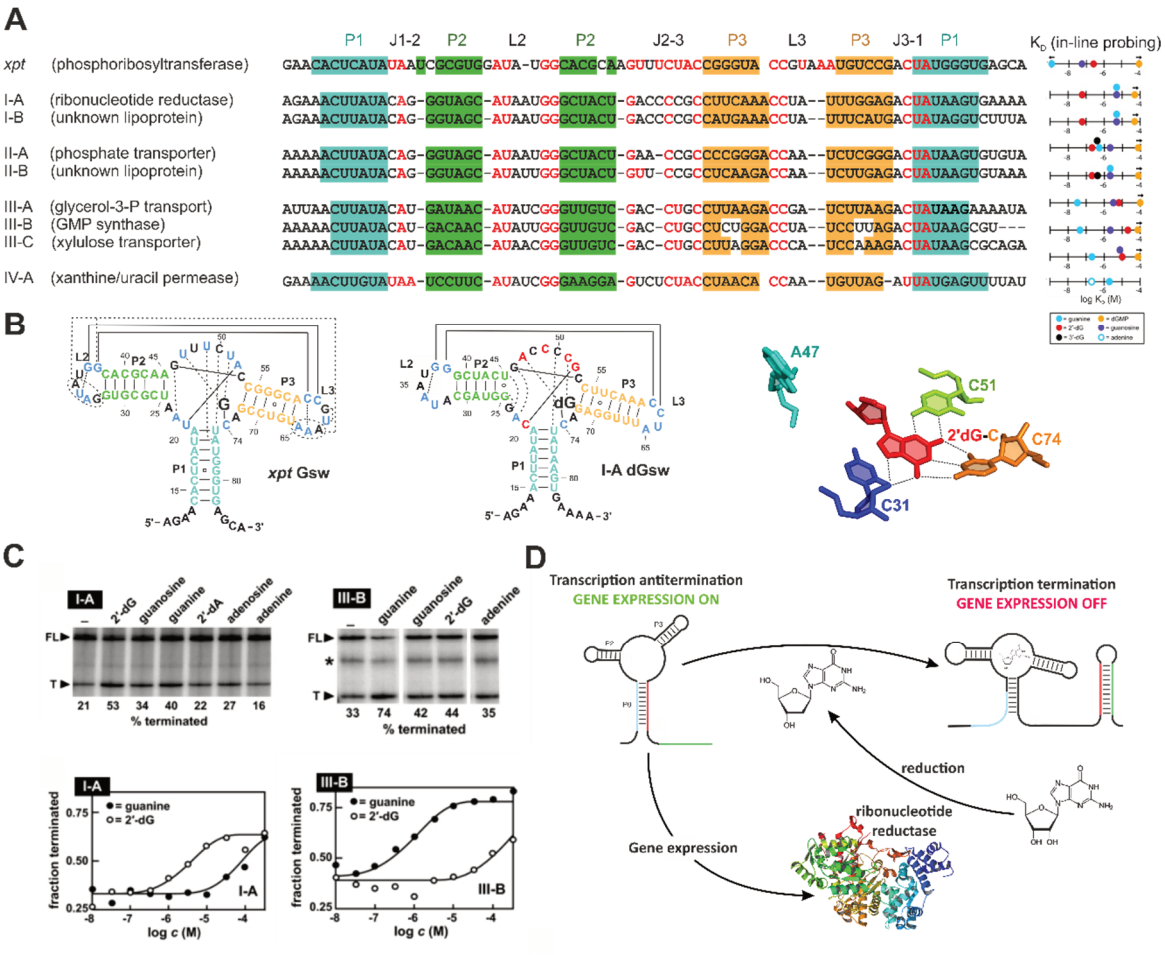


Figure 6: A) Sequence of the eight aptamers discovered in *M. florum* and comparison to the consensus sequence of the guanine-sensing riboswitch. P1 is highlighted in cyan, P2 in green, and P3 in orange. Conserved nucleotides from the *xpt* aptamer are marked in red. The K_D determined by in-line probing for guanine (blue circle), 2'dG (red circle), 3'dG (black circle), dGMP (orange circle), guanosine (purple circle) and adenine (open blue circle) are indicated on the right of the sequence according to Kim *et al.*[1] B) Comparison of secondary structure of the *xpt* guanine-sensing aptamer domain with the I-A 2'dG-sensing aptamer domain. Conserved nucleotides are color coded in blue in the *xpt* aptamer and deviations from this conservation in red in the 2'dG aptamer. Solid lines represent Watson-Crick base pairs and dashed lines represent non-canonical hydrogen bond interactions. The binding mode according to the published crystal structure (3SKI) of the I-A 2'dG riboswitch is shown on the right.[128] C) Transcription termination assays performed for the I-A 2'dG-sensing and the III-B guanine-sensing riboswitch. Gels show results of single round transcription assays in the presence of various purine metabolites. Graphs show the change in fraction terminated with increasing 2'dG or guanine concentration.[1] D) Biological role of the 2'dG-sensing riboswitch. At high 2'dG concentrations, transcription is terminated. At low 2'dG concentration transcription progesses to synthesize the coding strand for ribonucleotide reductase associated genes (RNR subunit R1 shown 1RLR[131]), which catalyze the reduction of ribonucleotides to deoxyribonucleotides.

Another interesting characteristic of the 2'dG-sensing riboswitch appears to be its strict discrimination against closely related metabolites. NMR studies have revealed that the *xpt* Gsw readily binds near-cognate ligands such as 2'dG, hypoxanthine and xanthine in solution, whereas the 2'dG riboswitch only binds 2'dG.[129] It has been proposed that the 2'dG riboswitch requires a more strict regulation against closely related metabolites due to the low cellular concentration of 2'dG (µM range in *E.coli*[130]) in comparison to guanine (mM in *E.coli*[130]). k_{on} rates, the critical parameter for kinetic riboswitch regulation (see Chapter 1.2.2), for 2'dG and guanine binding to the *xpt* Gsw are virtually identical. However, 2'dG may have a limited effect on regulation by *xpt* Gsw due to the low cellular 2'dG concentration, which is accompanied by slow ligand binding. Kinetics for 2'dG binding to the 2'dG riboswitch may be faster due to the less extensive structural rearrangement induced by ligand binding, which conforms with the more flexible nature of the binding pocked as described above.[129]

M.florum riboswitches of the type III selectively bind guanine over various nucleoside derivatives by 100-fold or more. The K_D, however, is about 100-fold lower than for the *xpt* Gsw. Nucleotide conservations in III-B riboswitches suggest many similarities in global architecture to the 2'dG riboswitch. For example, the shortened L3 and lengths of helices P2 and P3 in type III-B riboswitches are identical to the 2'dG riboswitch, suggesting that the global architecture of these riboswitches is a bacterium specific feature. In contrast, conserved nucleotides relevant for recognition of nucleobases opposed to nucleosides show more similarities to the *xpt* Gsw. In general, *mfl* type III guanine-sensing riboswitches appear to resemble a hybrid of guanine and 2'dG-sensing riboswitches, containing the critical U51 in J2/3 for N9 recognition, but lacking the moderately important U47 present in guanine and adenine riboswitches.

All of the aptamer domains discovered in *M. florum* are located upstream of a terminator stem, suggesting that all of these riboswitches regulate gene expression on the level of transcription. Kim *et al.* have performed *in vitro* transcription termination assays with *E.coli* polymerase for the I-A 2'dG-sensing riboswitch and the III-B guanine-sensing riboswitch. These transcription termination assays show roughly 53% of termination for the I-A 2'dG-sensing riboswitch at 25 µM 2'dG and 74% of termination for the III-B guanine-sensing riboswitch at 25 µM guanine (Figure 6C). In the absence of both ligands, transcription is still terminated by 21% for I-A dGsw and 33% for III-B Gsw, respectively.

This result implies that the time frame for antiterminator folding is in fact relevant and limited, since not all RNAs fold to the antiterminator conformation within the given window of time during transcription to prevent terminator formation. In addition, guanine causes a relatively large degree of termination for the I-A 2'dG-sensing riboswitch with 40% opposed to 53% for 2'dG. However, in line with its biological function, the riboswitch strikingly resents guanosine to a larger degree, despite guanosine being structurally more related to 2'dG than guanine. This result again emphasizes the importance of ligand binding kinetics for transcriptional riboswitch function, since the K_D determined for guanosine and guanine binding to I-A dGsw are identical. Precise termination efficiencies may differ *in vivo* in the presence of RNA folding proteins, such as NusA, ions or the native polymerase that in combination cause a difference in the kinetic framework.
Figure 6C further shows a dependency of terminated RNA fractions with increasing ligand concentration for the two riboswitches. The concentration needed to trigger half maximal modulation of transcription termination T_{50} is approximately 2 µM for the 2'dG riboswitch, which is much higher than the apparent K_D of 80 nM. In line with the parameters relevant for kinetically controlled riboswitches as outlined in section 1.2.4, the riboswitch therefore appears to be under kinetic control.

The I-A 2'dG-sensing riboswitch regulates the transcription of a polycistronic mRNA coding region including ribonucleotide-diphosphate reductase subunit alpha (Mfl528), a ribonucleotide reductase stimulatory protein (Mfl529) and ribonucleotide-diphosphate reductase subunit beta (Mfl530). The riboswitch is located 42 nucleotides upstream of the start codon initiating translation of the ribonucleotide reductase subunit alpha. The combination of these proteins are required to generate 2'-deoxyribonucleotide DNA building blocks from ribonucleotides. Consequently, expression of ribonucleotide reductase genes is repressed if the cellular concentration of the 2'-deoxyribonucleotide 2'dG is sufficiently high in a negative feedback loop (Figure 6D). The physiological role of the 2'dG-sensing riboswitch also explains why such a high degree of discrimination between the remarkably closely related metabolites guanosine and 2'dG is mandatory. Type II RNAs also have highest affinity for 2'dG but are less selective, possibly to deliberately allow a modulation of gene expression by the presence of a variety of metabolites.

Chapter II

Materials and Methods

2.1. Biochemical

2.1.1 DNA template preparation

DNA templates for *in vitro* transcription reactions were prepared by PCR from linearized or circular plasmid DNA. Plasmid DNA templates for I-A dGsw, I-B dGsw and III-B Gsw were purchased in a 4 µg scale from Genscript (New Jersey, USA). All DNA template sequences contain a T7 promotor sequence, a 5'-hammerhead ribozyme, and the desired RNA sequence and were cloned into pUC57 by EcoRI and PstI. The sequence for I-A dGsw and III-B Gsw starts at the predicted transcription start site according to R. Breaker (personal communication). The transcription start site for I-B dGsw was predicted by sequence alignment to I-A dGsw and III-B Gsw riboswitches based on -35 and -10 promotor sequences. The template sequence for I-A dGsw ends at the termination site. Template sequences for III-B Gsw and I-B dGsw end at the translation start codon. The plasmid DNA for dGsw^{C74U} was obtained by site-directed mutagenesis of the I-A dGsw plasmid DNA. Site-directed mutagenesis was performed using the QuickChange^{TM}-Kit from Agilent Technologies (Karlsruhe, Germany). Details on sequences and construct design can be found in Appendix A1. To amplify plasmid DNA for PCR, plasmids were transformed into competent DH5α *E.coli* cells (Invitrogen, Karlsruhe). 50 ng of plasmid DNA was added to 20 µL *E.coli* cells. Cells were left on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and cooled on ice. The cells were plated onto 100 µg/mL ampicillin agar plates. After overnight incubation at 37 °C, a single colony was transferred to 50 mL of LB medium with an ampicillin concentration of 100 µg/mL. Cell cultures were incubated at 37 °C and 160 rpm overnight and harvested by centrifugation at 5000 g for 30 minutes at 4 °C. Plasmids were purified by alkaline lysis following the protocol for Midi-kit or Mini-kit preparation provided by Qiagen (Hilden, Germany).

As template for PCR reactions, amplified plasmid DNA was used. For constructs that were transcribed with a 5'-hammerhead ribozyme, the standard T7 promotor primer (20 nt) was used as forward primer. The length of reverse primers was adjusted to match the melting temperature of the forward primer (54.3 °C). For all constructs that were transcribed without a hammerhead ribozyme (for rapid purification or transcription of fragments within the riboswitches), the T7 promotor primer was extended by a sequence segment of the RNA including two additional G's for transcription initiation to yield a total melting temperature of 74.8 °C. Exact sequences for all primers can be found in Appendix A2. Primers were purchased from Eurofins MWG Operon (Ebersberg, Germany). PCR reactions were

performed according to the standard protocol from New England Biolabs (NEB) with 10-fold increased amount of primers (1-3 µM of each primer, 0.2 ng/µL DNA template, 200 µM dNTPs) using homemade Phusion® High-Fidelity DNA Polymerase. The thermocycling program used for PCR reactions is listed in Table 1. Annealing temperature and cycles were optimized to obtain the highest yield combined with high purity of PCR products for *mfl* riboswitch DNA templates. The extension temperature resembles the optimal temperature for Phusion® High-Fidelity DNA Polymerase according to NEB.

Table 1:Thermocycling program used for PCR reactions of *mfl* riboswitch transcripts.

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	2 min	1
Denaturation	98 °C	30 sec	51
Annealing	54 °C	30 sec	
Extension	72 °C	1 min	
Hold	4 °C	-	1

2.1.2 RNA preparation

In vitro transcription reactions were performed from double stranded DNA templates obtained from PCR. Transcription reactions were performed in 200 mM of buffer (Tris-glutamic acid, pH 8.3), 20 mM of DTT, 2 mM of spermidine, 0.01% of Triton X-100, 15 ng/mL T7 polymerase (P266L, homemade)[132] in a volume 5-20 mL. Both $Mg(OAc)_2$ and NTPs were optimized to a final concentration of 10 mM each in 25-50 µL transcription assays. DMSO was optimized to 20% to avoid 3'-end inhomogeneities from run-off transcriptions.[8] Transcription reactions with unlabeled NTPs were performed with 0.2 U/mL and transcription reactions using isotope labeled NTPs with 0.5 U/mL of yeast inorganic pyrophosphatase (NEB, Frankfurt), respectively. All RNA constructs prepared by *in vitro* transcription are listed in Appendix A2.

Isotope labeled NTPs for *in vitro* transcription reactions and ^{15}N,^{13}C-labeled 2'-deoxyguanosine were purchased from Silantes (Munich). 2-aminopurine deoxyriboside was purchased from Berry & Associates (Dexter, Michigan, USA). RNA constructs prepared by solid-phase synthesis were obtained from Dharmacon (Boulder, Colorado, USA) (Appendix A3) and deprotected according to the protocol provided by the supplier.

2.1.3 DNA splinted ligation

Photocaged dGsw^{121} was prepared by splinted enzymatic ligation using T4 RNA ligase 2. T4 RNA ligase 2 catalyzes the ATP-dependent joining of fragments containing a 5'-monophosphate with a 3'-OH with a preference for RNA/RNA and RNA/DNA duplexes.[133] While RNA ligase 2 is prone to ligating monophosphorylated single stranded segments in a side reaction, T4 DNA ligase is more selective but exhibits an overall lower turnover and is therefore less suitable for NMR-scale ligation reactions. ^{15}N G/U labeled dGsw^{85} was prepared by *in vitro* transcription as described in section 2.1.1-2.1.4 and either extensively washed with water or purified by denaturing PAGE. Photocaged dGsw$^{86\text{-}121}$ was synthesized by solid-phase synthesis (Dean Klötzner, AK Heckel, Goethe University Frankfurt). The 31 nt long DNA splint (splint^{31}) was purchased from Eurofins MWG Operon (Ebersberg, Germany) in a 5 µmol scale (HPSF purity).

Table 2: Individual fragments used for splinted enzymatic ligation with T4 RNA ligase 2. ^{c}U correspond to NPE-photocaged Uridine residues and p to monophosphorylated 5'-ends.

Fragment	Sequence
^{15}N G/U dGsw^{85}	5'-AAU GAA UAU AAA AGA AAC UUA UAC AGG GUA GCA UAA UGG GCU ACU GAC CC CGC CUU CAA ACC UAU UUG GAG ACU AUA AGU GAA A-3'
caged dGsw$^{86\text{-}121}$	5'-pAAC CAC UCU UUA AUU AUU AAA G(^{C}U)U (^{C}U)CU (^{C}U)UU UAU GUC-3'
Splint^{31}	5'-AAT AAT TAA AGA GTG GTT TTT CAC TTA TAG T-3'

Ligation reactions were optimized in a 25 µL scale with respect to ATP, splint^{31} and ligase concentrations, and the ratio between splint^{31}, dGsw^{85} and dGsw$^{86\text{-}121}$.[134] The large scale ligation was performed in 120 x 1 mL scales containing 12.5 µM dGsw^{85}, 6.25 µM caged dGsw$^{86\text{-}121}$, 10 µM splint^{31}, 48 µg/mL T4 RNA ligase 2 (homemade), 10 mM $MgCl_2$, 2 mM ATP, 25 mM DTT in 50 mM Tris-Cl, pH 7.4. Ligation reactions were incubated at 37 °C for 1-2 h and 16 h at room temperature. Following ligation, the mixture was incubated with TURBO DNAse (0.02 U/µL) (Thermo Fisher Scientific) overnight at room temperature. Ligations reactions were purified by HPLC, lyophilized and exchanged into NMR buffer (50 mM KCl, 25 mM K_2HPO_4/KH_2PO_4, pH 6.2) using centrifugal concentrators (Vivaspin 20, MWCO 5000).

2.1.4 Gel electrophoresis

Agarose gel electrophoresis

The integrity of PCR products and plasmid DNA was analyzed by agarose gel electrophoresis. DNA samples were loaded in a volume of 10 µL containing 5 µL loading buffer (0.1% bromphenol blue, 0.1% xylene cyanol in 70% glycerine). Plasmid DNA was analyzed with 0.8% agarose gels and PCR products with 1.5-2% agarose gels. Separation of DNA products was achieved in 1x TAE buffer (40 mM Tris/acetate pH 7.5, 1 mM EDTA) at 120 V. Lengths of dsDNA fragments were assigned by comparison to the 2-log DNA marker from New England Biolabs (Frankfurt, Germany).

Denaturing polyacrylamide gel electrophoresis

The lengths of transcription products were analyzed by denaturing polyacrylamide gel electrophoresis. Denaturing gels were composed of 10-20% acrylamide (29:1 acrylamide:bisacrylamide), 7 M urea in 1x TBE buffer (90 mM tris/boric acid, 1 mM EDTA). The polymerization was initialized by adding 0.1% of APS and 0.1% of TEMED. RNA was loaded in a volume of 10-20 µL with 50% of loading buffer (99% formamide, 0.1% bromphenol blue, 0.1% xylene cyanol). High concentrations of urea in the gel and formamide in the loading buffer guarantee denaturing of the RNA to achieve RNA separation based exclusively on the RNA length. Gels were run for 30-40 minutes at 220 V.

Native gel electrophoresis

To investigate the fold of RNA constructs, including the formations of dimers, polyacrylamide gel electrophoresis was performed under native conditions. Native gels were composed of 10-20% acrylamide (29:1 acrylamide:bisacrylamide) in TA buffer (50 mM Tris/acetate, 100 mM NaOAc, pH 8.0). RNA was loaded in a volume of 10 µL with 5 µL of loading buffer (70% glycerine, 0.1% bromphenol blue, 0.1% xylene cyanol). Gels were run for 3-4 h at 40 V at 4 °C.

Visualization

Agarose gels were stained with GelRed™ and acrylamide gels were either visualized by UV shadowing or staining with GelRed™. GelRed is a sensitive fluorescent nucleic acid dye with strongly enhanced fluorescence properties when intercalated into RNA or DNA.

Stained Gels were visualized with a Gel iX Imager (Intas). UV shadowing was performed on UV fluorescent TLC plates that were excited at 254 nm.

2.1.5 RNA purification

RNA was purified by either PAGE, RP-HPLC or by extensive washing in centrifugal concentrators.

For HPLC purification of *in vitro* transcription reactions, the majority of non-incorporated nucleotides and enzymes were first removed by anion exchange column chromatography using DEAE-Sepharose (GE Healthcare, Uppsala, Sweden). DEAE-Sepharose (suspension in 20% ethanol) was equilibrated with 0.1 M NaOAc (pH 5.5). The RNA was eluted by applying a stepwise gradient of NaOAc (0.6 - 3 M). The content of RNA in collected fractions was analyzed by either UV-spectroscopy (section 2.2.1) or gel electrophoresis (section 2.1.4). Fractions containing RNA were either precipitated by dilution to 0.6 M NaOAc followed by precipitation with 4 volumes of ethanol (stored at -80 °C) or by concentration in centrifugal concentration. Subsequent HPLC purification was performed by Elke Stirnal (AK Schwalbe, Goethe University, Frankfurt) by applying samples to a PerfectSil RP18 column with a size of 10x250mm (preparative scale) or 4.6x250mm (analytical scale). RNA was eluted at 60 °C by applying an acetonitrile gradient (0-60%) in phosphate buffer (50 mM potassium phosphate pH 5.9, 2 mM tetrabutylammonium hydrogen sulfate). For removal of acetonitrile, RNA containing fractions were lyophilized. High salt concentrations were removed by submitting the dissolved RNA to centrifugal concentrators (Vivaspin, MWCO 2000-5000). Tetra-butyl-ammonium hydrogen ions bound to the backbone of the RNA were replaced by Li^+ by precipitation with five volumes of 2% (w/w) $LiClO_4$ in acetone. Li^+ ions were finally replaced by K^+ by buffer exchange in centrifugal concentrators using NMR buffer (50 mM KCl, 25 mM K_2HPO_4/KH_2PO_4, pH 6.2).

Ligation reactions were directly applied to a Kromasil RP18 10x250mm column without further concentration and eluted at room temperature by applying an acetonitrile gradient starting from trietylammonium acetate buffer (0.1 M triethylammonium acetate, pH 6.0). HPLC fractions were lyophilized and exchanged into NMR buffer.

For PAGE purification, transcription reactions were concentrated to ~1 mL by centrifugal filter units (Vivaspin 20, MWCO 2000-5000). PAGE purification was performed using

10% polyacrylamide gels (29:1 (w/w) acrylamide/bisacrylamide, 7M urea) in 1x TBE buffer. For transcription reactions containing ribozyme fragments, up to 600 OD were loaded on the gel and for transcription reactions yielding a single RNA fragment, up to 1600 OD were loaded on the gel in at least 30% of formamide. The gel was run for 30 minutes at 25 W to allow migration of the RNA into the gel matrix and subsequently for 4-6 hours at 50 W under cooling. The RNA was visualized by UV shadowing at 253 nm, excised from the gel and passively eluted in 0.6 M NaOAc (pH 5.5) at room temperature overnight. Eluted RNA was precipitated with five volumes of -80 °C ethanol and then twice with five volumes of 2% (w/v) $LiClO_4$ in acetone. The RNA was exchanged into NMR buffer using centrifugal filter units (Vivaspin 20, MWCO 2000-5000).
Extensive washing for rapid RNA screening was performed by applying the transcription mixture directly to centrifugal concentrators, washing the solution with 60 mL of water followed by buffer exchange into NMR buffer with an additional 40-60 mL of NMR buffer as described previously.[8]

RNA purified under denaturing conditions was folded by thermal denaturation of the RNA at high concentrations (0.2 - 0.5 mM) followed by tenfold dilution with water (0 °C) and incubation on ice for 1 h. Folded RNA was exchanged into NMR buffer and analyzed by native gel electrophoresis. For RNA purified by extensive washing under native conditions, no refolding protocol was applied prior to measurements.

2.1.6 Transcription termination assays

Multi-round transcription termination assays were performed with *E.coli* RNA polymerase and dsDNA obtained from PCR. The PCR reaction was performed from a circular plasmid containing the T7A1 promotor and the I-A dGsw sequence extended to 235 nt containing the start codon including 46 nucleotides of the coding sequence for ribonucleotide reductase subunit alpha (Appendix A1). PCR reactions were performed as described in section 2.1.1. Transcription reactions were performed with 5 vol-% PCR product, 0.1 U/µL *E.coli* RNA polymerase holoenzyme (NEB, Frankfurt), 150 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100, 500 µM of each NTP, 5 µCi [γ-^{32}P]UTP, 1 mM DTT in 40 mM Tris·HCl (pH 7.5) in presence and absence of 200 µM of 2'dG. Reactions were incubated for 2 hours at 37 °C and analyzed by 10% PAGE. Gels were visualized using a PhosphorImager and analyzed with the software ImageQuaNT.

2.2. Biophysical

2.2.1 UV spectroscopy

UV-spectroscopy is commonly applied to determine the concentration of DNA, RNA or nucleotides. Concentrations can be calculated based on the Beer-Lambert equation: $A = \varepsilon * c * d$, where A is the optical density, ε the wavelength dependent molar extinction coefficient in $Lmol^{-1}cm^{-1}$, c the concentration in $molL^{-1}$ and d the path length in cm. For double stranded DNA, the extinction coefficient is commonly approximated to 50 $ng*cm*\mu L^{-1}$. For RNA, the molar extinction coefficient was calculated by summarizing the molar extinction coefficient of individual bases. The following extinction coefficients were used for the individual bases and ligands:

Guanosine:	10400 $l*mol^{-1}*cm^{-1}$	(260 nm)
Cytidine:	6500 $l*mol^{-1}*cm^{-1}$	(260 nm)
Adenosine:	10400 $l*mol^{-1}*cm^{-1}$	(260 nm)
Uridine:	9100 $l*mol^{-1}*cm^{-1}$	(260 nm)
2'-deoxyguanosine:	13300 $l*mol^{-1}*cm^{-1}$	(254 nm)
2-aAminopurine 2'-deoxyriboside:	3600 $l*mol^{-1}*cm^{-1}$	(303 nm)

Final extinction coefficients used for all synthesized RNA constructs are listed in Appendix A2. The absorption of nucleic acids was measured on either a Varian UV/vis-spectrometer Cary 50, a NanoDrop UV/vis spectrometer ND-1000 (PEQLAB) or a NanoDrop One UV/vis spectrometer (Thermo Scientific). For measurements with the UV/vis-spectrometer Cary 50, quartz cuvettes with a path length of 1 cm and a volume of 1 mL (0.1 – 0.9 OD/mL) were used. The NanoDrop spectrophotometers use a path length of 0.2-1 mm, normalized to 1 cm path length, allowing measurements of 50 times more concentrated samples to obtain nucleic acid concentrations according to the Beer-Lambert law. Samples were applied in a 1-2 µL volume without further dilution.

2.2.2 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) allows direct assessment of heat generated or absorbed by interacting molecules. The difference in heat is measured by the power required to maintain the temperature difference between a reference cell and a sample cell

close to zero. ITC measurements yield a complete thermodynamic profile of biomolecular interactions in the mM to nM range including dissociation constants (K_D), the reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS). ITC measurements were performed on a MicroCal^{TM} VP-ITC (Northampton, MA, USA). RNA samples were added to the ITC cell in a volume of 1.4 mL in concentrations of 10-100 µM, while the ligand 2'dG was provided in the syringe in a volume of 500 µL in 20x concentration excess compared to RNA. Both components were provided in NMR buffer (50 mM KCl, 25 mM K_2HPO_4/KH_2PO_4, pH 6.2) and 3 mM $MgCl_2$. The ligand was injected in a 1-3 µL volume spaced by 120-240 seconds for equilibration with a total of 30-50 injections to reach saturation. The baseline for released heat from an injection was recorded and subtracted by titrating ligand into NMR buffer. Heat amplitudes of each injection were integrated and evaluated using the Origin ITC software (OriginLab, Northampton, MA, USA). The program uses the following iterative model to fit the change in released heat after each injection. The total heat content Q for each injection for a single set of identical binding sites is expressed as follows:

$$Q = \frac{nM_t \Delta H V_0}{2}\left[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{\left(1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t}\right)^2 - \frac{4X_t}{nM_t}}\right] \quad (1)$$

where n the number of binding sites, M_t is the concentration of the macromolecule in the cell, and X_t the concentration of the ligand in the syringe. M_t and X_t are provided in the fit. Q is first calculated by initial guessing of n, K and ΔH for the i^{th} injection. Then, the change in heat ΔQ(i) from injection i-1 to i is calculated according to:

$$\Delta Q(i) = Q(i) + \frac{dV_i}{V_0}\left[\frac{Q(i)+Q(i-1)}{2}\right] - Q(i-1) \quad (2)$$

where V_0 is the actual cell volume and dV_i is the change in volume after the injection. Calculated changes in heat content ΔQ(i) for each injection are compared to experimental values. Initial values of n, K and ΔH are improved by standard Marquardt methods, and the process is iterated until further iteration does not significantly improve the fit.

2.2.3 Stopped-flow spectroscopy

Stopped-flow spectroscopy allows monitoring of folding events in the millisecond to second timescale. Biomolecular interactions, such as ligand binding, can be monitored by rapid mixing followed by detection using diverse spectroscopic methods. Most commonly,

changes in UV absorption, fluorescence or CD caused by environmental changes are monitored.

Stopped-flow measurements to monitor the event of ligand binding of the fluorescent ligand 2-aminopurine 2'-deoxyriboside were performed using a π*-180 CDF-spectrometer (Applied Photophysics). Folding was initiated by mixing I-A $dGsw^{C74U}$ transcripts (2 µM) with 2-aminopurine 2'-deoxyriboside (2 µM – 512 µM) in a 1:1 volume ratio at 25 °C. Both RNA and ligand were provided in NMR buffer (50 mM KCl, 25 mM K_2HPO_4/KH_2PO_4, pH 6.2) and 3 mM $MgCl_2$. Fluorescence was monitored after a dead time of 1.8 ms over a course of 20-60 seconds. 2-aminopurine 2'-deoxyriboside was excited at 304.5 nm with a Xe lamp and monitored at >340 nm (E_{max}: 271 nm) using a wavelength cut-off filter. The fluorescence signal was detected using a cell path length of 2 mm over a bandwidth of 4 nm. Kinetic measurements were averaged over 10-15 single measurements. Kinetic traces were analyzed with the program sigma plot 12.5 (Systat).

Exponential fluorescence decays were fitted with single and/or double exponential functions: $y(t) = y_0 + a * e^{kt}$ and $y(t) = y_0 + a * e^{k_1 t} + c * e^{k_2 t}$, where y(t) is the time dependent fluorescence signal, y_0 the fluorescence in the ligand-bound complex, and c the fluorescence amplitude of free 2-aminopurine 2'-deoxyriboside, k_1 and k_2 the apparent rate constants and t the time.

Association and dissociation constants k_{on} and k_{off} were determined by recording fluorescence decays at various ligand concentration (2 µM – 512 µM) while keeping the RNA concentration constant. The ligand was applied in excess (> 16x) to keep the concentration of the free ligand during the course of the kinetics constant by approximation (pseudo-first-order conditions). Association constants k_{on} and dissociation constants k_{off} were determined by plotting determined apparent rate constants k_{app} against the ligand concentration [2'dAP] within the range of linear dependence according to: $k_{app} = k_{on} * [2'dAP] + k_{off}$

The equilibrium binding constant K_D was calculated according to: $K_D = \frac{k_{off}}{k_{on}}$

2.2.4 Simulations of co-transcriptional folding

Rates of refolding between aptamer and antiterminator conformations were calculated based on kinetic traces measured by real-time NMR for three bistable RNAs at 298 K. These experiments have revealed an exponential dependence between the number of base pairs that need to dissociate and the life-time of adopted structures in the interconformational equilibrium (Figure 7). Based on the linear fit of logarithmic rate constants versus the number of base pairs, rates for interconversion were approximated.

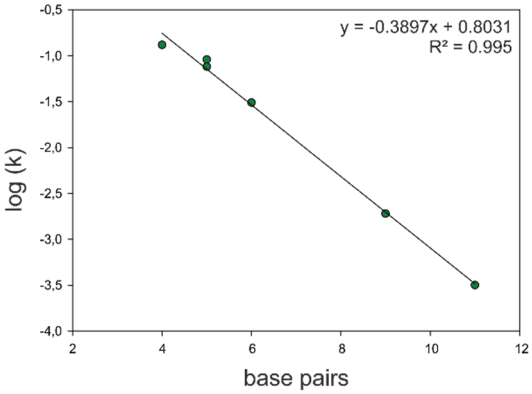


Figure 7: Logarithmic rate constants versus the number of dissociated base pairs including linear fit as derived previously.[135]

Kinetic Markovian simulations of co-transcriptional folding were performed with Wolfram Mathematica 8. Kinetics between the states were described by first order rate equations with a single rate constant k_{ij} for the inter-state transition i to j. The transition matrix $\mathbf{k}_{ij}$ assembled from rate constants was provided as input to describe statistical dynamics between Markov states (Appendix A4).

2.2.5 NMR-spectroscopy

Strategy for RNA secondary structure assignment

NMR-spectroscopy represents a powerful tool to obtain high-resolution structures of nucleic acids**.** Currently, the increase in spectral overlap combined with a continuous increase in line broadening with growing size of the macromolecule limits high resolution structures to a size of about ~50-70 nucleotides, depending on fold and homogeneity of the macromolecule. Particularly in case of RNA, NMR signals exhibit limited chemical shift dispersion due to the existence of only four structurally related ribonucleotide units along with the A-helix being the predominant structure element in double stranded regions.[136,137] Despite these limitations in size to obtain high-resolution structures, the application of NMR-spectroscopy extends beyond structure calculations to the study of complexes and folding dynamics.[138] The imino protons H1 of guanine and H3 of uracil are unique reporters

of RNA secondary structure. These signals can only be detected provided they are protected from solvent exchange in base pairing. Imino protons resonate between 10 and 15 ppm, substantially dispersed from aromatic protons and sugar protons, and facilitate determination of secondary structure, monitoring of conformational changes and binding interactions for nucleic acids >100 nt (Figure 8). In canonical Watson-Crick base pairs, imino protons resonate between 12-15 ppm and are shifted upfield to 10-12 ppm in U-U and G-U wobble base pairs. ^{1}H-^{15}N correlation spectra allow further differentiation between guanine H1 and uracil H3 protons due to a ~10 ppm chemical shift dispersion between N1 of guanine and N3 of uracil.[139]

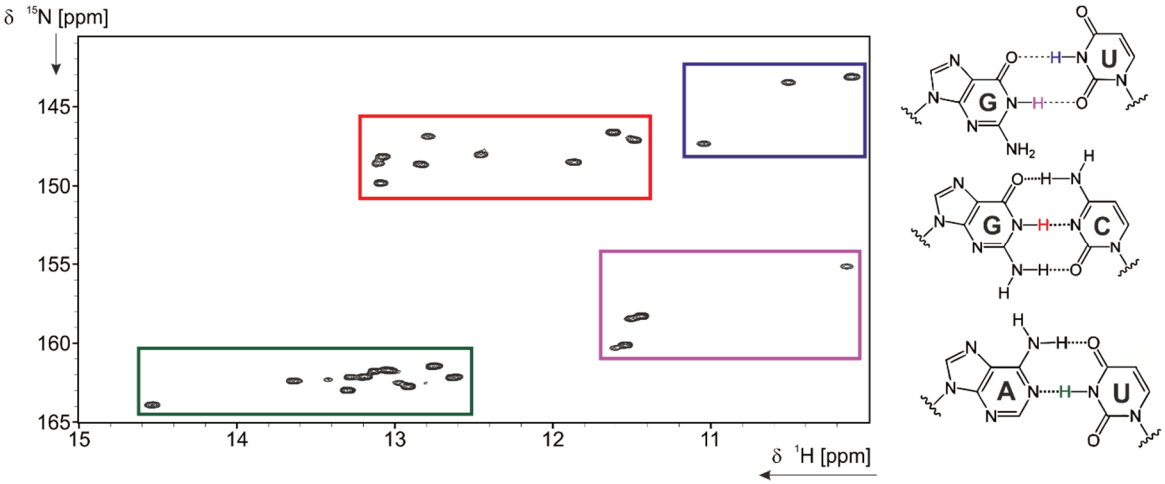


Figure 8: Exemplary ^{1}H-^{15}N correlation spectrum illustrating characteristic chemical shifts for imino protons in A-U (green), G-C (red) and G-U wobble (blue, magenta) base pairs.

The assignment of dGsw^{FL}, dGsw^{ON} including its individual domains was facilitated by combining ^{1}H,^{1}H-NOESY, ^{1}H,^{15}N-HSQC/^{1}H,^{15}N-TROSY, with ^{15}N/^{14}N-X-filtered ^{1}H,^{1}H-NOESY and HNN-COSY experiments (see below). The imino protons of the 144 nt dGsw^{FL} were assigned in a 'divine-and-conquer' approach. The relatively large size of the RNA causes broadening of many cross peaks in NOESY spectra beyond detection as a result of decreased T2 relaxation times. Therefore, dGsw^{FL} was dissected into its two individual domains aptamer and terminator hairpin and the assignment of the individual domains was aligned to imino proton signals of the full-length construct. The antiterminator helix of the 122nt dGsw^{ON} construct could be identified due to its partial identity with the terminator helix. For the more detailed examination of structural heterogeneity in dGsw^{ON}, dGsw^{ON} was assembled from fragments with different ^{15}N labeling patterns (see chapter

III) and the final secondary structure assignment was facilitated by using a combination of different $^{15}N/^{14}N$-X-filtered NOESY spectra.

$^{1}H,^{1}H$ NOESY

The $^{1}H,^{1}H$ NOESY (Nuclear Overhauser Exchange Spectroscopy) experiment is based on the nuclear Overhauser effect (NOE) between ^{1}H-spins. The NOE is a magnetization transfer by through space dipolar coupling that allows correlation of protons within a distance of 5 Å. The effect correlates with r^{-6}, where r is the distance between the two atoms.[140] In NOESY spectra, cross peaks from one imino proton to imino protons of adjacent base pairs, both inter- and intercatenar, can be observed in helical regions.[136,141,142] $^{1}H,^{1}H$-NOESY spectra were recorded using a jump return echo water suppression scheme.[143] The proton carrier frequency was set to the resonance frequency of the solvent (4.7 ppm) in the direct dimension and to 8.5-9 ppm in the indirect dimension. Spectra were recorded with a spectral width of 24 ppm in the direct dimension and 12 ppm in the indirect dimension. For RNA constructs larger than 100 nucleotides, spectra were recorded with a mixing time of 70-100 ms and for smaller RNA constructs (30-70 nt) with a mixing time of 100-150 ms. Spectra were recorded with an acquisition time of 70-100 ms representing 1024-3072 points in the ^{1}H dimension. In the indirect dimension, 348-800 points were recorded depending on RNA size and spectrometer frequency. The relaxation delay between experiments was set to 1.2-1.8 s and depending on the RNA concentration, 32-256 scans were recorded.

ω1,ω2-X-filtered NOESY

X-filtered NOESY spectra were recorded to assign $dGsw^{ON}$ by filtering ^{1}H protons attached to $^{15}N/^{14}N$ or ^{13}C. X-filter techniques exploit the evolution of $^{1}J_{XH}$ coupling during a spin echo period, which allows selective rotation of one of the two magnetizations back to the z-axis.[144] X-filtered experiments were recorded as described above using a jump return echo scheme.

For $^{15}N/^{14}N$-X-filtered experiments, an X-filter scheme was implemented in the direct dimension (ω2). ^{1}H-^{15}N or ^{1}H-^{14}N components of magnetization were selected by variations in the phase cycle. For the assignment of adenine H2 in $dGsw^{ON}$ (^{13}C-A,^{15}N-full), a NOESY experiment was recorded implementing a ^{13}C X-filter in the indirect dimension

(ω1), where only cross peaks of imino protons to adenine H2 and diagonal peaks of adenine H2 and H8 are detected. X-filter pulse programs are provided in Appendix A5.

^{1}H,^{15}N-HSQC, SF-[^{1}H,^{15}N]-HMQC, BEST-[^{1}H,^{15}N]-TROSY

Guanine H1 and Uracil H3 signals in NOESY spectra were resolved by recording ^{15}N-HSQC (Heteronuclear Single Quantum Coherence), ^{15}N-HMQC (Heteronuclear Multiple Quantum Coherence) or ^{15}N-TROSY (Transverse Relaxation Optimized Spectroscopy) spectra. These ^{1}H,^{15}N correlation spectra yield an imprint of RNA secondary structure by resolving most imino protons even in larger RNA molecules.

HSQC experiments were recorded to obtain ^{1}H-^{15}N correlation spectra of ^{15}N isotope labeled RNA constructs <70 nt. ^{1}H,^{15}N-HSQC experiments correlate protons bound to ^{15}N by transferring magnetization between ^{1}H and ^{15}N through $^{1}J_{NH}$ coupling using INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) schemes and represent the most common pulse scheme applied to obtain ^{1}H-^{15}N correlation spectra.[145] HSQC spectra were recorded using a soft WaterGATE suppression scheme.[146] Experiments were recorded with relaxation delays of 1s.

SF-HMQC experiments were recorded to obtain ^{1}H-^{15}N correlation spectra of unlabeled RNA constructs <50 nt. SOFAST (band Selective Optimized Flip-Angle Short Transient) experiments exploits HMQC-type coherence transfer, which requires lower radio frequency pulses compared to HSQC experiments.[147,148] Further, HMQC experiments enable Ernst-angle excitation,[149] which is a pulse flip angle that yields maximum signal intensity in the least amount of time. The use of band-selective ^{1}H pulses further shortens interscan delays significantly, since ^{1}H T_1 relaxation of excited nuclei proceeds much faster if the majority of nuclei within the sample, particularly the solvent, are maintained at equilibrium.[150] The combination of these features allow very short interscan delays, perfectly suited to acquire a large amount of scans required for samples that contain ^{15}N only in natural abundance (0.4%). Natural abundance ^{15}N SF-HMQC experiments were recorded with >1024 scans and an interscan delay of 0.3 s.

BEST-TROSY experiments were recorded to obtain ^{1}H,^{15}N correlation spectra of ^{15}N isotope labeled RNA constructs larger than ~70 nt. TROSY experiments are designed to

exploit constructive or destructive interactions of dipole-dipole (DD) induced T_2 relaxation with chemical shift anisotropy (CSA) induced T_2 relaxation in non-decoupled $^{1}H,^{15}N$ correlation spectra.[151] The experiment selects the component of the $^{1}J_{NH}$ coupling multiplet with the smallest line width caused by partial cancellation of CSA and DD relaxation. The resulting signals are shifted by -47 Hz in the direct dimension and +47 Hz in the indirect dimension compared to broadband decoupled HSQC spectra. Due to relatively large size and corresponding long T_2 times of the RNAs investigated in this thesis, TROSY experiments contribute significantly in reducing the line width and increasing the resolution. TROSY spectra were recorded in an extended SOFAST manner by implementing Band Selective Short Transient (BEST) pulses.[152–154] Spectra were recorded with an interscan delay of 0.3-0.5 s.

All ^{15}N-correlation experiments were recorded with a nitrogen carrier frequency of 153 ppm, resembling the intermediate resonance frequency between guanine N1 and uracil N3. Spectra were recorded with a spectral width of 24 ppm with 1024-3072 points in the direct dimension and 25 ppm with 128-256 points in the indirect dimension. TROSY experiments were recorded with up to 512 points at high spectrometer frequencies.

HNN-COSY

HNN-COSY spectra allow monitoring of a ^{2h}J-coupling over a hydrogen bond N-H···N through a sequential $^{1}J_{HN}$ INEPT magnetization transfer followed by a $^{2}J_{NN}$ COSY magnetization transfer.[155,156] This hydrogen bonding pattern is present in all Watson-Crick base pairs and facilitates direct observation of base pairing interactions in RNA. The HNN-COSY spectrum for dGsw^{ON} was recorded with a nitrogen carrier frequency of 185 ppm as intermediate frequency of acceptor (adenosine/cytidine) and donor (guanosine/uridine) bases. The spectrum was recorded with a spectral width of 24 ppm and 2048 points in the direct dimension and 100 ppm and 256 points in the indirect dimension. The delays for HN transfer were shortened to 2.3 s (J_{NH} ~110 Hz) and for the NN-transfer to 10.5 ms (J_{NN} ~24 Hz) to reduce T_2 relaxation during a single experiment. 1024 scans of the experiment were recorded with an interscan delay of 1.1 s using a soft WaterGATE suppression scheme.[146]

Screening of ligand binding

1 mM of ^{15}N-labeled 2'dG was added to ~500 µM samples obtained from buffer exchange of 2.5 mL transcriptions. ^{1}H-^{15}N-filtered experiments were recorded to monitor the ligand imino proton signal integral at varying Mg^{2+} concentrations (2 mM, 3 mM, 4 mM and 5 mM).

Light-induced real-time NMR experiments

Time resolved NMR experiments were performed in a pseudo 2D experiment using the jump return echo water suppression scheme.[143] Kinetics were triggered by a TLL connection to a laser set up (Paladin Advanced 355-8000) operating at 8 W. Laser irradiation within the NMR tube was achieved by directing the laser beam to a Shigemi tube plunger, modified to achieve conical light distribution within the sample, via a quartz fiber. The laser beam retains ~60-70% of its initial power at the end of quartz fiber. Experiments were optimized to achieve a maximum in time resolution. Laser irradiation was performed for 500 ms in the presence of ligand and for 1 s in the absence of ligand. The signal intensity was maximized by Ernst angle excitation and an interscan delay of 0.3 s.[149] Measurements were averaged over 2 transients (ligand-bound) or 10 transients (ligand-free).

NMR samples and experimental setup

NMR samples were prepared in a concentration ranging from 100 µM to 800 µM of RNA in NMR buffer (50 mM KCl, 25 mM K_2HPO_4/KH_2PO_4, pH 6.2). Measurements were performed in Shigemi susceptibility matched microtubes in a volume of 260-300 µL with 5-10% of D_2O added. All experiments were recorded at 283-298 K. Exact temperatures for each experiment are listed in the respective Figure captions. For ^{15}N- and ^{13}C-broadband decoupling, GARP sequence were used.[157]

NMR experiments were performed on Bruker spectrometers listed in Table 3. Experiments were analyzed using the Bruker Biospin software TopSpin 2.1.-3.5. Assignments were performed using the software Sparky 3.114.[158]

Table 3: Bruker spectrometers including equipped probeheads used for all NMR measurements.

Spectrometer	**probehead**
AVIII 600 HD	Prodigy 5mm TCI $^{1}H,^{15}N,^{13}C$ Z-GRD
AVII 600	5 mm TCI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD
AVIII 600	5 mm TXI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD
AVIII 700 HD	5 mm QCI cryo $^{1}H,^{15}N,^{13}C,^{31}P$ Z-GRD
AV 800	5 mm TXI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD
AVIII 800 HD	5 mm TCI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD
AV 900	5 mm TXI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD
AVIII 950	5 mm TCI cryo $^{1}H,^{15}N,^{13}C$ Z-GRD

Chapter III

NMR Characterization of the I-A 2’dG-sensing Riboswitch

3.1. Structural characterization of the I-A 2'dG-sensing riboswitch

3.1.1 NMR characterization of the full-length riboswitch

Previous studies performed on the I-A type 2'dG-sensing riboswitch have focused on ligand binding, ligand selectivity and the architecture of the aptamer domain investigated by NMR-spectroscopy[126,129] and X-ray crystallography[128] (see Chapter 1.2.4). Experiments performed on the full-length riboswitch including its ligand-dependent allosteric modulation and mechanism of regulation are limited to transcription termination assays performed by Kim *et al.*[1] According to these assays, gene expression of ribonucleotide reductase genes is repressed at high concentrations of 2'dG by formation of a transcriptional terminator. The work presented in this thesis centers around the full-length 2'dG riboswitch from the bacterium *Mesoplasma florum* and aims to decipher the mechanism of action by transcriptional riboswitches based on this model system. The sequence of the full-length riboswitch ($dGsw^{FL}$) was provided by Breaker and coworkers starting from the transcription start site and ending at the termination site (personal communication).

First, $dGsw^{FL}$ was investigated by NMR-spectroscopy to characterize conformations relevant for the regulation of the downstream gene expression. The secondary structure of $dGsw^{FL}$ was characterized in both absence and presence of ligand by assigning the imino protons in ^{1}H-^{1}H-NOESY spectra (Figure 9). Due to the size of the RNA, secondary structure motifs within the riboswitch (aptamer domain and expression platform) were characterized individually. The assignment of individual domains was then transferred to the full-length riboswitch for verification of independent folding. Figure 1B shows an overlay of NOESY spectra of $dGsw^{FL}$ (black) and a model terminator hairpin (40mer, in red) recorded in the absence of ligand at 283 K. In the NOESY spectrum of $dGsw^{FL}$, primarily cross peaks corresponding to the terminator hairpin can be detected. Within the aptamer domain, only the cross peak emerging from the strong NOE transfer between imino protons in the G70-U56 wobble base pair can be observed. The remaining aptamer cross peaks are broadened beyond detection in $dGsw^{FL}$ at 283 K due to its large size (144 nt), and consequently larger correlation time τ_c. The increased signal intensity for terminator signals suggests accelerated motion in comparison to the aptamer domain and implies decoupled motion between the two domains. Increasing the temperature to 298 K leads to a loss in signal intensity for terminator signals and an increase in signal intensity for aptamer domain signals. The increase in the amount of aptamer signals can be explained

by a decrease in the correlation time, while the decrease in the amount of terminator signals is associated with an increase in the imino proton solvent exchange. Changes induced by ligand binding to the aptamer domain of dGsw^{FL} were therefore investigated at 298 K. Figure 9C shows a NOESY spectrum of dGsw^{FL} at 298 K in the presence of 2'dG and Mg^{2+}. In general, ligand binding to dGsw^{FL} can be detected also in the absence of Mg^{2+}, but loop-loop interactions and the formation of a stable binding pocket require the addition of Mg^{2+}. In the presence of Mg^{2+}, ligand binding is significantly enhanced by stabilizing the parallel alignment of helices P2 and P3 and consequently preorganizing the binding pocket towards the ligand bound state. The assignment of the aptamer domain in the presence of ligand was adapted from Wacker *et al.*[126]

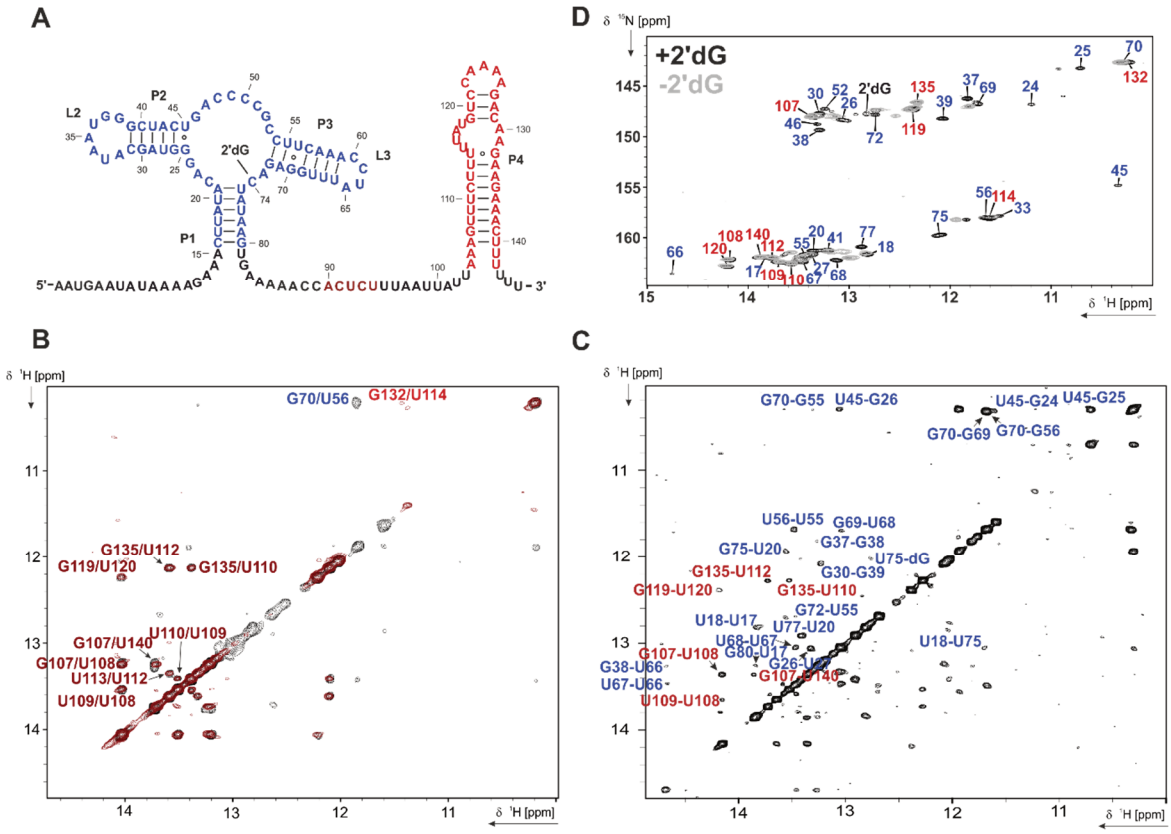


Figure 9: NMR characterization of dGsw^{FL} in presence and absence of ligand. A: Secondary structure of dGsw^{FL} restrained by NMR. B: Overlay of NOESY spectra of dGsw^{FL} in black (450 µM, 800 MHz) and the terminator model helix in red (40mer, 700 µM, 600 MHz). The spectra were recorded at 283 K. C) NOESY spectrum of 400 µM dGsw^{FL}, 2 eq. 2'dG, 12 eq. $MgCl_2$ recorded at 950 MHz and 298 K. D) Overlay of ^{15}N-TROSY spectra of dGsw^{FL} in the absence of ligand (grey, 400 µM RNA, 12 eq. $MgCl_2$) and in the presence of ligand (200 µM RNA, 12 eq. $MgCl_2$) The spectra were recorded at 800 MHz and 298 K. Aptamer signals are color coded in blue and terminator signals are color coded in red.

Comparison of ^{15}N-TROSY spectra (Figure 9D) of dGsw^{FL} recorded in the presence (black) and absence (grey) of ligand reveal the presence of terminator signals (in red) not only in the ligand-bound state but also in the ligand-free riboswitch form. The effects of ligand binding to dGsw^{FL} is limited to the formation of a compact ligand binding pocket including parallel alignment of helices P2 and P3, while the terminator helix remains completely independent of this structural rearrangement. A ligand-induced allosteric conformational switch that would allow transcription to proceed cannot be observed.

The terminator conformation represents the lowest free energy state independent of ligand binding, a feature of kinetically controlled transcriptional riboswitches, which was also previously observed for both the *xpt* guanine-sensing OFF riboswitch and the *pbuE* adenine-sensing ON-riboswitch from *B.subtilis*.[3,106,159,160] As described in Chapter 1.2.2, regulation of transcription by riboswitches inevitably comprises a kinetic component, where both ligand binding and conformational transitions within the expression platform are tightly coupled to the speed of transcription.[2,3,108] As a consequence, dynamic and structural investigations on fully transcribed, thermodynamically equilibrated, transcriptional riboswitches yield limited insight into the mechanism of regulation. The functional state leading to gene expression (ON-state), which cannot be observed in the full-length riboswitch form, represents a transient, metastable state that is adopted during transcription as nascent transcripts are extended by single nucleotides. Gene expression is facilitated by kinetic trapping of this genetic ON-state beyond the time point of regulatory decision. The following section addresses the structural characterization of the conformation that allows gene expression to proceed.

3.1.2 NMR characterization of the genetic ON-state

Typically, in transcriptional OFF-riboswitches, the genetic ON-state involves an alternative base pairing of strand T, that forms the terminator helix TH in the presence of ligand, with a sequence segment in the aptamer domain to form an antiterminator helix (Figure 10). Since the formation of the extremely stable terminator hairpin TH prevents formation of the alternative antitermator helix, truncation of strand H should energetically favor the formation of the antiterminator helix and allow characterization of the genetic ON-state. Mfold folding predictions[112] of this transcriptional intermediate (dGsw^{ON}, 122nt) suggest that the antiterminator helix PT forms with segment P (blue, Figure 10). Further, it is predicted that in the ON-state, the 3'-strand of helix P1 is involved in the formation of an

additional helix P5. To verify mfold folding predictions, the secondary structure of dGsw^{ON} was investigated by NMR-spectroscopy. Extensive analysis of NOESY spectra of dGsw^{ON} at different mixing times (50 - 200 ms) and TROSY spectra yield a preliminary assignment, covering approximately 70% of detected imino protons. Helices P2 and P3 of the ligand-free aptamer domain could be identified to also form in the ON-state. In addition, the antiterminator helix PT could be identified due to its partial identity with the terminator helix TH, which cannot form in the transcriptional intermediate dGsw^{ON} due to the absence of strand H. However, helix P5 as predicted by mfold, could not be identified.

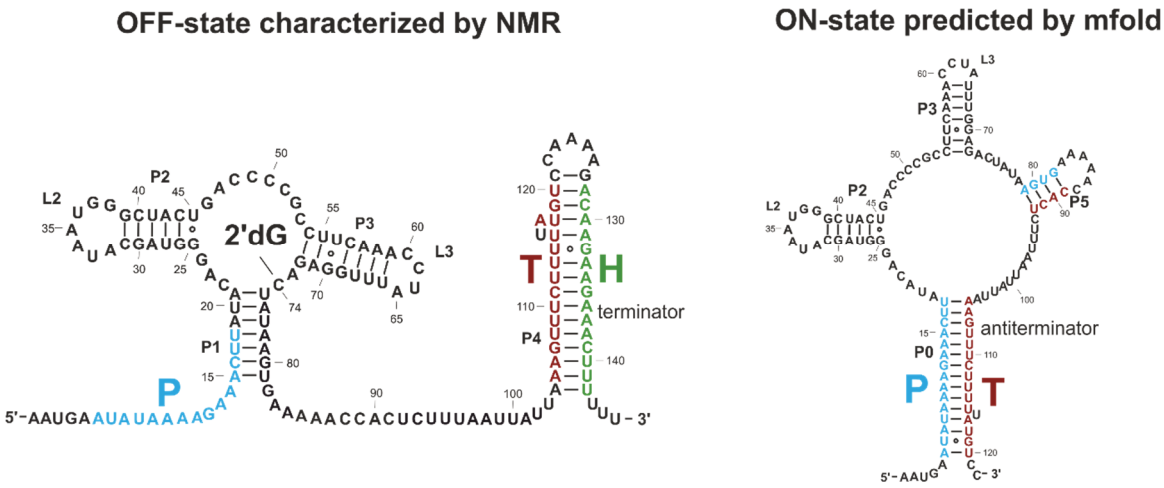
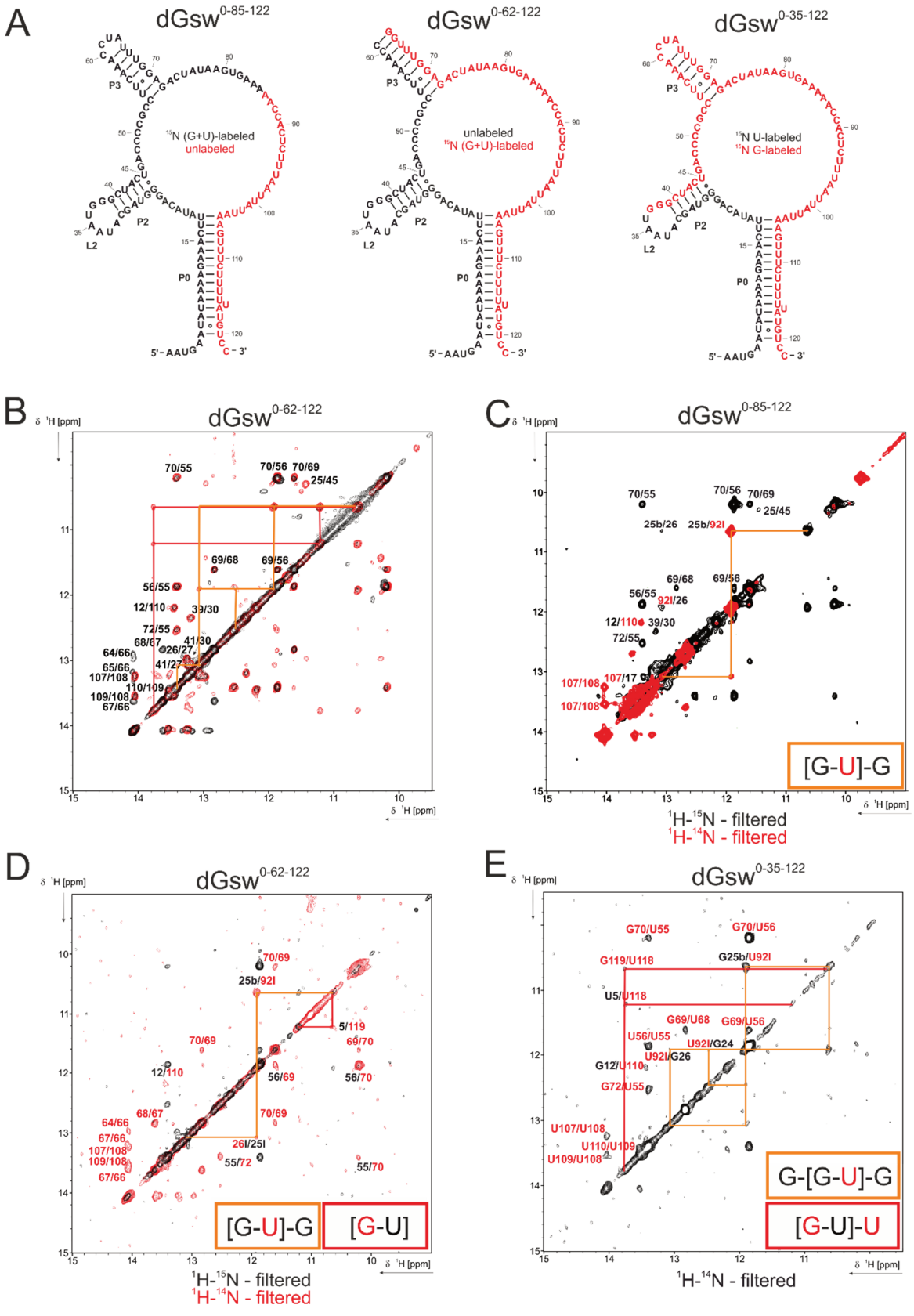


Figure 10: Predicted allosteric conformational modulation of the I-A 2'dG responsive riboswitch according to mfold.[112] In the presence of ligand, the switching strand T (in red) anneals with the terminator strand H (in green) to form terminator helix TH. In the absence of ligand, the aptamer stabilizing strand P (in blue) is predicted to anneal with the terminator strand H (in green) to form the antiterminator helix PT. The predictions also hypothesize the formation of an additional helical segment P5 in the genetic ON-state.

To complete the missing assignment, dGsw^{ON} was assembled from fragments with different ^{15}N-labeling patterns without further ligation. In total, three constructs were prepared: dGsw$^{0\text{-}85\text{-}122}$, dGsw$^{0\text{-}62\text{-}122}$ and dGsw$^{0\text{-}35\text{-}122}$ (Figure 11A). The structural integrity of the individual constructs was examined by comparing NOESY spectra of assembled dGsw^{ON} with intact dGsw^{ON}. An overlay of NOESY spectra for the structurally most divergent construct dGsw$^{0\text{-}62\text{-}122}$ with dGsw^{ON} is shown in Figure 11B. dGsw$^{0\text{-}62\text{-}122}$ appears to be destabilized compared to dGsw^{ON}, leading to reduced signal intensity of signals corresponding to unassigned NOESY pathways (red and orange pathways). However, cross peaks corresponding to these pathways can still be detected and assigned partially (Figure 11D). Additional signals appearing in the NOESY spectrum of dGsw$^{0\text{-}62\text{-}122}$ can be assigned to the artificially implemented G-C base pairs in helix P3. For construct dGsw$^{0\text{-}85\text{-}122}$, the

antiterminator helix is destabilized and signals within the red assignment pathway are broadened beyond detection.



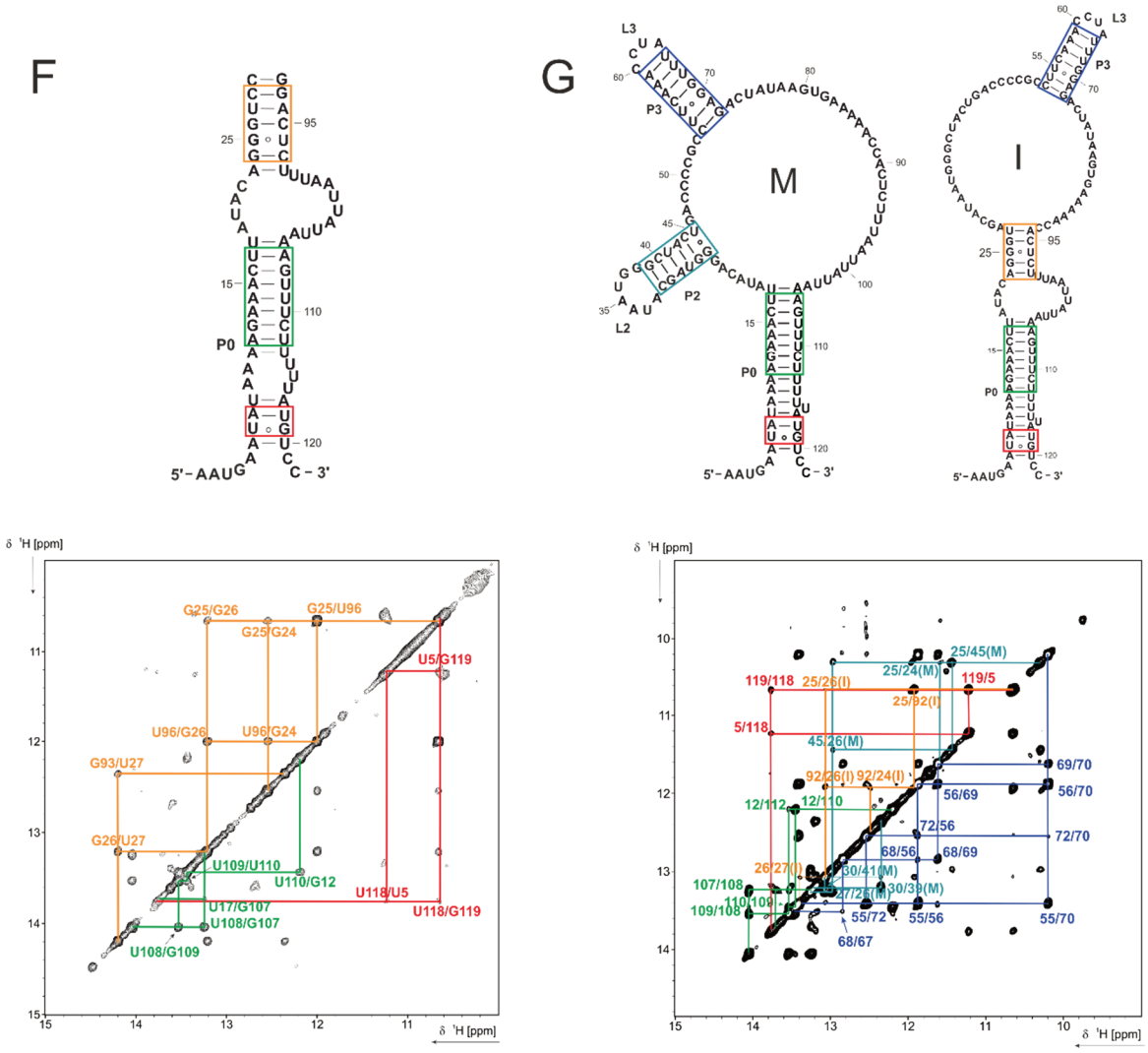


Figure 11: Secondary structure determination of dGsw^{ON}. A) Secondary structure representation of assembled dGsw^{ON} constructs (dGsw$^{0\text{-}85\text{-}122}$, dGsw$^{0\text{-}62\text{-}122}$, dGsw$^{0\text{-}35\text{-}122}$) from two fragments without ligation. Isotope labeling patterns of the individual fragments are indicated. B) Overlay and assignment of NOESY spectrum of dGsw$^{0\text{-}62\text{-}122}$ (black) and dGsw^{122} (red) for verification of correct folding. The two unassigned NOESY pathways that do not correspond to either helix P2, P3 or P0 are indicated in red and orange, respectively. C) Overlay of NOESY ^{15}N(black)/^{14}N(red)-X-(ω2)-filtered spectra of dGsw$^{0\text{-}85\text{-}122}$. The spectra were recorded at 900 MHz, 283 K with a mixing time of 150 ms on 400 µM of each RNA fragment. D) Overlay of NOESY ^{15}N(black)/^{14}N(red)-X-(ω2)-filtered spectra of dGsw$^{0\text{-}62\text{-}122}$. The spectra were recorded at 800 MHz, 283 K with a mixing time of 100 ms on 500 µM of each RNA fragment. E) NOESY spectrum of dGsw$^{0\text{-}35\text{-}122}$ ^{14}N-X-(ω2)-filtered. The spectrum was recorded at 700 MHz and 283 K with a mixing time of 100 ms and 400 µM of each RNA fragment. Unassigned pathways highlighted in red and orange are shown in each NOESY spectrum including the secondary structure information derived from each of the experiments. F) Assembling of the proposed elongated antiterminator according to results shown in B-E and corresponding NOESY spectrum for verification. The spectrum was recorded 900 MHz, 283 K on 500 µM of each RNA fragment. G) Final assignment of dGsw^{ON} and identified secondary structures (M: multi loop antiterminator conformation; I: interior loop antiterminator conformation) in conformational exchange. Continuous pathways of segments of P0 (red, green, orange), P2 (cyan), and P3 (blue) are color coded accordingly in both NMR spectra and secondary structures.

Nevertheless, the construct contributes to the identification of the orange assignment pathway. $dGsw^{0\text{-}35\text{-}122}$ represents the most stable of the assembled constructs and is conformationally homologous to $dGsw^{ON}$. The final assignment for the missing NOESY pathways was achieved by recording $^{1}H\text{-}^{15}N$- or $^{1}H\text{-}^{14}N$-(ω2)-X-filtered NOESY experiments, by which inter- and intra
strand NOE transfers can be distinguished. The combination of X-filtered NOESY analysis of the three assembled ON-state constructs, as shown in Figure 11B-E, reveal that G24, G25 and G26 are involved in an alternative base pairing interaction competing with helix P2 formation and yield a final hypothesis for the conformation of $dGsw^{ON}$. The results suggest that the genetic ON-state is formed by two structures in conformational exchange. The multi-loop conformation (M) (Figure 11G) retains helix P2 and P3 from the aptamer domain, while in the interior-loop conformation (I), helix P2 is destroyed to form an elongated, stabilized antiterminator helix PT. To cross-validate the presence of the elongated antiterminator helix, this helical segment was artificially reconstructed from two fragments, while the upper segment was stabilized by two additional G-C base pairs (Figure 11F). The imino proton NOESY pattern of this assembled model antiterminator helix is in agreement with the NOESY spectrum of $dGsw^{ON}$. The artificially implemented G89 appears as an additional signal in the NOESY spectrum of the model helix. Significant chemical shift perturbations can only be observed for U27 and derive from the additional G-C base pairs.

The complete NMR characterizations of $dGsw^{ON}$ was performed at 283 K in the absence of Mg^{2+}, since neither temperature nor Mg^{2+} have significant effects on the conformation adopted by $dGsw^{ON}$ (Figure 13). Addition of Mg^{2+} causes minor chemical shift perturbations, which are negligibly small compared to Mg^{2+}-induced changes within the aptamer domain (Figure 13A). Further, chemical shift perturbation cannot be linked to a specific structural element within $dGsw^{ON}$ (G69:P3, G12:P0, G39:P2, U68:P3). Signals corresponding to the antiterminator helix PT (G25I, G119, and U5) may appear more pronounced in the presence of Mg^{2+} but Mg^{2+}-induced signal broadening leads to a general decrease in signal intensity. Increasing the temperature to 298 K also leads to an overall loss in imino proton signal intensity, despite the large size of the RNA (Figure 13B).

Direct G-C and A-U base pairing within $dGsw^{ON}$ was investigated by the HNN-COSY spectrum depicted in Figure 13C. This experiment resolves the signal overlay of G72/G24(I), which cannot be resolved in the TROSY spectrum depicted in Figure 13A.

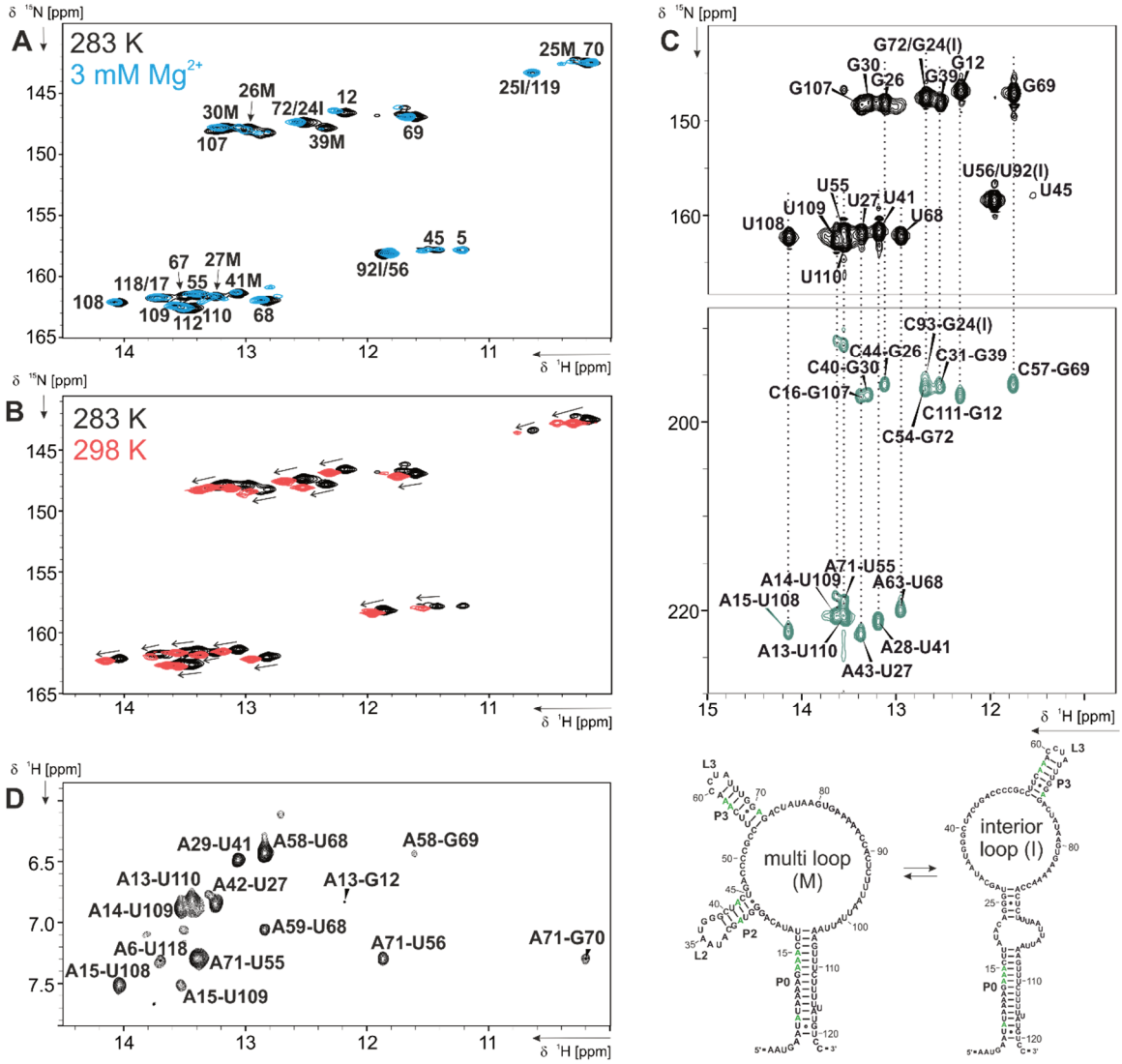


Figure 13: A) Overlay of TROSY spectrum of dGsw^{ON} recorded at 283 K and 900 MHz (black) with a TROSY spectrum of dGsw^{ON} in the presence of 3 mM Mg^{2+} recorded at 283 K. B) Overlay of TROSY spectrum of dGsw^{ON} recorded at 283 K and 900 MHz (black) with dGsw^{ON} recorded at 700 MHz and 298 K (red). C) HNN-COSY spectrum of dGsw^{ON} recorded at 900 MHz and 283 K. D) ^{13}C-X-(ω1)-filtered NOESY spectrum showing only cross peaks between imino protons and adenine H2. The spectrum was recorded at 800 MHz and 283 K.

Further, adenine H2 could be assigned for helices P2 and P3 and parts of the antiterminator helix PT by a ^{13}C-filter in ω1 in a NOESY spectrum on ^{13}C-adenine labeled dGsw^{ON} (Figure 13D). However, the elongated helix PT in the interior loop conformation consists only of three stable base pairs G24-C91, G25-U92 and G26-C93 and cross peaks to neither adjacent adenine residues A90 or A23 can be detected due to a relatively weak base pairing with U27 and U94, respectively. Therefore, the relative equilibrium population between the two conformations cannot be conveniently quantified on the basis of non-exchangable protons. In order to examine exchange rates between the interior loop and multi loop

conformation and to possibly determine the equilibrium population between conformations I and M, ^{15}Nzz exchange experiments were performed at two distinct temperatures (283 K, 298 K) and different exchange delays (70 ms, 150 ms, 200 ms, 300 ms) in the presence of Mg^{2+} (3 mM) and in absence of Mg^{2+} at 700 MHz. However, a cross peak between G25I and G25M could not be detected in any of the recorded experiments. Therefore, the two conformations likely exhibit life times within the seconds regime, a timescale that is too slow for ZZ-exhange experiments to detect.

3.1.3 Allosteric conformational switch in *Mesoplasma florum* riboswitches

2'dG-sensing riboswitches from *Mesoplasma florum* belong to the class of purine-sensing riboswitches. Comparison of the allosteric conformational switch found in the I-A 2'dG-sensing riboswitch to well-studied purine-sensing riboswitches[71] reveals a discrepancy in the ligand-dependent rearrangement of mutually exclusive conformations. In well-studied purine-sensing riboswitches, strand A is involved in antiterminator formation (Figure 14), leading to a small separation in nucleotides between the two complementary strands below 33nt.[71] In case of the I-A 2'dG-sensing riboswitch, strand P is involved in antiterminator formation, and therefore the separation of the two strands is much larger (85 nt for the multi loop antiterminator conformation). In both the *xpt* guanine-sensing riboswitch from *B.subtilits* and the I-A 2'dG riboswitch, segments A (*xpt* Gsw) and P (I-A dGsw) are composed of a free segment located 5' (I-A dGsw) or 3' (*xpt* Gsw) adjacent to the aptamer domain and a segment involved in helix P1 formation.

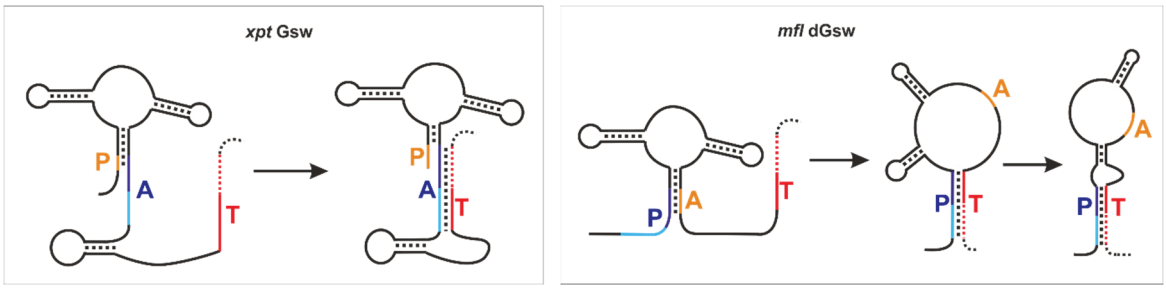


Figure 14: Schematic representation of the rearrangement between strands P (5'-aptamer strand), A (aptamer stabilizing strand) and T (switching strand) to form the antiterminator helix at low ligand concentrations during transcription. In the *xpt* guanine-sensing riboswitch from *B.subtilis*, the PA aptamer interactions dissociates to form the antiterminator helix between strands A and T. In the I-A 2'dG-sensing riboswitch from *M.florum*, the PA interaction dissociates to form the antiterminator helix between strands P and T.

In the *xpt* Gsw, 5'-3' synthesis of the complementary antiterminator strand T allows association of part of the antiterminator helix during transcription prior to 3'-P1 melting. If the 5'-end of P1 is involved, either melting of P1 precedes association of the antiterminator helix or strand T needs to be completely synthesized to initiate refolding. Consequently, the structural rearrangement in well-studied purine-sensing riboswitches appears to be more adapted to the progress of transcription.

In order to investigate if the structural rearrangements found in the I-A *mfl* dGsw is a universal feature of 2'dG-sensing riboswitches or bacterium specific, the allosteric conformational modulation of two additional *mfl* riboswitches (I-B dGsw, III-B Gsw) was investigated (I-B aptamer by Irene Bessi, III-B and I-B full-length riboswitches by Alexander Mazur and Strahinja Lucic). The I-A dGsw and I-B dGsw represent the only identified riboswitches to bind 2'dG with high selectivity and reject similar related purine derivatives.[1] The III-B Gsw binds guanine selectively, however, by two orders of magnitude weaker compared to the *xpt* Gsw from *B.subtilis* (see Chapter 1.2.4). According to theoretical folding predictions, the aptamer domain of both I-B and III-B riboswitches is located prior to a putative terminator hairpin and gene expression is regulated on the level of transcription. For the III-B Gsw, single round transcription termination assays confirm ligand-dependent transcription termination.[1] Figure 15 shows secondary structures including a preliminary assignment in $^{1}H,^{15}N$-correlation spectra of respective OFF and ON-states of I-B dGsw and III-B Gsw (Bachelor Thesis, Strahinja Lucic). The regulatory OFF-state including aptamer domain (PA) and terminator hairpin (TH) could be characterized by NMR completely. The terminator helix TH is adopted also in the absence of ligand for both I-B dGsw and III-B Gsw riboswitches, consistent with previous observations for kinetically controlled riboswitch regulation. The genetic ON-state of these two riboswitches was investigated in the transcriptional intermediates I-B $dGsw^{121}$ and III-B Gsw^{95}. NMR analysis confirms the predicted ON-state for III-B Gsw. However, for I-B dGsw, NMR analysis suggests the presence of an additional conformation, similar to the I-A dGsw, which was not further characterized.

In both riboswitches, the antiterminator helix forms between strands P and T in analogy to the *mfl* I-A dGsw, suggesting this allosteric rearrangement to be a common feature in *mfl* riboswitches.

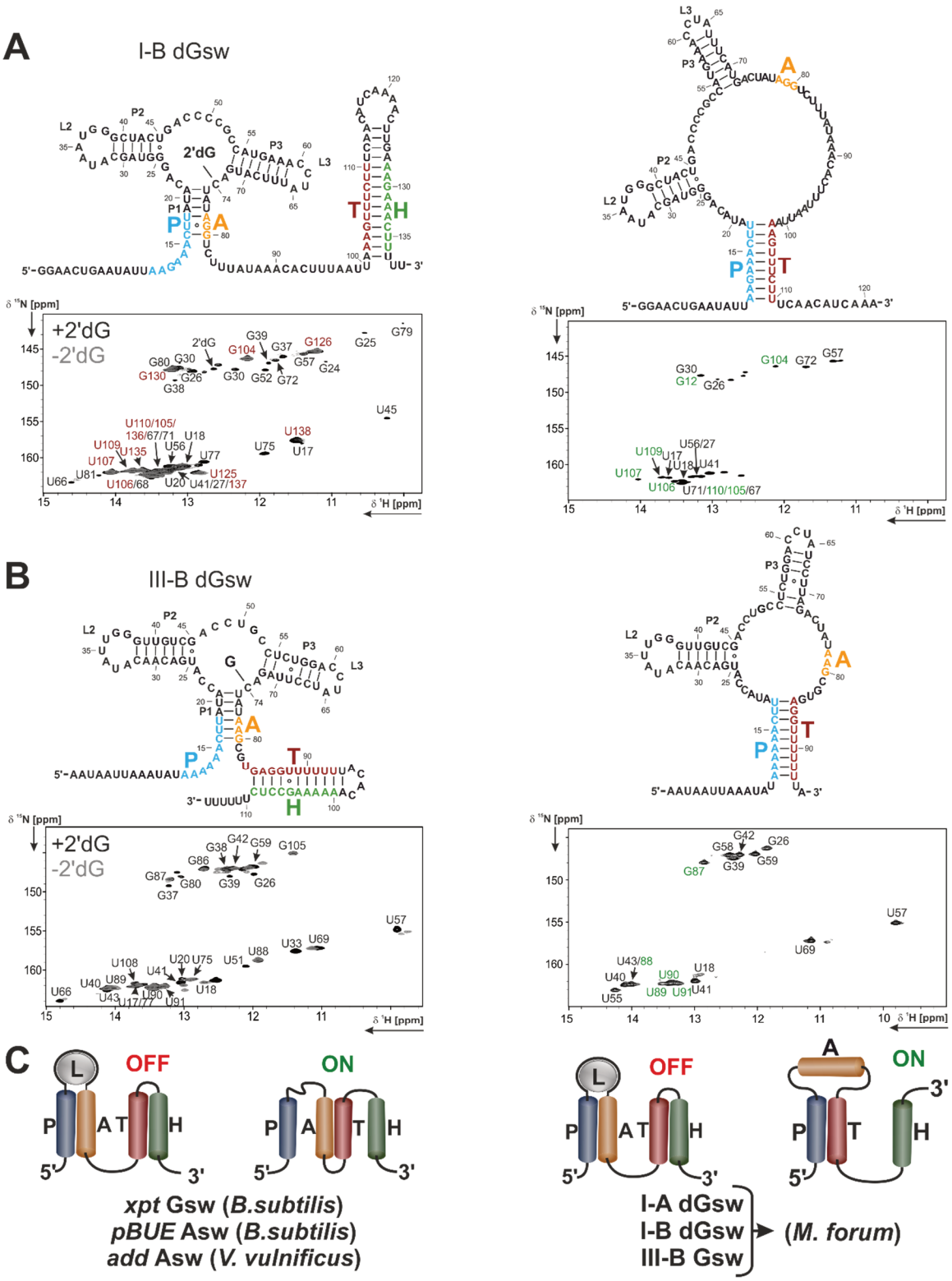


Figure 15: Secondary structure and TROSY assignment of I-B dGsw (A) and III-B Gsw (B). Full-length riboswitches were characterized in both presence (black TROSY spectrum) and absence of ligand (grey TROSY spectrum) on the left. Transcriptional intermediate ON-state representatives were characterized on the right in absence of ligand. C) Schematic representation comparing the ligand-dependent rearrangement of mutually exclusive conformations involving the four strands P (5'-aptamer strand), A (aptamer-stabilising strand), T (switching strand) and H (terminator strand) in common purine-sensing riboswitches (left) and *M. florum* riboswitches (right).

Regarding III-B Gsw, it appears to be remarkable that gene regulation can be controlled kinetically considering the extremely short distance between aptamer domain and terminator helix (3 nt). Typically, the average distance between aptamer domain (PA) and terminator helix (TH) exceeds 20 nts, and the time for the polymerase to transcribe these nucleotides generates a critical time span to allow ligand association to the aptamer domain and prevent antiterminator (PT) formation. The aptamer domain is commonly followed by a pause site to extend this time span. In III-B Gsw, ligand binding must occur directly after the aptamer domain is transcribed to ensure termination of transcription. Despite the shorter distance between aptamer domain and terminator hairpin, *in vitro* transcription termination assays of III-B dGsw show that this riboswitch achieves 75% transcription termination at high guanine concentrations, while I-A dGsw only reaches a termination efficiency of 60% at high 2'dG concentrations. Both riboswitches exhibit a similar K_D in the range of 80-100 nM. The high termination efficiency of III-B Gsw may also be a consequence of its short terminator hairpin (26 nt for III-B Gsw compared to 37 nt for I-A dGsw and 35 nt for I-B dGsw), which could significantly reduce the window of opportunity for ON-state folding. This hypothesis would be supported by the 33% of termination observed in the absence of ligand for III-B Gsw opposed to 21% for I-A dGsw.

In comparison to the guanine-sensing riboswitch regulating the *xpt* gene in *B.subtilis*, III-B Gsw exhibits a K_D that is one order of magnitude larger. This change in affinity may be significant for III-B Gsw to strictly discriminate against hypoxanthine, while both guanine ($K_D \leq 5$ nM) and hypoxanthine ($K_D \sim 50$ nM) bind to *xpt* Gsw.[35] III-B Gsw regulates the expression of GMP synthase (representative for the *guaA* motif in *B.subtilis* in Figure 4), while the *xpt* Gsw regulates the expression of xanthine phosphoribosyltransferase. The different selectivity may be correlated to GMP synthase generating GMP from XMP instead of guanine, provided guanine concentrations are low, whereas XMP synthesis is repressed if either hypoxanthine, xanthine or guanine concentrations are high.

3.2. Rapid NMR screening of RNA secondary structure

3.2.1 Introduction

In Chapter 3.1, the three allosterically responsive conformations leading to either progression (ON-state) or termination (OFF-state) of transcription were identified by NMR. According to structural investigations on $dGsw^{FL}$, the full-length riboswitch exclusively adopts the functional OFF-state. However, during transcription, folding towards the most stable conformation continuously evolves with the addition of new nucleotides.[5,6] Therefore, the conformation adopted at a particular time during transcription may not represent a thermodynamically equilibrated state, and transiently adopted states may not reach equilibrium within the given window of time before the thermodynamic equilibrium shifts towards a different conformation. A methodological approach that would allow monitoring of co-transcriptional folding from a structural perspective has not yet been developed. Both speed and low RNA concentrations during *in vitro* transcription complicate *in-situ* analysis of the transcriptional progress by NMR. Alternatively, the knowledge of folding states of transcriptional intermediates, interconversion rates between conformational states and the speed of transcription, including potential pause sites, could be combined to simulate how riboswitch conformations evolve during transcription. The former issue can be addressed by investigating the folding state of each transcriptional intermediate at single nucleotide resolution to mimic the transcriptional progress from a thermodynamic perspective. However, established approaches for RNA preparation for NMR-spectroscopy are time-consuming with a low throughput. In this section, a method to rapidly synthesize RNA at single nucleotide resolution for NMR-spectroscopic analysis is described.[8]

RNA preparation for NMR spectroscopy includes DNA template preparation, RNA transcription and its purification. In general, RNA is transcribed by T7 polymerase from linearized plasmids or PCR templates.[161–163] To synthesize a multitude of DNA templates that differ by a single nucleotide in length, the method of choice for DNA template preparation is PCR. PCR allows simultaneous synthesis of all DNA templates with the reverse primers being the only parameter that needs to be adjusted. It is commonly known that T7 RNA polymerase produces inhomogeneous 3'-ends by appending one to three random nucleotides to the 3'-end[162,164–166] and to circumvent this issue, self-cleaving ribozyme cassettes[167–171] have been widely applied. However, implementation of ribozymes requires more extensive DNA template preparation and decelerates the

purification process. To screen exact sequence dependent conformational transitions at single nucleotide resolution, 3'-end homogeneity of transcripts is imperative. A possible solution to obtain 3'-homogeneous transcripts from PCR templates is addressed in section 3.2.2.

Finally, ion-pair reversed phase HPLC or polyacrylamide gel electrophoresis represent the standard procedures to purify RNA for NMR-spectroscopy.[172,173] These methods are time-consuming with a limited throughput and therefore inefficient to facilitate rapid synthesis of a large amount of different RNA constructs. Section 3.2.3 describes an alternative approach to synchronize removal of perturbing transcriptional additives under native conditions to allow rapid structural screening by NMR.

3.2.2 Effect of DMSO on transcription yield and homogeneity

To address the issue of transcript homogeneity obtained from run-off transcription reactions with T7 RNA polymerase, the effect of various transcription additives (Triton X-100, DMSO, PEG 6000) was screened in transcription assays. Of the three tested additives, DMSO proved to be the most effective regarding improved transcript homogeneity and yield. No increase in 3'-end homogeneity could be observed by the addition of PEG 6000, while the addition Triton X-100 had a minor effect on transcript homogeneity.

In general, transcriptions of the I-A 2'dG-sensing riboswitch exhibit an unusually high degree of 3'-heterogeneity. Inhomogeneous additions by T7 polymerase are presumably caused by the high A/U content, which exceeds 75% in the expression platform. DMSO transcription optimizations were first performed on constructs $dGsw^{104\text{-}112}$, representing the riboswitch segment during which the antiterminator helix is continuously stabilized (Figure 16A, green), and $dGsw^{134}$-$dGsw^{137}$, representing the segment during which the terminator helix is continuously stabilized (Figure 16A, red).

Figure 16 shows gels of transcriptions performed in the presence and absence of 20% of DMSO. In the absence of DMSO, RNA fragments with the correct length are only transcribed from five DNA templates, namely $dGsw^{107}$, $dGsw^{112}$, $dGsw^{135}$, $dGsw^{138}$ and $dGsw^{139}$. The remaining transcription products exhibit a high degree of heterogeneity, with the main products being >+4 nt longer than the desired transcript length. Since all transcripts were co-transcribed with a 5'-hammerhead ribozyme, inhomogeneities can be unambiguously assigned to the 3'-end. In the presence of 20% of DMSO, the homogeneity of transcription products significantly improved and the corresponding RNA length appears

to conform to the desired transcript length. However, due to the size of the RNA, the gel cannot resolve single nucleotides.

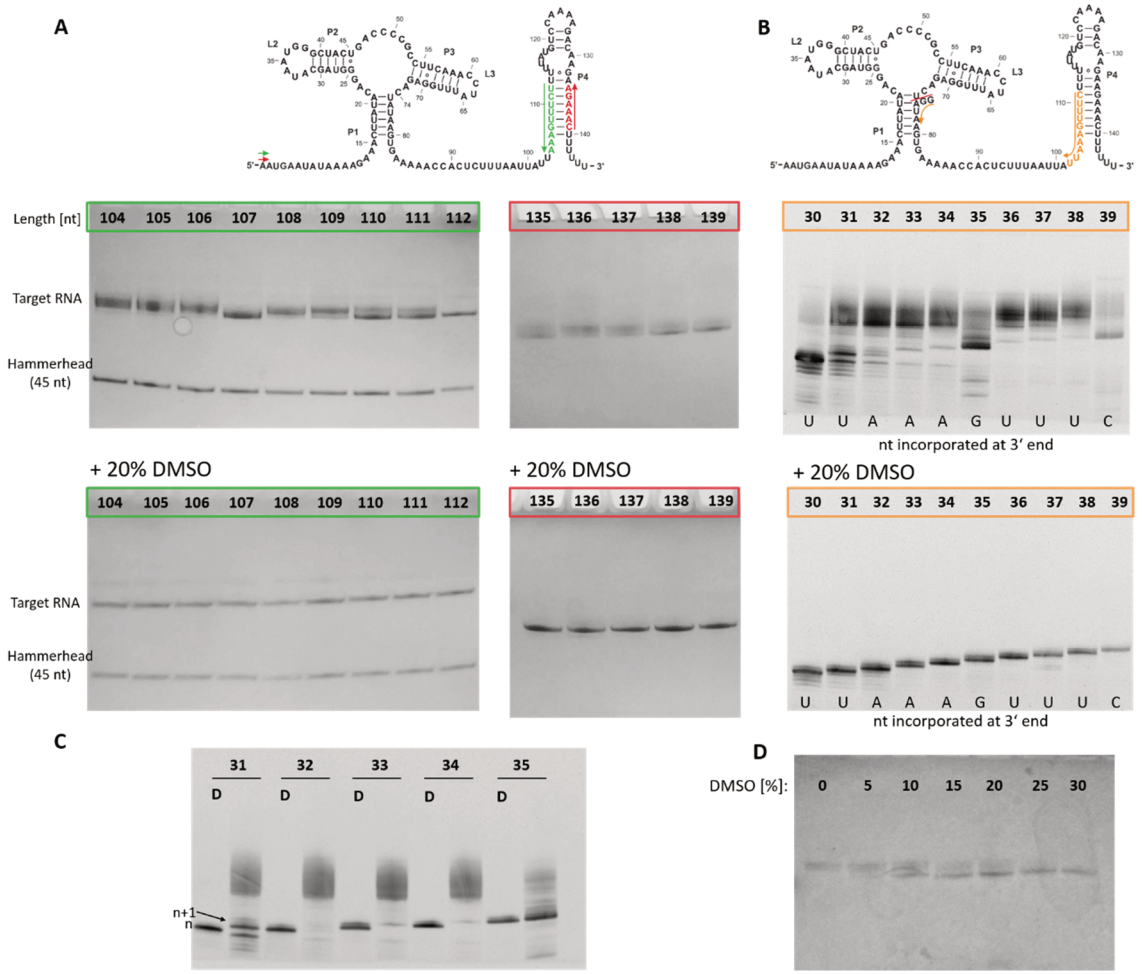


Figure 16: A) 12% denaturing polyacrylamide gels showing transcription products for dGsw^{104}-dGsw^{112} (green) and dGsw^{134}-dGsw^{138} (red) of transcription reactions performed under standard conditions (upper gels) and in the presence of 20% of DMSO (lower gels). B) 20% denaturing polyacrylamide gels showing transcription products of dGsw$^{75\text{-}102}$-dGsw$^{75\text{-}111}$ (orange) in the absence (upper gel) and presence (lower gel) of DMSO. Segments transcribed within dGsw are color coded accordingly in the depicted secondary structures. C) Direct comparison of the DMSO effect for five selected RNA transcripts shown in B. Lanes marked with a D correspond to transcription reactions performed in the presence of 20% of DMSO. D) DMSO optimization for the 36 nt long transcript shown in B.

To resolve single nucleotides, smaller segments of the riboswitch were transcribed starting from nucleotide 75 and ending at residues identical to constructs investigated in Figure 16A (Figure 16B). To allow transcription initiation by T7-polymerase, the T7 promotor including two additional guanine residues were artificially implemented within the forward primer. Gel analysis of the smaller fragments (30-39 nt) shows that in the presence of DMSO, the majority of fragments are transcribed with a high degree of homogeneity with the exact sequential length. When comparing the lengths of the different fragments that are

designed to differ by a single nucleotide, n+1 fractions can be detected for most constructs. For the 33nt length construct, even in the presence of DMSO, the extend of n+1 addition still exceeds 40%. However, n+2 products are not transcribed. Figure 16C shows a direct comparison of transcription reactions performed in the absence and presence of DMSO for five selected constructs (31 nt - 35 nt). In the absence of DMSO, a minor fraction corresponding to n+1 transcripts can be detected, while RNA transcripts with the correct length are not transcribed. Only constructs with a length of 35 nt and 39 nt are transcribed with the desired length in the absence of DMSO. Figure 16D shows DMSO optimizations (0-30%) on the 36 nt transcript. According to these optimization, 20% of DMSO in the transcription mixture appears to be optimal regarding both yield and homogeneity.

When investigating a potential correlation between the sequence and random nucleotide additions, it becomes apparent that incorporation of a G (35 nt transcript) and C (39 nt transcript) considerably reduces the addition of non-native nucleotides by T7 polymerase. Further, a decrease in heterogeneity can be observed for the 30 nt transcript. This decrease in heterogeneity could be attributed to additional C residues located prior to dGsw^{93} that stabilize the DNA/RNA duplex during transcription and that are located prior to the subsequent 13 nt long A/U stretch ranging from dGsw^{94}-dGsw^{106}.

Transcription reactions using DMSO as a cosolvent exhibit a significant improval in transcript homogeneiety obtained from PCR products. However, transcription reactions from templates containing long stretches of A/T residues still contain n+1 products; with the extent varying from 10-40%. To investigate sequential folding at single nucleotide resolution, transcripts with exact sequences with homogeneities larger than 90% are required.

It has been reported that the use of 2'-methoxy (2'-OMe) modified primers at the last two nucleotides of the 5'-end significantly reduces non-native addition of +1 and +2 nucleotides by T7 polymerase.[7] Therefore, the effect of 2'-methoxy modified primers was compared to and combined to the use of DMSO as a cosolvent. The comparison was performed on aptamer constructs of the I-A dGsw (dGsw^{77}-dGsw^{81}) and is shown in Figure 17. Transcription products of aptamer constructs are less heterogeneous compared to constructs that contain the expression platform. The decreased transcription heterogeneity can be attributed to the larger G/C content in the aptamer domain (38% vs. 25% in the expression platform). Figure 17A shows that the 2'-methoxy modified primers cause a significant reduction in the transcription yield. For most constructs, the transcription yield is reduced

by approximately 50%, while for dGsw[79] and dGsw[80] the decrease in yield exceeds 80%. ≤+4 nucleotide additions to the 3'-end can successfully be reduced by implementing 2'-OMe modifications in the DNA template. However, the DNA modification does significantly affect the transcription of longer side products. Figure 17B compares the effect of DMSO to the use of 2'-OMe modified DNA. The addition of DMSO prevents the formation of longer transcripts much more efficiently than the use of 2'-OMe modified primers. Only for constructs dGsw[75], dGsw[76], and dGsw[80], a minor n+1 fraction is transcribed in the presence of DMSO. The gel shown Figure 17C compares the addition of DMSO as a cosolvent and in combination with 2'-OMe modified DNA. According to this comparison, 2'-OMe modified DNA in the presence of DMSO completely prevents n+1 additions also for dGsw[75], dGsw[76], and dGsw[80].

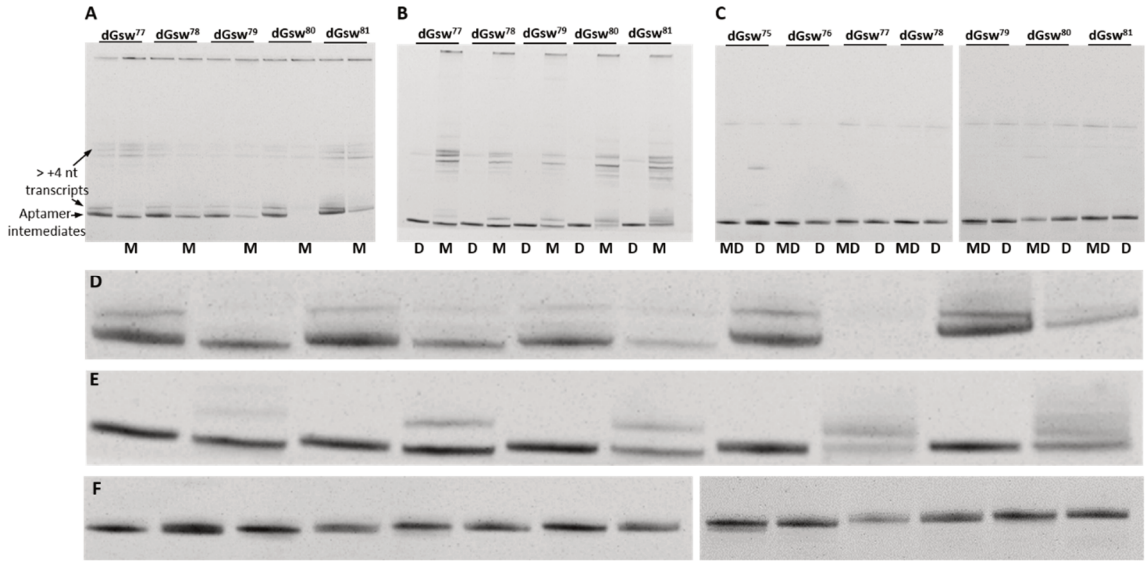


Figure 17: 20% polyacrylamide gels investigating the 3'-end homogeneity of I-A dGsw aptamer transcripts. Comparison of run-off transcriptions from: A: 2'-methoxy modified DNA (M) and standard PCR products, B) 2'-methoxy modified DNA (M) and standard PCR products in 20% of DMSO (D), C) standard PCR products in 20% of DMSO and from 2'-methoxy modified DNA in 20% of DMSO (MD), D, E, F) enlarged section of the target sequences of the gels shown in A), B), and C). The Figure was adapted from Helmling *et al.*[8]

In summary, the effect of 2'-OMe modified transcription templates appears to be limited to n+1 and n+2 additions by T7 polymerase, as reported. Regarding these n+1 and n+2 additions, 2'-OMe DNA is superior to the addition of DMSO as a cosolvent. However, the formation of longer byproducts produced during transcription is more efficiently inhibited by DMSO. As reported previously, these longer transcription products may be caused by RNA primed DNA template extension.[174,175] In contrast, unspecific n+1 and n+2 additions are non-templated nucleotide additions.[161,166] Therefore, these two sources of transcript

inhomogeneity derive from unrelated events and can be successfully inhibited by the combination of 2'-OMe modified DNA with DMSO to yield highly pure and homogeneous transcripts from run-off transcriptions.

3.2.3 Native high-throughput purification

The transcription of highly homogeneous and pure RNA transcripts allows acceleration of the purification procedure, since no separation of RNA fragments is required. In contrast to time-consuming HPLC or PAGE purification, the transcription mixture can be directly applied to centrifugal devices, washed with water and exchanged into NMR buffer (method developed by Sara Keyhani). This procedure removes NTPs, transcription additives and ions. Due to their size, enzymes (T7 RNA polymerase, inorganic pyrophosphatase) and the DNA template cannot be removed in centrifugal concentrators and remain in the final sample. However, the concentration of those components is below the NMR detection limit (2.4 µM for T7 RNAP and <8.5 µM for the DNA template). Figure 18 shows 1D NMR spectra comparing HPLC purified dGsw^{80} with dGsw^{80} exchanged with different amounts of NMR buffer (60 mL and 120 mL corresponding to dilution factors of <10^{-8} and 10^{-16}, respectively).

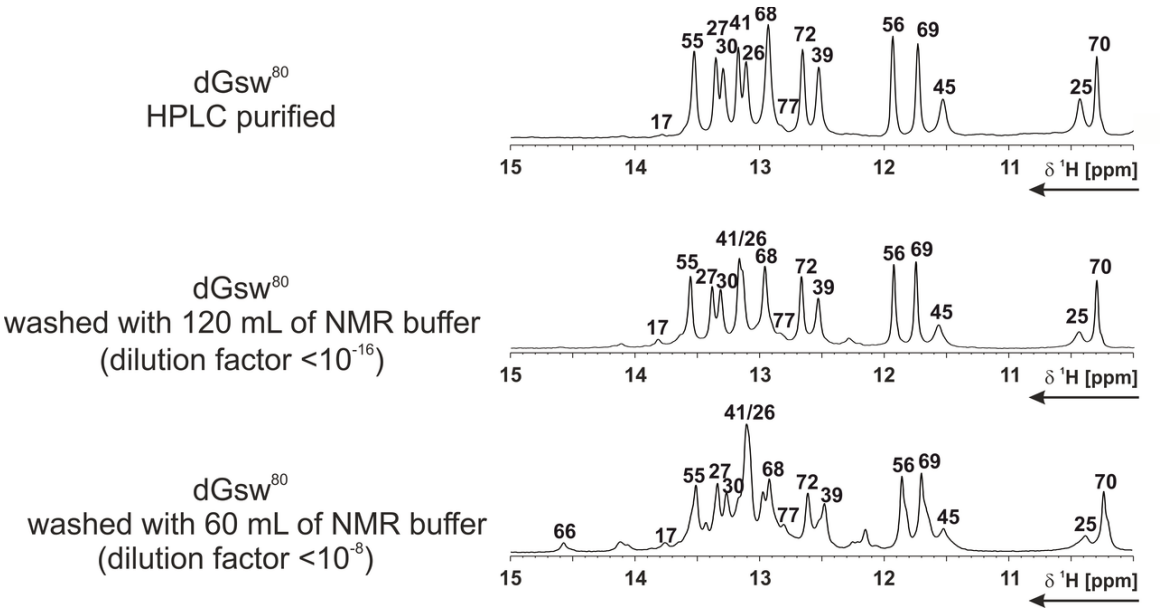


Figure 18: Comparison of ^{1}H-NMR spectra recorded of HPLC purified dGsw^{80}, dGsw^{80} washed with 60 mL of NMR buffer and dGsw^{80} washed with 120 mL of NMR buffer. The spectra were recorded at 800 MHz and 298 K.

The imino proton pattern of buffer exchanged and HPLC purified spectra can be considered identical provided that the transcription mixture is washed with sufficient amount of water (> 120 mL). In case of dGsw[80], a washing volume of 60 mL is not sufficient to remove RNA binding additives present in the transcription mixture. Transcription additives, such as Mg^{2+} and spermidine, are known to bind to RNA and lead to increased thermal stability.[110,176] This interaction can cause chemical shift perturbations and lead to alterations in the imino proton pattern. The final amount of washing buffer required for complete removal of RNA binding ions depends on their binding affinity to the RNA and is therefore construct specific. In particular, 2'dG riboswitch constructs with a fully synthesized P1 helix bind Mg^{2+} and spermidine more tightly to prepare the binding pocket for ligand binding.[120,126] Therefore, these constructs require more extensive washing compared to shorter constructs within the aptamer domain or fragments containing the expression platform.

3.2.4 Overview of the screening procedure

Figure 19 schematically summarizes the procedure that allows rapid screening of RNA secondary structure by NMR. The procedure is composed of three steps: 1) PCR to prepare DNA templates for transcription intermediates can be performed simultaneously, with the reverse primer representing the only parameter that needs to be adjusted; 2) The multiple PCR products are directly applied for overnight transcriptions without further DNA purification; 3) Following overnight transcription, transcription mixtures are directly applied to centrifugal concentrators and buffer exchanged. Depending on the size of the RNA and consequently the required membrane, extensive buffer exchange can take up to 10 hours. The whole procedure of NMR sample preparation can be performed within a total of two days. In addition, the RNA retains its native fold and is not denatured during the purification procedure. The number of RNA constructs that can be prepared simultaneously by this approach is limited by the availability of centrifuges. In most cases, 1D NMR screening is sufficient to monitor sequence-dependent changes in folding based on the imino proton pattern, and can be performed within one day. Following 1D NMR screening, selected constructs can be further isotope labeled and analyzed by multidimensional NMR experiments. Since buffer exchange is relatively time-consuming, thermodynamic equilibration of secondary structures after transcription is inevitable. However, screening of equilibrated structures still contributes significantly to improve understanding of co-

transcriptional folding pathways; in particular, because potential transient and/or misfolded states can be detected by this approach.

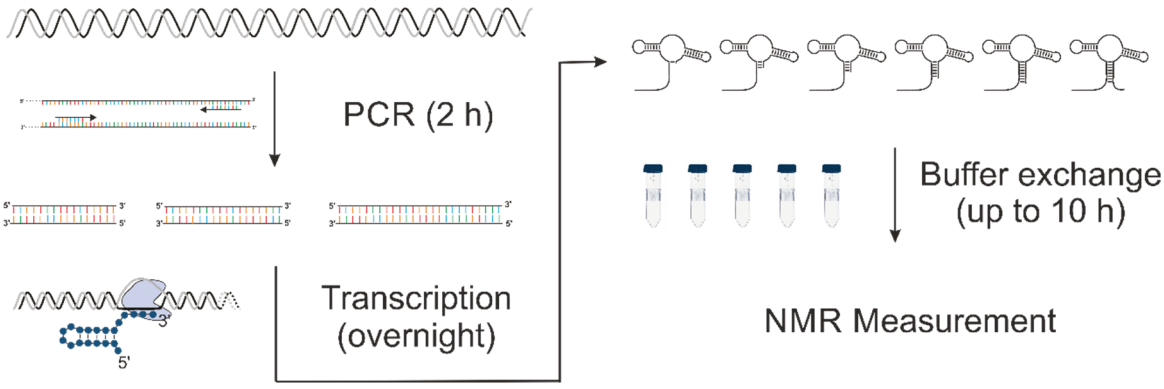


Figure 19: Schematic representation of the three step screening approach. Multiple DNA templates are prepared simultaneously within two hours followed by overnight transcription. Transcription mixtures are buffer exchanged in centrifugal concentrators and directly measured by NMR.

Chapter IV

Dissecting ON- and OFF-Function through Transcriptional Intermediates

4.1. Introduction

The method described in Chapter 3.2 enables rapid synthesis of multiple transcriptional intermediates at single nucleotide resolution and consequently allows the study of various parameters relevant for riboswitch regulation as a function of the transcript length. In this chapter, transcriptional intermediates are investigated both structurally and with respect to their ligand binding properties by NMR-spectroscopy, ITC and stopped-flow fluorescence spectroscopy. Investigations performed on transcriptional intermediates are divided into the subsections: "ON-function" (4.2) and "OFF-function" (4.3). In Chapter 4.2, secondary structures of transcriptional intermediates are characterized by NMR-spectroscopy in the absence of ligand to monitor conformational transitions between the strands P, A, T and H, as the RNA is extended by single nucleotides. In this context, the effect of temperature and Mg^{2+} on folding states adopted by transcriptional intermediates was investigated. Based on the results of NMR experiments, Markovian simulations were performed to simulate co-transcriptional folding in the absence of ligand. In Chapter 4.3, the ligand affinity of individual transcriptional intermediates was investigated by both NMR and ITC. These experiments were supplemented with kinetics of ligand binding to examine the event of ligand binding in the context of transcription. Chapter 4.4 describes real-time NMR measurements of antiterminator folding monitored in both absence and presence of ligand by means of photo-labile protection groups in the expression platform. Based on obtained kinetic rates and data derived from experiments described in sections 4.2 and 4.3, co-transcriptional folding was simulated in both absence and presence of ligand to describe how the events of ligand binding and structural rearrangement are fine-tuned with the speed of transcription to modulate gene expression in a ligand dependent manner. In Chapter 4.5, the various experiments are further related and discussed.

4.2. ON function (-2'dG)

Conformational transitions between the helical segments PA, PT and PA-TH in relation to the transcript length were fist predicted theoretically using the mfold web server.[112] Figure 20 shows ΔG values for the ON-state (represented by the formation of helices P0, P2 and P3) and the OFF-state (represented by the formation of helices P1, P2, P3 and P4), as the RNA is extended by single nucleotides. According to these theoretical predictions, the terminator conformation also represents the lowest free energy state in the full-length

riboswitch in absence of ligand, while the antiterminator conformation represent the lowest free energy state at transcript lengths dGsw[110]-dGsw[137]. These theoretical predictions are remarkably close to our experimental findings, as described below.

In the secondary structure depiction in Figure 20B, the 29 selected constructs chosen for NMR structural profiling based on these theoretical ΔG values are shown. Transcripts dGsw[75]-dGsw[80] were synthesized to investigate at which RNA length the PA interaction is stably formed, and to determine the shortest riboswitch length required for ligand binding. Constructs dGsw[104] to dGsw[122] were synthesized to identify the length required to switch from the PA aptamer interaction to the antiterminator helix PT. dGsw[134] to dGsw[139] were selected to define the transcriptional length, at which the thermodynamic equilibrium shifts from the PT helix to the terminator conformation PA-TH.

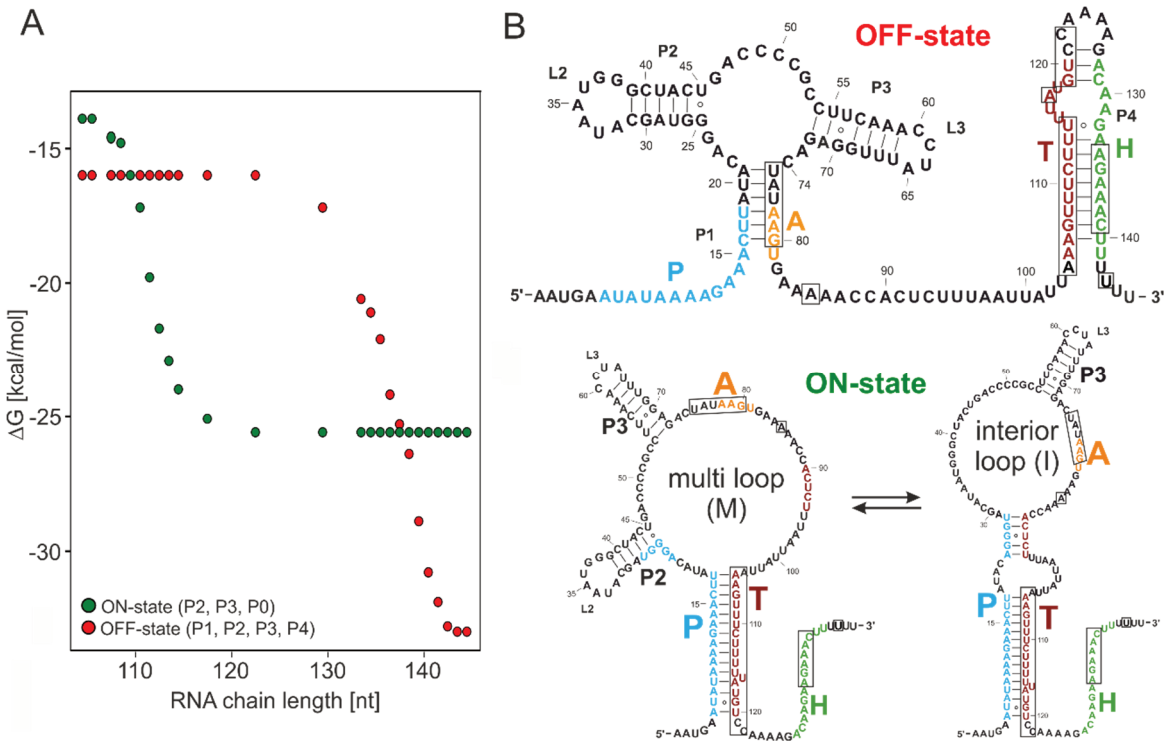


Figure 20: A) ΔG values of the ON-state (green) and OFF-state (red) according to mfold[112] for transcriptional intermediates of various lengths. B) Secondary structure of dGsw[FL] illustrating the 29 transcribed intermediate constructs: aptamer transition constructs dGsw[75]-dGsw[80], PT transition constructs dGsw[104]-dGsw[114], dGsw[117], dGsw[119]-dGsw[122] and PA-TH transition constructs dGsw[134]-dGsw[139], dGsw[144] by segments framed with black boxes.

For kinetically controlled riboswitches, the rearrangement from PT to PA-TH during transcription is considered insignificant, since it is assumed to occur only after the

polymerase has advanced beyond the point of regulatory decision. However, the precise RNA length, at which the antiterminator conformation $P^{A}T$ no longer represents the lowest free energy state, represents a barrier for the folding event from PA to $P^{A}T$. Once the polymerase proceeds past this threshold, ON-state folding can no longer occur.

4.2.1 Structural profiling of transcriptional intermediates in the absence of ligand

Selected transcriptional intermediates of dGsw (indicated in Figure 20), were synthesized as described in Chapter 3.2. Secondary structures were subsequently screened in 1D NMR spectra and the imino proton pattern of each intermediate was assigned to specific structural motifs by means of 2D NMR analysis presented in Chapter 3.1. According to structural characterizations performed in Chapter 3.1, imino protons in weakly base paired segments, in particular closing base pairs of helix PT, are more pronounced at lower temperatures. However, at 283 K, the bulk of imino protons cannot be resolved sufficiently in 1D NMR spectra due to extensive line broadening caused by the decrease in temperature. Therefore, 1D NMR screening within strands T and H were performed at 298 K. Screening of transcriptional intermediates within strands A and helix P1 was performed at both 298 K and 283 K due to the smaller size of the aptamer domain and the relatively weak interactions within helix P1 that contains primarily A/U base pairs. At 298 K and in the absence of Mg^{2+}, the imino proton pattern observed in 1D NMR spectra of $dGsw^{75}$-$dGsw^{80}$ can be considered identical and includes helices P2 and P3. In the absence of Mg^{2+}, imino proton reporter signals for helix P1 (U17, U77 and U18) can be observed only at 283 K in the fully transcribed aptamer domain (Figure 21A-B). Therefore, formation of a stable P1 helix requires formation of the closing G-C base pair G80-C16. Further, formation of a fully stabilized P1 cooperatively stabilizes helix P2, which can be observed by the appearing imino proton signal G39. G39, the closing base pair of helix P2 further indicates the presence of weak loop-loop interactions between helices P2 and P3 even in the absence of ligand and Mg^{2+} at 283 K. Mg^{2+} binding to riboswitch aptamer domains is commonly known to play an important role in stabilizing the aptamer domain fold, in particular with respect to tertiary interactions.[120,126] Therefore, 1D NMR spectra of $dGsw^{76}$-$dGsw^{81}$ and $dGsw^{85}$ were recorded also in the presence of 3 mM Mg^{2+} at 298 K, the physiological temperature of *Mesoplasma florum* (Figure 21C). In the presence of Mg^{2+}, P1 formation can be observed by the reporter signal U77 also at 298 K but requires nucleotide 79 to be

synthesized to form a 5 base pair helix. Reporter signals U17 and U18 can be detected once the aptamer domain is elongated past nucleotide 81. However, U17 appears to be strongly shifted in dGsw^{81} compared to dGsw^{85}. Parallel alignment of helices P2 and P3 in the presence of Mg^{2+} coincide with helix P1 formation and also require 4-5 base pairs in P1 to form. These tertiary interactions can be observed by reporter signals U66, G39, G24 and G25.

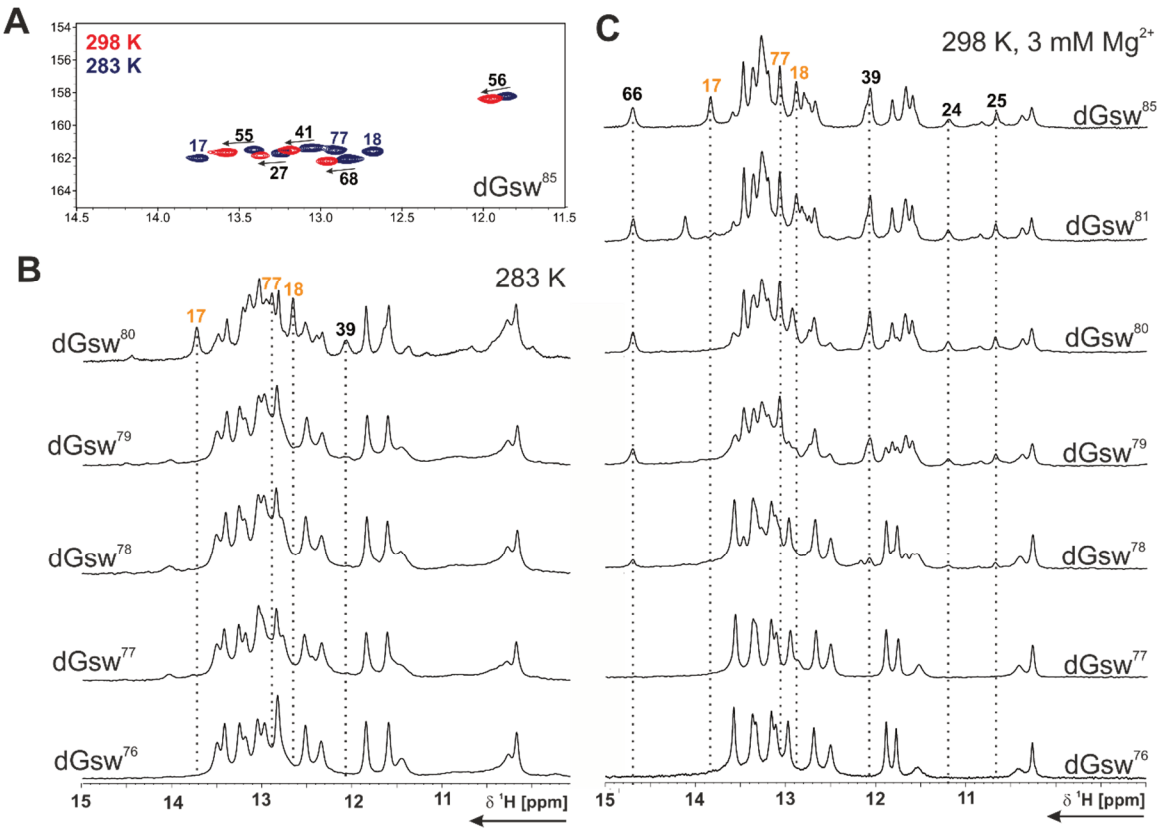


Figure 21: A) Overlay of ^{15}N-HSQC spectra showing imino protons of uridine residues of dGsw^{85} recorded at 283 K (blue) and 298 K (red) and 950 MHz. B) 1D NMR spectra of dGsw^{76}-dGsw^{80} recorded at 283 K and 800 MHz. C) 1D NMR spectra of dGsw^{76}-dGsw^{81} and dGsw^{85} recorded at 298 K and 900 MHz in the presence of 3 mM Mg^{2+}.

Figure 22A,B shows 1D NMR spectra obtained from secondary structure screening of transcripts larger than dGsw^{104}. Characteristic reporter signals to identify the helical interactions PA, PT and TH are color-coded in orange, green and red, respectively. The transition from PA to PT was monitored by isolated reporter signals U18 (PA, orange), G12 (PT, light green) and U108 (PT, dark green).

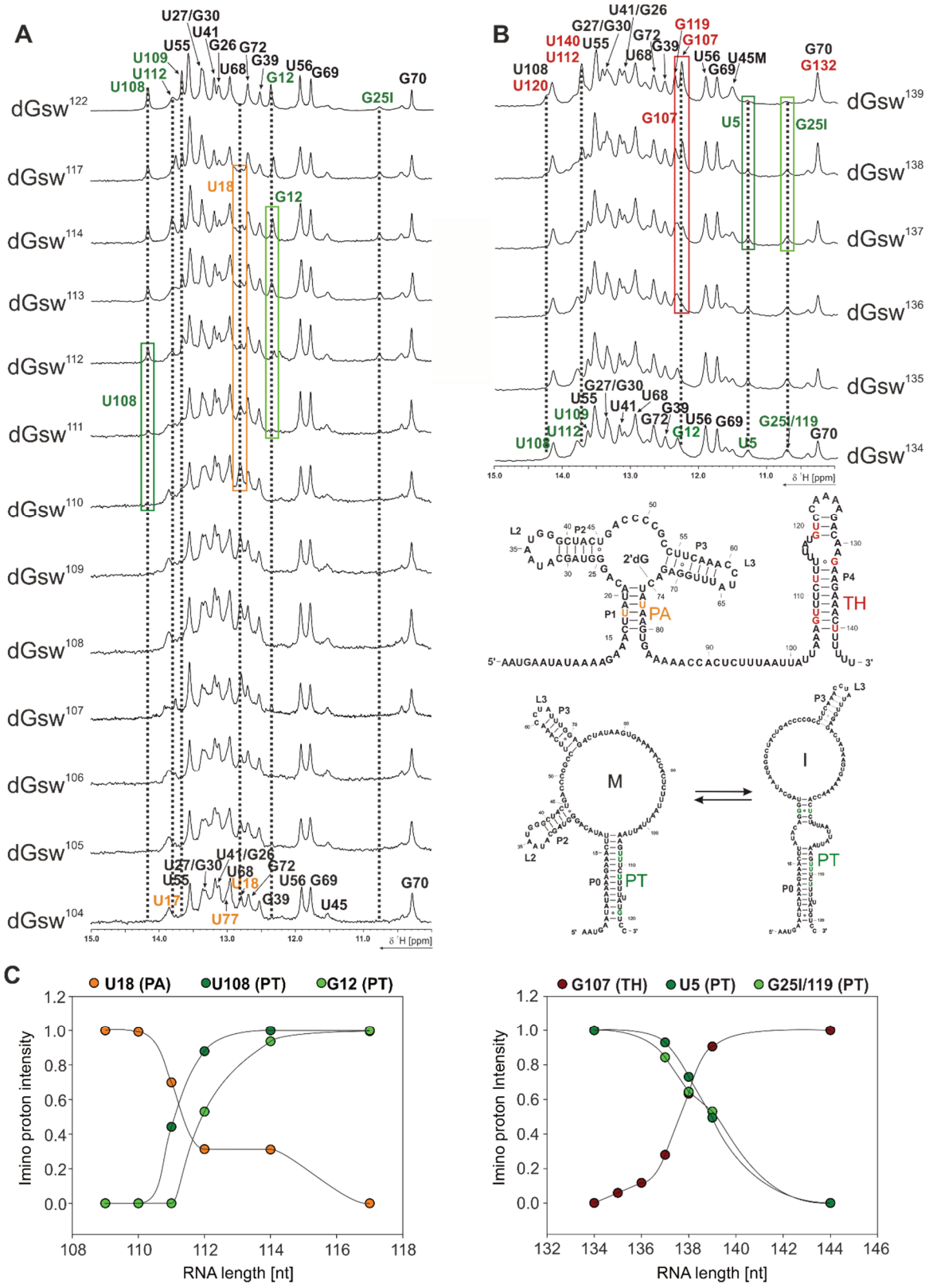


Figure 22: Imino proton region of 1D NMR spectra of 19 transcriptional intermediates showing the transition from PA to PT (A) and from PT to PA-TH (B). Spectra for dGsw^{104} - dGsw^{117} were recorded at 800 MHz and spectra for dGsw^{134} - dGsw^{144} were recorded at 700 MHz. All spectra were recorded at 298 K. Reporter signals for the PA interaction are highlighted in orange, for the PT helix in green and for the TH helix in red. C) Normalized imino proton intensity of reporter signals for PA (orange), PT (green) and PT (red) highlighted by boxes in A and B color coded accordingly.

The transition from PT to PA-TH was monitored by the isolated reporter signals U5 (PT, dark green), G25I (PT, light green) and G107 (TH, red). The signal intensity of these isolated reporter signals in relation to the transcript length is shown in Figure 22C. The PA interaction can be observed up to transcript dGsw^{110}. The transition to form the PT interaction occurs within the extension of three additional nucleotides up to dGsw^{113}. Constructs dGsw^{111} and dGsw^{112} represent bistable transcripts and show reporter signals for both PA and PT. Compared to mfold predictions shown in Figure 20, experiments show that three additional base pairs are required to form a stable antiterminator helix PT. The thermodynamic equilibrium shifts from PT to PA-TH between dGsw^{137} and dGsw^{140}, which conforms with the theoretical predictions.

In addition, it was investigated if Mg^{2+}-induced structural stabilizations are sufficiently strong to shift determined transcript lengths, at which structural transitions between strands P, T and H occur. Mg^{2+} titration experiments on transcript dGsw^{114}, which forms a stable PT helix, are shown in Figure 23A. This experiment shows that gradual addition of Mg^{2+} leads to a significant loss in signal intensity of PT reporter signals G12, U108, U109 or G25I. Mg^{2+} addition only causes signal broadenings and mostly affects chemical shifts for helix P3 signals, which according to results shown in Chapter 3.1, can form weak tertiary interactions with P2 in the antiterminator conformation M. Therefore, addition of Mg^{2+} alone cannot induce a shift in thermodynamic equilibrium from PT to PA (Figure 23A).

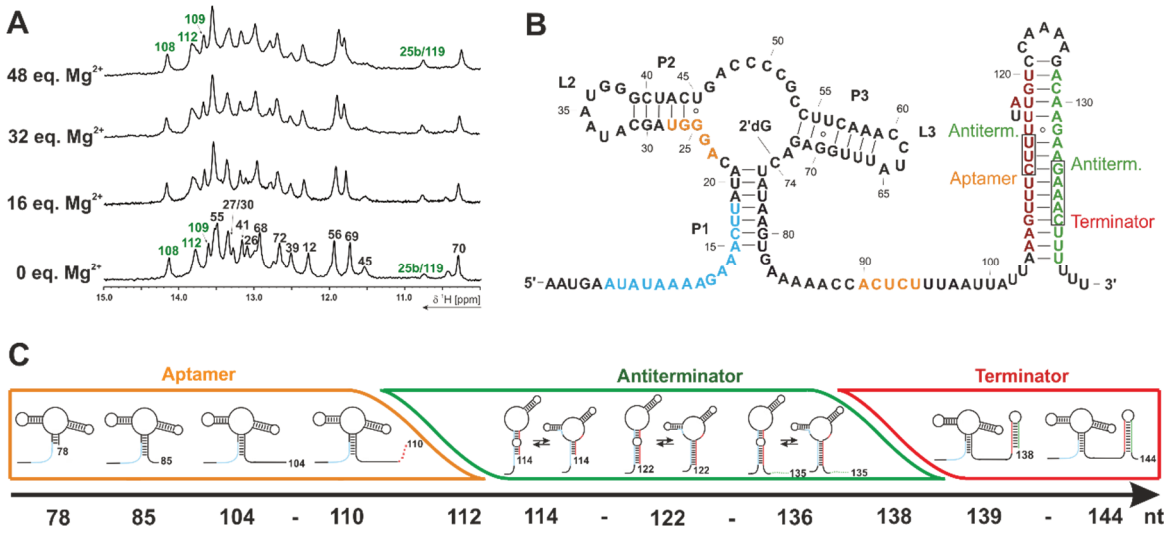


Figure 23: A) Titration of dGsw^{114} with Mg^{2+} at 800 MHz and 298 K. B) Secondary structure of dGsw^{FL} illustrating shifts in lowest free energy states from PA to PT and PT to PA-TH in black boxes. C) Schematic representation of equilibrium state conformations adopted by transcriptional intermediates highlighting the window of opportunity for ON-state folding of 23-24 nts.

The structural screening approach described in this section yields lowest-free energy states for particular transcript lengths. These structures may not represent structures adopted by nascent transcripts as they emerge from the transcription elongation complex, which are rarely at equilibrium. However, the identification of these lowest free energy states restricts the folding event from PA to PT to occur during the transcription of $dGsw^{113}$-$dGsw^{137}$. Folding to the antiterminator conformation $P^{A}T$ must occur within the transcription of these 23-24 nucleotides to ensure gene expression in the absence of ligand (summarized in Figure 23B-C). In the following section, co-transcriptional folding within this derived window of opportunity for ON-state folding is simulated.

4.2.2 Markovian simulations of co-transcriptional folding

Co-transcriptional folding can be simulated based on 9 discrete Markov states. These Markov states represent three distinct RNA conformations at three distinct RNA lengths (Figure 24). The model describes the folding rearrangement from PA (aptamer, state 1), which is always transcribed and folded first, to PT (antiterminator conformations M: state 2, and I: state 3). Folding is initiated at a transcript length of 113 nts ($dGsw^{113}$). While nascent transcripts fold from PA to PT, the RNA is further elongated to a length of 137 nt at the rate of transcription. According to experiments presented in the previous section, during transcription of nts 113 to 137, the antiterminator conformation forms at thermodynamic equilibrium. Therefore, transcripts 113 and 137 represent the boundaries of the 24 nt antiterminator folding window. From transcript length 137, the RNA elongates to 144 nt, the point of regulatory decision. During transcription of these 7 nts, the terminator conformation PA-TH represents the lowest free energy state. Therefore, folding to $P^{A}T$ from PA is arrested and folded antiterminator states $P^{A}T$ can equilibrate slowly to the terminator conformation PA-TH.

The Markov state model (MSM) describes the stochastic dynamics between these 9 states based on a transition matrix assembled from corresponding transition rates. Interconversion rates were first calculated theoretically based on the number of base pairs required to dissociate in order to enter a state of transition. These rates were derived from real-time NMR experiments on RNA structures in interconformational conversion equilibria that revealed an exponential dependence between the number of base pairs required to

dissociate and rates of interconversion.[177] Calculated interconversion rates between PA, P^{A}T (I and M), and PA-TH were used as input for the transition matrix for Markovian simulations.

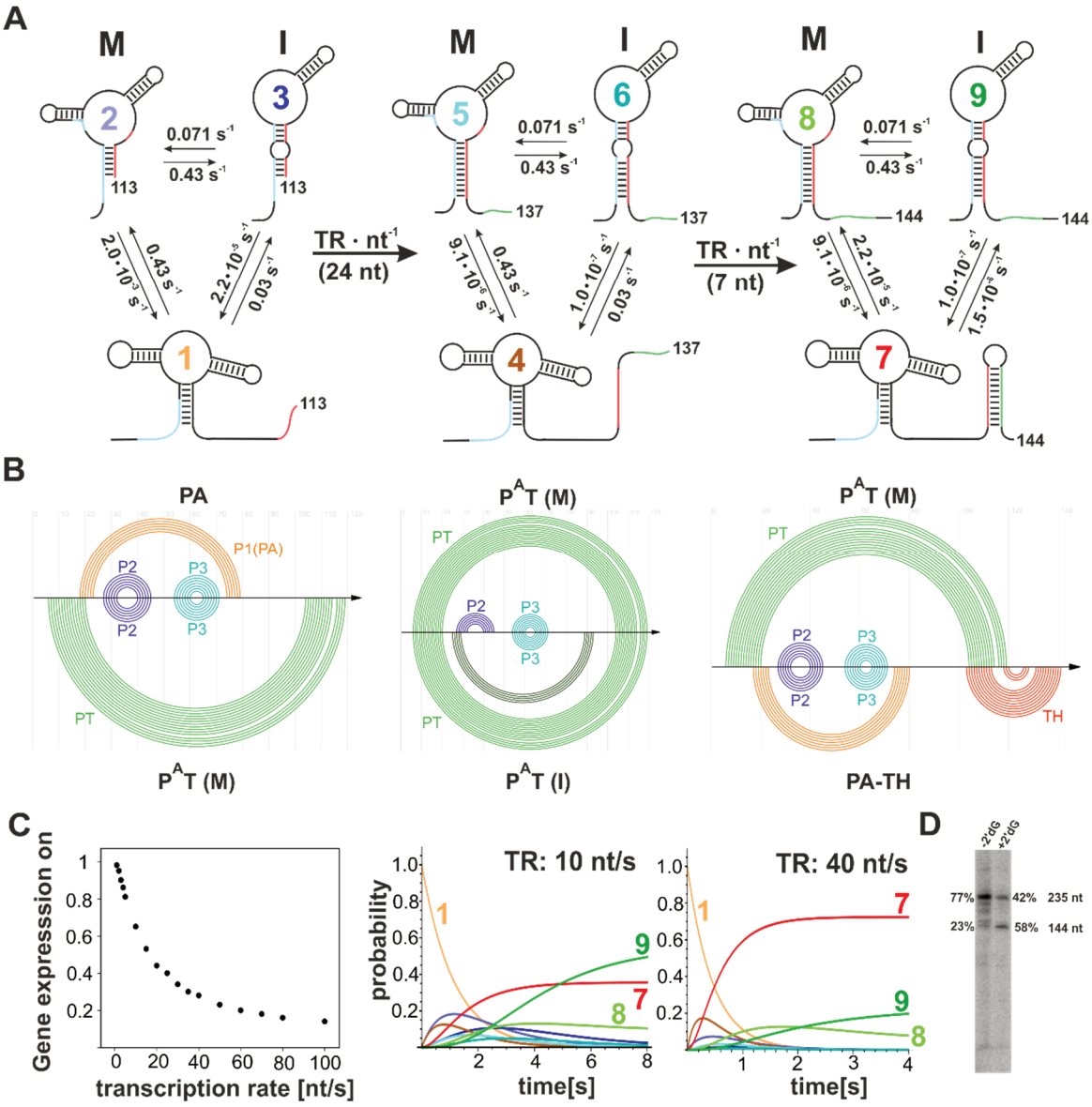


Figure 24: A) Theoretical model for co-transcriptional folding in the absence of ligand including interconversion rates between folding intermediates and the rate of transcription. B) Arc diagram of RNA secondary structure[178] visualizing the changes in base pairing for the structural rearrangement from aptamer (PA) to the antiterminator conformation P^{A}T (M), from P^{A}T (M) to P^{A}T (I) and from P^{A}T (M) to PA-TH. C) Graphical illustration of the fraction of ON-state obtained at the regulatory decision point depending on the rate of transcription including two exemplary simulations at distinct transcription rates (10 and 40 nt/s). D) Multi-round transcription assay of dGsw[235] with *E.coli* polymerase. The 235 nt transcript corresponds to the run-off transcript and the 144 nt transcript to the fraction of terminated transcript. Individual percentages of abortion and full-length fragment were quantified with the software ImageJ and corrected for the amount of U residues within the two fragments.

Changes in base pairing induced by conformational transitions are illustrated in the arc diagrams shown in Figure 24B. According to these secondary structure plots, the transition from PA to $P^{A}T$ (M) requires dissociation of three base pairs within helix P1 and should be fast with a rate of ~0.43 s^{-1}. Folding to $P^{A}T$ (I) from $P^{A}T$ (M) requires three base pairs within helix P2 and should be similarly fast. In contrast, direct folding from PA to $P^{A}T$ (I) requires dissociation of both helix P1 and helix P2. Therefore, calculated transition rates are slow with 0.03 s^{-1}. Consequently, rate calculations suggest fast folding from PA to $P^{A}T$ (M), followed by equilibration to $P^{A}T$ (I). Further, once either of the antiterminator conformations $P^{A}T$ is adopted, refolding to PA-TH during transcription of nts 137 and 144 should be slow with rates of $<10^{-5}$ s^{-1}, since it requires complete dissociation of the PT helix (15 bp for $P^{A}T$ (M) and 20 bp for $P^{A}T$ (I)).

Results of Markovian simulations of co-transcriptional folding are shown in Figure 24C. Since rates of transcription in the bacterium *M. florum* have not been identified yet, transcription rates were estimated to 10-90 nt/s, as found for bacterial polymerases.[9–11] Figure 24C shows a graphical illustration of the fraction of ON-state obtained at the regulatory decision point depending on the rate of transcription. In addition, two exemplary simulations performed at transcription rates of 10 nt/s and 40 nt/s are shown. Slow transcription rates coincide with a larger population of ON-state obtained at the regulatory decision point due to the increase in time required by the polymerase to synthesize nts 113-137. Consequently, an increase in the rate of transcription decreases the temporal window of opportunity for folding from PA to $P^{A}T$ and decreases the fraction of ON-state obtained at the regulatory decision point. Even at slow transcription rates of 10 nt/s the antiterminator conformation $P^{A}T$ (I) does not completely equilibrate within the time frame of transcription (light green line). After transcription of nts 113 to 144 (3.1 s), it only folds to ~35% compared to its equilibrium value. At a transcription rate of 40 nt/s, only $P^{A}T$ (M) is adopted after the time frame of transcription (0.8 s). Transcription termination assays performed with *E. coli* RNA polymerase show that approximately 70%-80% of all transcripts progress past the point of regulatory decision in the absence of ligand *in vitro* (Figure 24D). Comparison to our simulations would suggest a transcription rate of 8-12 nt/s, corresponding to the lower end of reported rates for bacterial polymerases. These theoretical calculations and simulations are further compared to experimental refolding rates from PA to $P^{A}T$ determined by real-time NMR in section 4.4.

4.3. OFF function (+ 2'dG)

In this section, transcriptional intermediates were investigated with respect to their ligand binding properties. Ligand binding is expected to trap the PA interaction within the aptamer domain to prevent antiterminator folding $P^{A}T$ during synthesis of $dGsw^{113}$-$dGsw^{137}$, as determined in the previous section. In analogy to structural profiling studies performed in the absence of ligand (section 4.2.1), the distinct window of opportunity for ligand binding was mapped, investigating both the aptamer length required to bind 2'dG and if the degree of 2'dG binding is altered as the RNA is extended by single nucleotides until transcript length 113. The extent of ligand binding by transcriptional intermediates was investigated by NMR for construct lengths highlighted in blue in Figure 25 (section 4.3.1). In particular, ligand binding was monitored during synthesis of P1 (nts 75-80), between nts 80-112 (PA fold), nts 114 to 137 (PT fold) and nts 135 to 144 (PA-TH fold). Further, NMR experiments were supplemented by ITC equilibrium K_D measurements of eight selected transcriptional intermediates highlighted in cyan circles in Figure 25 (section 5.3.2).

According to findings described in Chapter I, ligand binding never reaches equilibrium during transcription and ligand binding is accelerated by cellular ligand concentrations that exceed K_D values by 3 orders of magnitude.[3,106] Ligand binding kinetics were investigated by stopped-flow fluorescence spectroscopy to investigate if kinetics change as the RNA is elongated by single nucleotides for 10 transcriptional intermediates, selected according to results presented in sections 4.3.1 and 4.3.2, and highlighted in red circles in Figure 25 (section 4.3.3).

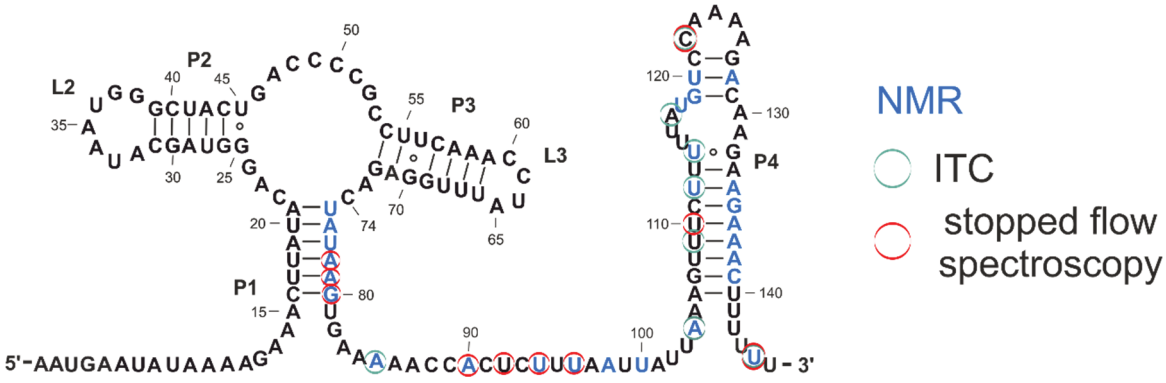


Figure 25: Secondary structure of $dGsw^{FL}$ illustrating intermediate sequences investigated with respect to their ligand binding affinity by NMR-spectroscopy (blue) and ITC (cyan circles), and their ligand binding kinetics by stopped-flow fluorescence (red circles).

4.3.1 Ligand-binding efficiency of transcriptional intermediates by NMR

Ligand binding efficiencies by transcriptional intermediates were assessed by ^{15}N-labeling of 2'dG to allow filtering of the imino proton signal of 2'dG from signals corresponding to unlabeled riboswitch RNA. Since imino protons can be detected only if they are protected from solvent exchange, the signal intensity of the imino proton of 2'dG is directly related to the fraction of ligand bound to the RNA. However, elongation of the RNA is accompanied by an increase in size and successive signal broadening. Therefore, relative 2'dG H1 signal intensities yield distorted results, and to avoid this issue, the 2'dG H1 signal was integrated for the various transcriptional intermediates (Figure 26A). Ligand binding to transcriptional intermediates was investigated at different Mg^{2+} concentrations ranging from 2 mM - 5 mM (2 - 10 eq.) (Figure 26B)

Ligand binding is most pronounced for constructs $dGsw^{80}$-$dGsw^{93}$ that contain a fully synthesized aptamer. Between $dGsw^{93}$ and $dGsw^{112}$ ligand binding continuously decreases by 30% despite structural profiling studies of transcriptional intermediates in the absence of ligand suggesting the aptamer domain to represent the only conformation adopted >10% within this sequential segment. Possibly, RNA elongation beyond the aptamer domain expands the conformational landscape, and multiple states populated below 10% form within the structural ensemble, which cannot be detected by NMR. Elongation of the RNA by two additional nucleotides to $dGsw^{114}$ leads to a further decrease in the ligand binding efficiency by 20%. Ligand addition to constructs that adopt the antiterminator conformations $P^{A}T$ at thermodynamic equilibrium in the absence of ligand ($dGsw^{114}$-$dGsw^{137}$) induces a conformational switch to PA by ~20-40%, with the ligand-bound population reaching a low of 20% once the PT helix is completely synthesized. Continuous stabilization of the terminator helix TH between $dGsw^{134}$-$dGsw^{144}$ leads to an increase in the ligand binding efficiency; however only up to 70% in $dGsw^{FL}$ compared to the isolated aptamer domain.

Effects of Mg^{2+} on ligand binding are strongest for constructs with a truncated P1 helix. For these constructs, stabilizations by Mg^{2+} to induce a parallel alignment of helices P2 and P3 likely compensates for destabilizations of these tertiary interactions caused by P1 truncation. Once the P1 helix is completely formed, the effect of Mg^{2+} on ligand binding becomes insignificant. Only the ligand-induced switch from the antiterminator helix PT to the aptamer domain (PA) can be increased by ~10% by addition of Mg^{2+}. In section 4.2.1 it was shown that Mg^{2+}-binding alone is insufficient to induce a conformational

rearrangement from PT to PA above the detection limit of ~10%; however, in combination with 2'dG, Mg^{2+} appears to capture ligand binding sufficiently to cause a shift in equilibrium between PT and PA.

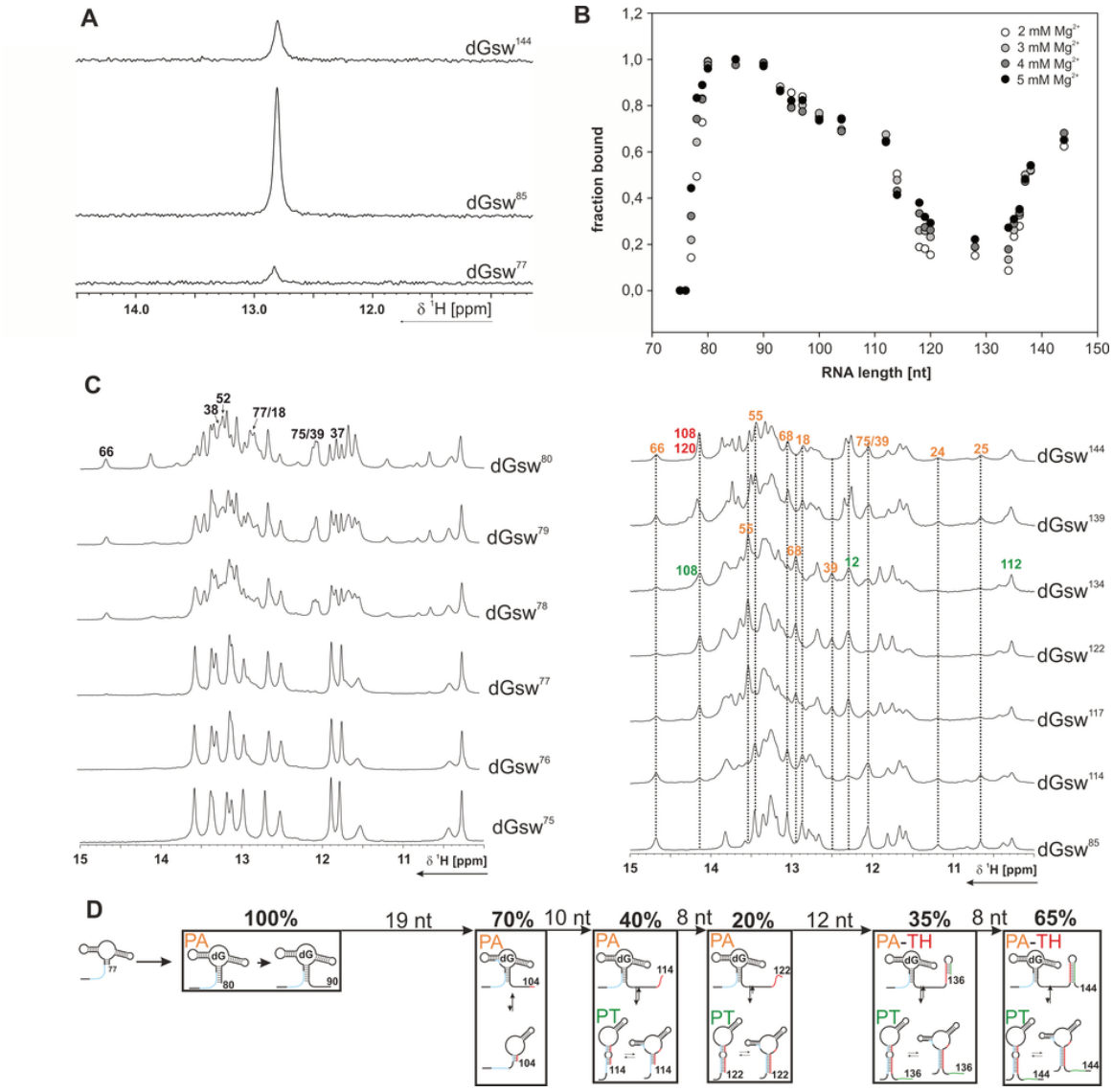


Figure 26. NMR analysis of ligand binding to transcriptional intermediates. A) Exemplary 1D ^{15}N-X-filtered 1D spectra filtering the imino proton of 2'dG bound to dGsw for dGsw^{77}, dGsw^{85} and dGsw^{144}. Spectra were recorded on 500 µM RNA samples in the presence of 1 mM of 2'dG and 2 mM of Mg^{2+} at 600 MHz and 298 K. B) Relative integrals of 2'dG imino proton signals shown in A) depending on the transcript lengths at Mg^{2+} concentration of 2 mM (white), 3 mM (light grey), 4 mM (dark grey) and 4 mM (black). C) 1D NMR spectra of transcriptional intermediates in the presence of 2 eq. of 2'dG at 3 mM Mg^{2+} recorded for 14 exemplary transcriptional intermediates at 600 MHz and 298 K. Aptamer reporter signals (PA) are highlighted in orange, antiterminator reporter signals (PT) in green and terminator reporter (TH) signals in red D) Graphical representation of secondary structures adopted at different transcript length according to results shown in B).

Figure 26C (left) shows a more detailed 1D NMR analysis of ligand binding to aptamer constructs with a truncated P1. While P1 formation cannot be detected prior to dGsw^{80} in the absence of ligand, 2'dG binds to the aptamer domain at a length of 78 nt. This result suggests that a P1 helix consisting of four base pairs is required to enable ligand binding. Segment P required to form the antiterminator helix PT contains the three lower base pairs in P1, and once formed, leaves a 3 base pair P1 helix. Such a P1 helix would be too instable to form by one nucleotide, further demonstrating the strict fine-tuning of conformation equilibria between mutually exclusive conformation in riboswitches. I-B and III-B *mfl* riboswitches exhibit the same pattern regarding the number of nucleotides involved in P1 helix formation in strand P (Figure 15, Chapter 3.1).

4.3.2 Ligand binding affinity of transcriptional intermediates by ITC

ITC measurements to obtain equilibrium K_D values were performed for eight selected constructs representing the fully synthesized aptamer domain with maximum ligand binding efficiency (dGsw^{85}), constructs that adopt the aptamer domain according to NMR analysis but exhibit reduced ligand binding affinity (dGsw^{104}, dGsw^{109}), PA-PT intersection constructs (dGsw^{112}, dGsw^{114}) and the full-length riboswitch with a fully formed terminator hairpin (dGsw^{FL}) (Figure 25). The results of ITC measurements for these constructs are shown in Figure 27 and their derived thermodynamic parameters are listed in Table 4. The lowest K_D value of 250 nM was obtained for the aptamer domain (dGsw^{85}), which also exhibits the highest ligand binding efficiency according to NMR. The K_D for 2'dG obtained from these ITC measurements for the aptamer domain is ~3 fold higher than the reported K_D derived from in-line probing experiments of 80 nM.[1] Differences in K_D values can be explained by changes in the experimental conditions, such as the much higher excess of Mg^{2+} over RNA in in-line probing experiments (20 mM Mg^{2+}:2 nM RNA compared to 3 mM Mg^{2+}:10 µM RNA).

Aptamer constructs that contain parts of strand T (dGsw^{104}, dGsw^{109}) exhibit slightly reduced apparent K_D values of 310/320 nM. The decrease in K_D from dGsw^{85} to dGsw^{109} is relatively modest compared to the decrease in ligand binding efficiency of 20-30%, as determined by NMR. ITC experiments directly measure thermodynamics of 2'dG binding to binding competent species within the structural ensemble. Misfolded species that do not bind 2'dG lead to an overestimation in the RNA concentration provided in the fit. However,

concentration errors in the ITC cell (RNA) have a limited effect on determined K_D values, while K_D values are largely affected by errors in the syringe concentration (2'dG). In contrast, NMR measures ligand binding as a combination of binding competent and incompetent species. Therefore, combining these ITC results with NMR results reveals the presence of misfolded structures within the structural ensemble.

Within the two constructs framing the conformational transition from aptamer (PA) to the antiterminator helix PT ($dGsw^{112}$ and $dGsw^{114}$), the K_D decreases from 370 nM to 3.6 µM. In $dGsw^{114}$, the aptamer domain (PA) is populated <10%, and therefore ITC measurements are no longer biased towards 2'dG binding to binding competent species and directly reflect the K_D of 2'dG binding to $dGsw^{114}$ with a partially formed PT helix. Once the antiterminator helix PT is fully stabilized ($dGsw^{117}$, $dGsw^{122}$), K_D values further decrease significantly to >10 µM. The course of induced heat in relation to the amount of injected ligand does not contain an inflection point and therefore the K_D cannot be reliably derived for these two constructs. $dGsw^{FL}$ exhibits a K_D value of 370 nM, which is identical to the K_D value determined for $dGsw^{112}$, and therefore in line with NMR experiments.

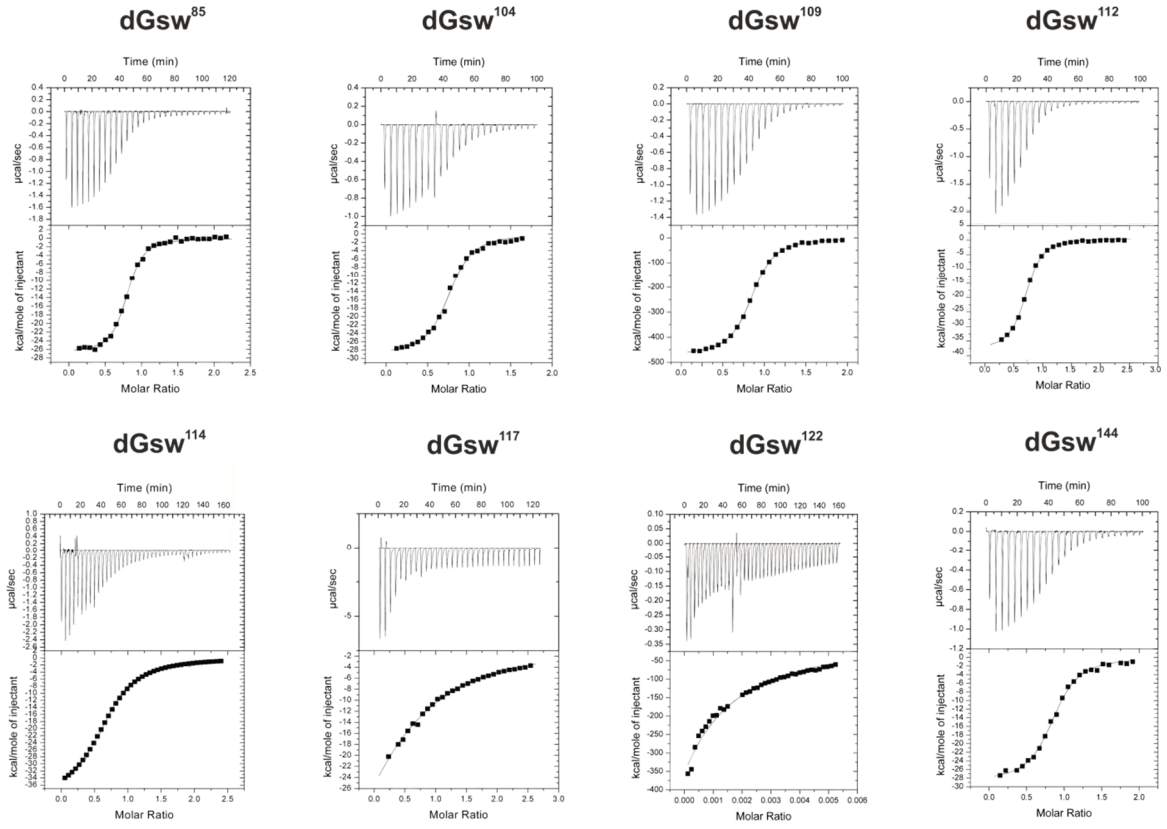


Figure 27: ITC curves of $dGsw^{85}$ (12 µM), $dGsw^{104}$ (15 µM), $dGsw^{109}$ (15 µM), $dGsw^{112}$ (15 µM), $dGsw^{114}$ (40 µM), $dGsw^{117}$ (80 µM), $dGsw^{122}$ (80 µM) and dGsw144 (20 µM) titrated with 2'dG (20x excess in concentration over RNA). Upper graphs display the change in heat induced by each injection with ligand. Lower graphs display integrations of heater power including fits to a standard single-site binding model.

Table 4: Thermodynamic parameters K_D and ΔG obtained from ITC measurements of transcriptional intermediates. Errors were calculated based on error propagations of standard deviations.

Construct	dGsw^{85}	dGsw^{104}	dGsw^{109}	dGsw^{112}	dGsw^{114}	dGsw^{117}	dGsw^{122}	dGsw^{144}
K_D [μM]	0.25 ± 0.02	0.32 ± 0.03	0.31 ± 0.01	0.37 ± 0.02	3.6 ± 0.2	> 10	> 10	0.37 ± 0.03
ΔG [kcal mol^{-1}]	-9.0 ± 0.2	-8.9 ± 0.4	-8.9 ± 0.1	-8.9 ± 0.4	-7.5 ± 0.8	-	-	-8.8 ± 0.3

Stoichiometry values (N) vary from 0.7 to 0.9. The deviation from the ideal value of 1.0 can be explained by errors in the RNA concentration. Accurate determinations of the binding stoichiometry require exact concentrations resembling only active conformations of the RNA provided in the fit, which would demand more time-consuming procedures for concentration determinations, such as RNA hydrolysis followed by UV measurements. Further, RNA constructs were prepared by the rapid screening approach described in Chapter 3.2,[8] leading to a slight overestimation in the RNA concentration and consequently lower stoichiometry values. However, as described above, errors in the ITC cell concentration have a relatively small impact on K_D values. The formation of a 1:1 complex of 2'dG with a properly folded aptamer domain was proven by NMR experiments performed in Chapter 3.1 and 4.3.1.

4.3.3 Ligand binding kinetics of transcriptional intermediates by stopped-flow spectroscopy

Analysis of ligand binding kinetics for a multitude of transcriptional intermediates was performed by stopped-flow fluorescence spectroscopy. To monitor changes in fluorescence upon ligand binding, the Watson-Crick base pairing C74-2'dG was mutated to U74 to allow binding of fluorescent 2-aminopurine 2'-deoxyriboside (2'dAP). The fluorescence of the nucleobase 2-aminopurine (2AP) is strongly quenched upon binding by changing environments, such as stacking upon adjacent nucleotides.[116,179–181] According to the base pairing architecture derived from the published crystal structure of the I-A aptamer domain (Figure 28A),[127] 2'dAP is expected to bind well to the C-74-U mutant of dGsw (dGsw^{C74U}). The base pairing pattern of the ligand with C51 is not altered by the mutation. However, a decline in ligand affinity is expected due to the weaker nature of the A-U base pair in dGsw^{C74U} compared to the G-C base pair in dGsw. Previous studies report a C74-to-U74 mutated guanine riboswitch to readily accept adenine as a ligand, although 2,6-DAP binds more tightly to these aptamers compared to adenine.[78,92] For the I-A dGsw, it is reported

that dGsw^{C74U} does not bind 2'-dA (K_D > 1 mM) but binds 2'-d-2,6-DAP with a K_D of 8 µM, emphasizing the importance of hydrogen bond interactions of the exocyclic amine group in both G and 2,6-DAP with C51.[1]

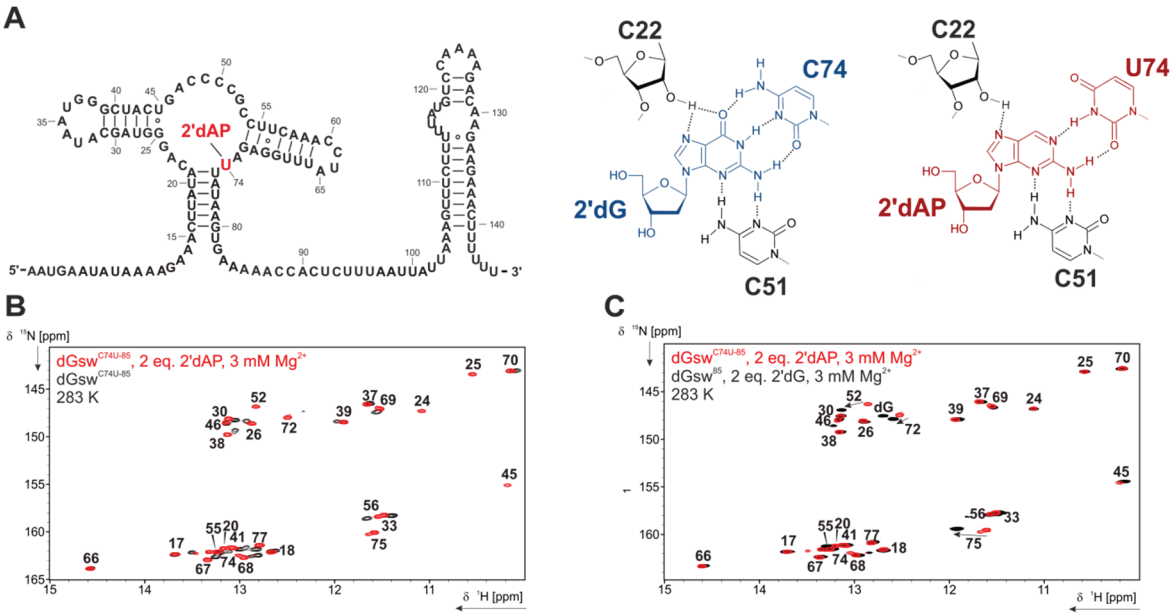


Figure 28: A) Secondary structure of dGsw^{C74U} highlighting the C to U mutation at position 74 in red and comparison of the base pairing pattern in the binding pocket of dGsw bound to 2'dG^{127} with dGsw^{C74U} bound to 2'dAP. B) Overlay of ^{15}N-TROSY spectra of dGsw$^{C74U-85}$ recorded in the absence (black) and presence of 2 eq. of 2'dAP (red) and 3 mM Mg^{2+} at 800 MHz and 298 K. C) Overlay of ^{15}N-TROSY spectra of dGsw$^{C74U-85}$ (red) and dGsw^{85} (black) recorded at 800 MHz and 283 K. Arrows indicate chemical shift perturbations caused by the mutation in the binding pocket.

Prior to kinetic measurements, the integrity of the ligand binding pocket was investigated by NMR-spectroscopy. The overlay of ^{15}N-TROSY spectra of dGsw^{C74U} recorded in the absence and presence of ligand at 3 mM Mg^{2+} (Figure 28B) shows that 2'dAP binds to dGsw^{C74U} and forms a homogeneous ligand-bound state, suggesting a K_D of <100 µM. Ligand binding leads to the appearance of reporter signals G24, G25, G52 and U66, also detected in dGsw. In Figure 28C, the conformation of 2'dAP-bound dGsw$^{C74U-85}$ (red) is directly compared to 2'dG-bound dGsw^{85}. Mutations in the binding pocket lead to chemical shift perturbations of G52, G72 and U75. An additional Uracil signal appears at ~13.3 ppm, which likely corresponds to the mutation site U74. The remaining signals are virtually identical in dGsw, demonstrating homology in the global fold.

Kinetic stopped-flow experiments were performed on eight selected constructs: dGsw$^{C74U-78}$ and dGsw$^{C74U-79}$ both contain a truncated P1 helix, and according to NMR experiments represent the first two constructs able to bind the ligand; dGsw$^{C74U-80}$,

dGsw$^{C74U-85}$, dGsw$^{C74U-90}$, dGsw$^{C74U-94}$, dGsw$^{C74U-96}$, dGsw$^{C74U-100}$, dGsw$^{C74U-110}$ form the aptamer domain (PA) but NMR experiments show a continuous decrease in ligand binding efficiency; dGsw$^{C74U-122}$ was investigated as a reference construct to fully adopt the antiterminator conformation $P^{A}T$. k_{on} and k_{off} rates were obtained by varying the 2'dAP concentration to keep the Mg^{2+} over RNA ratio constant. Experiments were performed at an RNA concentration of 1 µM while an excess of 2'dAP (16, 32, 48, 64, 96, and 128 µM) was applied to guarantee pseudo first order conditions. By applying an excess of ligand over RNA, a potential pre-equilibrium in the riboswitch RNA can be monitored. If the RNA was applied in excess and ligand binding competent species were populated >10%, only binding to these species would be observed.

Figure 29A-F shows kinetic traces recorded at 24 eq. of 2'dAP added to dGsw^{C74U} constructs at varying lengths. Constructs with a fully synthesized aptamer domain quench fluorescence by approximately 2% in the presence of 24 eq. of 2'dAP (Figure 29C-D). Due to the low proportion of fluorescence quenched and the corresponding high error, a correlation between the extent of fluorescence quenched and the population of ligand-bound state cannot be reliably derived from these experiments. However, fluorescence quenching is significantly decreased for dGsw$^{C74U-78}$, dGsw$^{C74U-79}$ and dGsw$^{C74U-122}$ to 0.4%, 0.6% and 0.2%, respectively (Figure 29A,B,F). For dGsw$^{C74U-78}$ and dGsw$^{C74U-79}$, the decrease in fluorescence quenching may be caused by incomplete stacking in P1. While ligand binding requires four base pairs in P1, by NMR a stable P1 helix can be detected only from dGsw^{80} on. For dGsw$^{C74U-122}$, the degree of fluorescence quenching can be directly correlated to its low ligand-bound population, in line with NMR experiments (20%). Kinetic traces for RNA constructs beyond the length of 80 nt are well described by single exponential decays. The exponential decay of constructs with a truncated P1 helix (dGsw$^{C74U-78}$ and dGsw$^{C74U-79}$) is better described by bi-exponential function fits. The bi-exponential fit includes a rate that resembles ligand-binding rates obtained for the larger constructs and a second rate that is one order of magnitude slower. The second rate possibly describes the compact formation of the ligand binding pocket including tertiary loop-loop interactions, which may occur on a slower timescale due to truncations in helix P1, while the initial encounter complex formation is equally fast for these constructs.

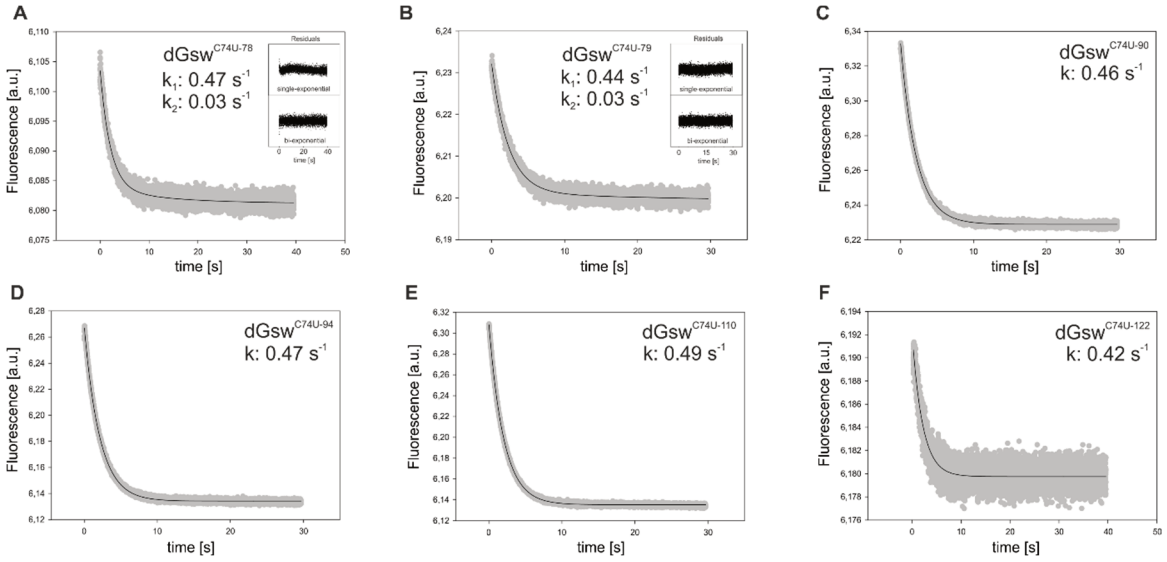


Figure 29: Kinetic traces of 1 µM dGsw$^{C74U-78}$ (A), dGsw$^{C74U-79}$ (B), dGsw$^{C74U-90}$ (C), dGsw$^{C74U-94}$ (D), dGsw$^{C74U-110}$ (E) and dGsw$^{C74U-122}$ (F) mixed with 24 eq. of 2'dAP. (A) and (B) show inserted windows with residuals depending on time to compare single- and bi-exponential fits.

Table 5 summarizes all obtained apparent rate constants at different ligand concentrations and derived k_{on}, k_{off} and K_D values. Figure 30A shows exemplary linear fits to obtain k_{on} and k_{off} for dGsw$^{C74U-79}$ and dGsw$^{C74U-94}$. In Figure 30B, the dependence of k_{on} and k_{off} on the transcript lengths is shown. Ligand association rates k_{on} can be considered to be in a similar range for aptamer constructs dGsw$^{C74U-80}$-dGsw$^{C74U-110}$, while dGsw$^{C74U-78}$, dGsw$^{C74U-79}$ and dGsw$^{C74U-122}$ exhibit reduced k_{on} rates. For dGsw$^{C74U-78}$ and dGsw$^{C74U-79}$, association rates are reduced by a factor of five. In the context of transcription, ligand recognition may be initiated while the P1 helix is not fully transcribed. However, the proposed refolding rate of 0.03 s^{-1} to form a compact binding pocket involving all related tertiary interactions is on a slower timescale than the transcription of two additional nucleotides to dGsw$^{C74U-80}$. dGsw^{122} exhibits only a ~2-fold reduced association rate compared to all aptamer folded constructs. However, determined kinetics represent binding to the low population of binding competent species adopted at a length of nt 122, which is also reflected in the low signal to noise of the kinetic experiments.

K_D values determined by stopped-flow fluorescence are reduced by two orders of magnitude in comparison to K_D values obtained from ITC and by one order of magnitude compared to binding of 2'-d-2,6-DAP to dGsw^{C74U}. A general decrease in the K_D, caused by the switch from G-C to A-U base pair, is expected. Therefore, determined rates of ligand association may be underestimated by these experiments. However, considering the much larger excess of ligand over RNA in the cell, ligand binding rates likely exceed 1 s^{-1}, as

determined at 128 eq. of 2'dAP. Further, stopped-flow experiments show that, while populations change as the RNA is elongated by single nucleotides, kinetics of ligand binding to binding competent species within the structural ensemble do not change. The data demonstrate that all conformations that adopt the aptamer domain conformation (PA) exhibit similar ligand binding kinetics. Therefore, the amount of transcripts that are trapped in the OFF-state during co-transcriptional ligand binding is not regulated by a change in ligand binding kinetics but by a shift in the population of the folded aptamer domain.

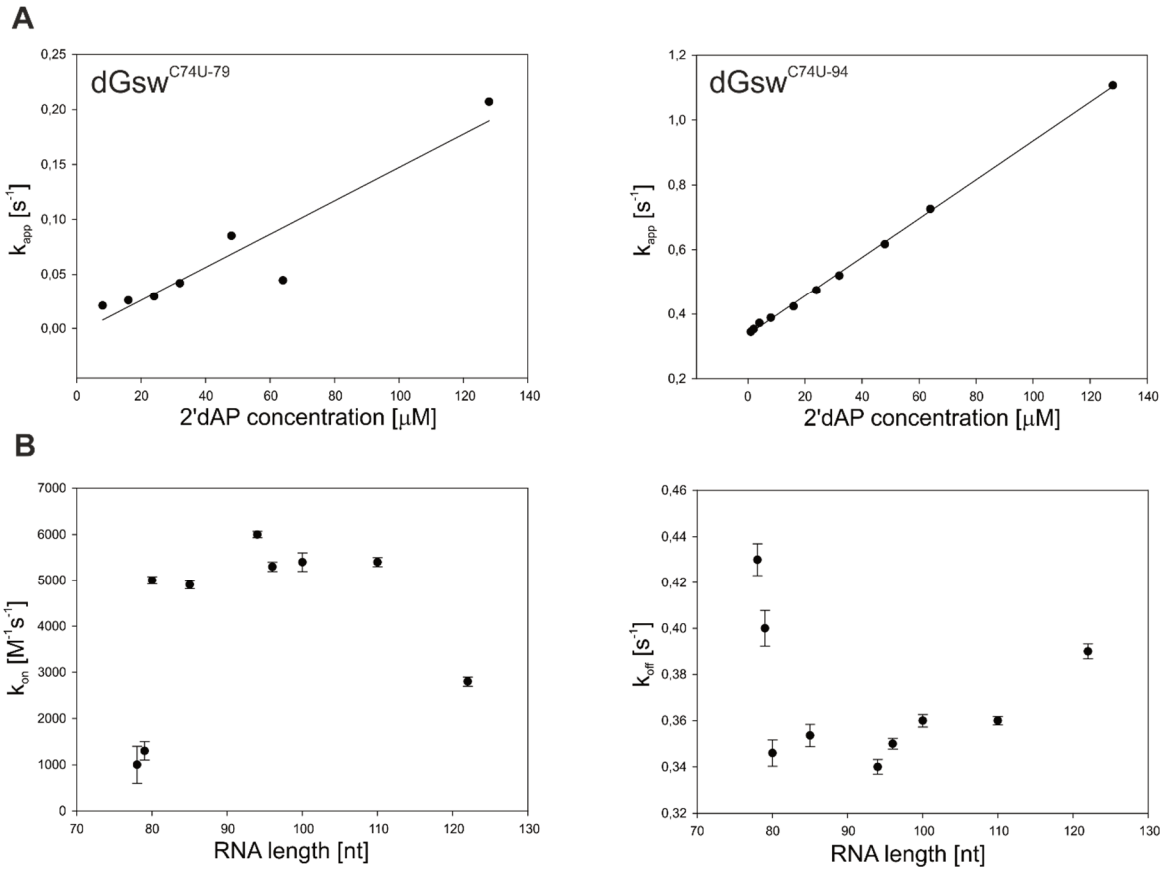


Figure 30: A) Exemplary linear fits of rate constants under pseudo-first order conditions to obtain k_{on} and k_{off} according to $k_{app} = k_{on} * [2'dAP] + k_{off}$ for dGsw$^{C74U-79}$ and dGsw$^{C74U-94}$. B) k_{on} (left) and k_{off} (right) rates of 2'dAP binding to dGsw^{C74U} at varying transcript lengths.

Table 5: Apparent rate constants k_{app} for selected dGsw^{C74U} constructs (1µM) at six different ligand concentrations determined by single ($y(t) = y_0 + A * e^{k_{app}t}$) and bi-exponential decay ($y(t) = y_0 + A * e^{k_{1,app}t} + B * e^{k_{2,app}t}$) functions. k_{on}, k_{off} were determined as described in Figure 29 and K_D values were determined according to $K_D = \frac{k_{off}}{k_{on}}$.

	k_{app} [s^{-1}]						k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [µM]
[2'dAP] →	16 µM	24 µM	32 µM	48 µM	64 µM	128 µM			
$dGsw^{C74U-78}$	0.436 0.016	0.477 0.080	0.492 0.028	0.476 0.047	0.502 0.211	0.54 -	$1.0*10^3$	0.43	430.0
$dGsw^{C74U-79}$	0.412 0.026	0.436 0.029	0.442 0.042	0.486 0.085	0.483 0.045	0.552 0.207	$1.3*10^3$	0.40	307.7
$dGsw^{C74U-80}$	0.429	0.466	0.503	0.583	0.671	0.967	$5.0*10^3$	0.35	70.0
$dGsw^{C74U-85}$	0.373	0.393	0.434	0.518	0.685	0.948	$4.7*10^3$	0.36	76.6
$dGsw^{C74U-90}$	0.445	0.481	0.525	0.603	0.688	0.922	$8.9*10^3$	0.37	35.6
$dGsw^{C74U-94}$	0.423	0.474	0.520	0.617	0.726	1.107	$6.0*10^3$	0.34	56.6
$dGsw^{C74U-96}$	0.446	0.466	0.520	0.602	0.696	0.968	$5.3*10^3$	0.35	66.0
$dGsw^{C74U-100}$	0.441	0.485	0.527	0.614	0.707	0.945	$5.4*10^3$	0.36	66.7
$dGsw^{C74U-110}$	0.448	0.498	0.538	0.621	0.705	0.995	$5.4*10^3$	0.36	66.7
$dGsw^{C74U-122}$	0.480	0.428	0.420	0.477	0.571	0.936	$5.5*10^3$	0.23	41.8

4.4. Kinetics of antiterminator folding monitored by real-time NMR

4.4.1 Strategy and introduction

Markovian simulations of co-transcriptional folding based on calculated refolding rates between PA, PT and PA-TH, as presented in section 4.2.2, roughly estimate how folding intermediates can interconvert during transcription. In this section, a strategy was developed to allow monitoring of the folding event from aptamer (PA) to the antiterminator helix PT by real-time NMR. Kinetics of antiterminator folding were then compared to calculated rates used in the previous model in section 4.4.5. In addition, previously performed simulations were only performed in the absence of ligand, since the effect of ligand binding could not be reliably included in theoretical rate calculations. The experimental setup allows also determination of refolding rates from PA to PT in the presence of ligand, where PA is stabilized.

To monitor the formation of the PT helix in real time, nucleotides within strand T were protected from forming canonical base pairs with strand P by photolabile protection groups (Figure 31A). These protection groups can be removed by laser irradiation at 355 nm within the NMR spectrometer to initiate folding from PA to PT.[182–186] Due to the relatively large

size of dGsw^{ON} (122 nt), photocaged dGsw^{121} was prepared by enzymatic splinted ligation. The RNA was dissected into the following two fragments: dGsw^{85}, which forms the aptamer domain (blue segment in Figure 31B), and dGsw$^{86-121}$, which represents part of the expression platform including strand T (red segment in Figure 31B).

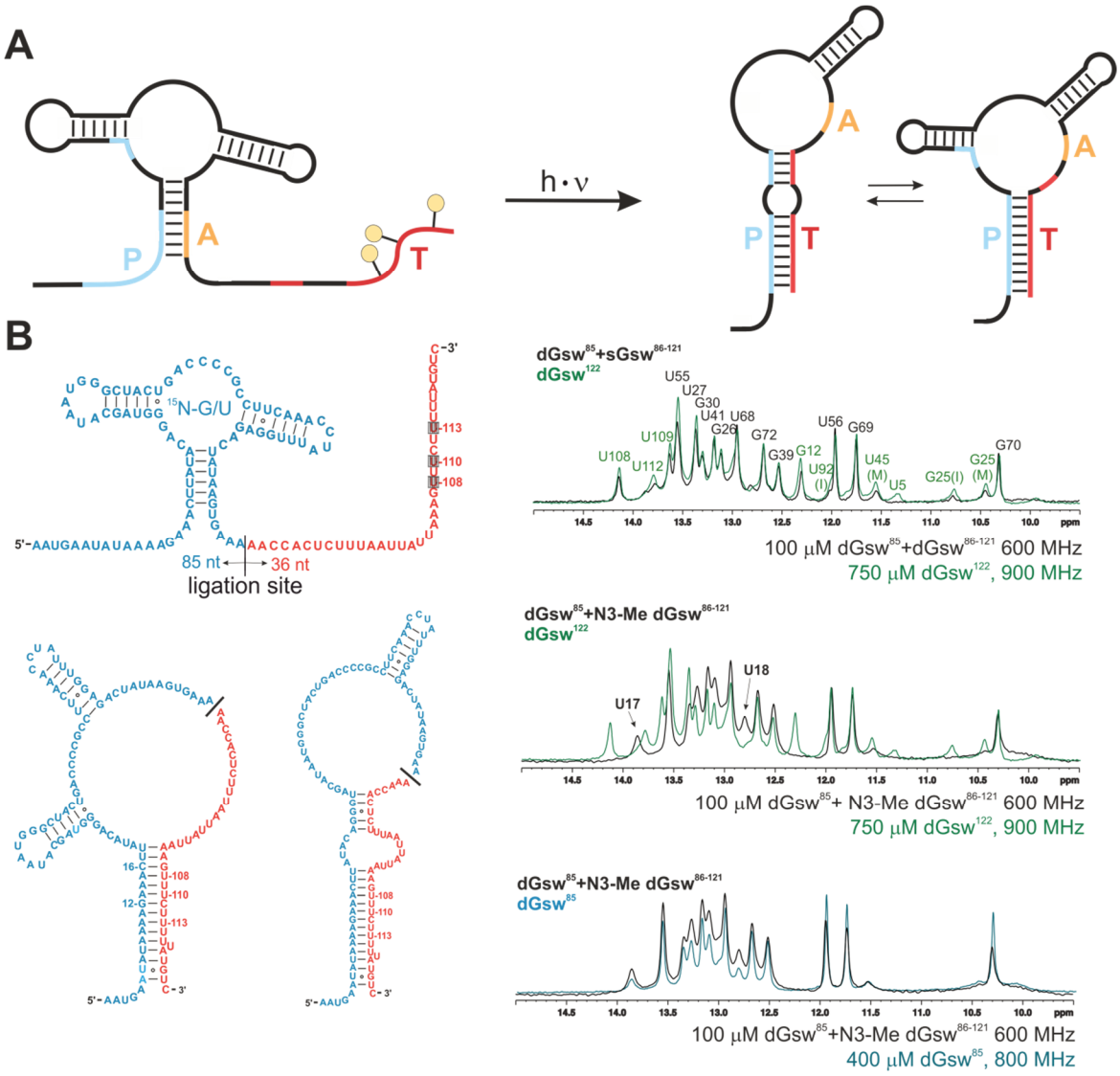


Figure 31: A) Schematic representation of the strategy applied to monitor antiterminator P^{A}T folding by real-time NMR. Strand T is protected by three photolabile protection groups (illustrated by yellow circles). Light-induced cleavage of these photolabile protection groups triggers refolding from PA to PT. B) Design of photocaged dGsw^{121}, composed of ^{15}N-G/U labeled dGsw^{85} obtained from *in vitro* transcription and dGsw$^{86-121}$ modified at Uridine residues 108, 110, and 113. 1D NMR spectral overlays of the imino proton region show a comparison of intact dGsw^{122} (in green) and dGsw^{85} (in blue) with assembled dGsw^{121} from fragments dGsw^{85}+dGsw$^{86-121}$ and dGsw^{85}+N3-methylated dGsw$^{85-122}$ (in black). Spectrometer frequencies and RNA concentrations are indicated for each spectrum and all spectra were recorded at 298 K.

dGsw^{85} was prepared with ^{15}N-labeled GTP and UTP by *in vitro* transcription to allow higher resolution NMR analysis. dGsw$^{86\text{-}121}$ was prepared by solid-phase synthesis to allow implementation of the photolabile protection groups. The smaller fragment dGsw$^{86\text{-}121}$ contains 22 free nucleotides 5' of the photocaged residues to enable DNA splint annealing. Due to the time-consuming and costly procedure of preparing photocaged dGsw^{121} in NMR scale, the approach was first cross validated by N3-methylation of three Uridine residues within strand T (U108, U110 and U113) (Figure 31B). Modified residues were chosen based on helical segment stability. According to PT imino proton signal intensities in dGsw^{ON}, the helical segment G107-C111:G12-C16 represents the most stable segment within the antiterminator helix PT and interference in its base pairing is most likely to inhibit antiterminator formation (P^AT). N3-methylated dGsw$^{86\text{-}121}$ was obtained commercially and tested for its association affinity to dGsw^{85} in comparison to unmodified dGsw$^{86\text{-}121}$. This initial validation was performed by adding the two fragments together in a 1:1 ratio without further ligation. 1D spectra in Figure 31B show that unmodified dGsw$^{86\text{-}121}$ hybridizes to dGsw^{85} to form the antiterminator helix PT, identified by the appearance of reporter signals highlighted in green. The imino proton pattern of assembled dGsw^{121} closely resembles the spectrum of intact dGsw^{ON}/dGsw^{122} with the exception of signal U5. U5 is located at the lower end of the PT helix, and its imino proton may be more accessible to solvent exchange due to destabilizations in the PT helix caused by the lack of covalent connection between the two fragments. In contrast, the 1D NMR spectrum of dGsw^{85} in the presence of N3-methylated dGsw$^{86\text{-}121}$ shows no indication for PT helix association and lacks corresponding reporter signals, while reporter signals for helix P1 (U17 and U18) can be detected. In this spectrum, only reporter signals corresponding to the aptamer domain (PA) can be identified. Alternative base pairing interactions both within N3-methylated dGsw$^{86\text{-}121}$ or between the two fragments cannot be observed. These results strongly suggest that the antiterminator conformation P^AT can be inhibited by protection of the Watson-Crick base pairing site within Uracil residues 108, 110 and 113.

4.4.2 Preparation of photocaged dGsw^{121} by enzymatic splinted ligation

Photocaged dGsw$^{86\text{-}121}$ (c-dGsw$^{86\text{-}121}$), containing three 1-(2-nitrophenyl)ethyl (NPE) protection groups at uridine residues 108, 110 and 113, was provided by Dean Klötzner (AK Heckel, Goethe University Frankfurt, Germany) (red segment in Figure 32A). dGsw^{85}

and c-dGsw$^{86-121}$ were pre-annealed to a 31 nt long DNA splint (splint^{31}) and ligated using T4 RNA ligase 2. Ligation reactions were optimized to obtain the largest turnover of c-dGsw$^{86-121}$, which was achieved by applying dGsw^{85} in 2-fold excess (Figure 32B-C). Optimal ligation conditions yield ~80% ligation product with respect to c-dGsw$^{86-121}$. The increased hydrophobicity of c-dGsw^{121} and c-dGsw$^{86-121}$ introduced by the three NPE modifications allows separation of all RNA fragments including splint^{31} by RP-HPLC (Figure 32C). In the RP-HPLC chromatogram shown in Figure 32C, AMP/ATP elute first at a retention time of 15 minutes, followed by dGsw^{85} and the splint^{31} at ~21 minutes. dGsw^{85} and the splint^{31} could not be completely separated as shown in the gel depicted in Figure 32C and were directly applied in a mixture for further ligation reactions. The desired product c-dGsw^{121} elutes at a retention time of 26 min, followed by elution of the smaller, but more hydrophobic, dGsw$^{86-121}$ at 29 min. After purification and buffer exchange, 0.6 µmol of c-dGsw^{122} were obtained (54%).

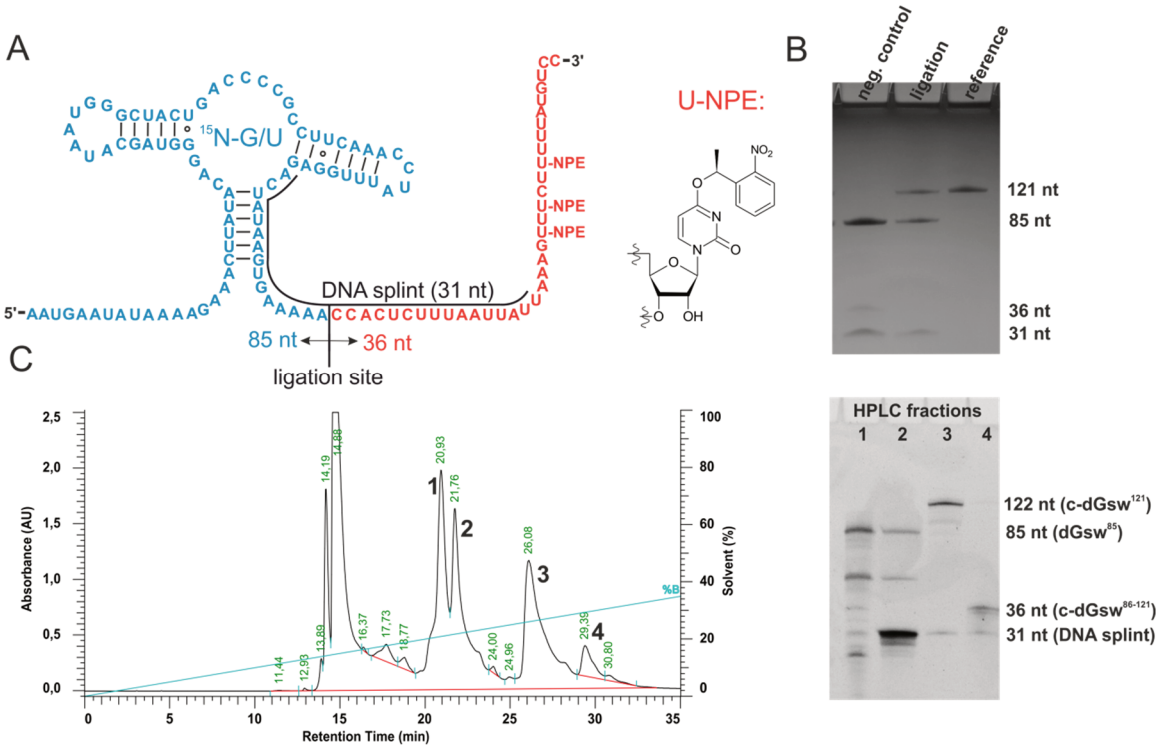


Figure 32: A) Strategy for preparation of c-dGsw^{121} by enzymatic splinted ligation. ^{15}N G/U labeled dGsw^{85} (blue segment) was annealed by a 31 nt long DNA splint to c-dGsw$^{86-121}$ (red segment) and ligated with T4 RNA ligase 2. B) 12% denaturing polyacrylamide gel of the ligation reaction under optimized conditions using a 2:1.6:1 ratio of dGsw^{85} : splint^{31} : dGsw$^{86-121}$. The gel shows an additional negative control and a reference lane loaded with intact dGsw^{122}. C) HPLC-chromatogram with a fixed wavelength at 260 nm (left). Solvent A: 0.1 M triethylammonium acetate, pH 6.0; Solvent B:Acetonitrile. The peak numbering refers to HPLC fractions loaded on the 12% denaturing polyacrylamide gel shown on the right.

4.4.3 Structural integrity of photocaged dGsw^{121} before and after laser irradiation

The structural integrity of c-dGsw^{121} was investigated by NMR to ensure homogeneous folding of the aptamer domain (PA) and inhibition of helix PT formation. To provide for efficient cleavage of the photolabile protection groups, the RNA concentration for kinetic experiments should not exceed ~100 µM. Therefore, the Mg^{2+} concentration was reduced to 600 µM to keep the Mg^{2+}:RNA ratio constant and comparable to all NMR experiments previously performed on dGsw.

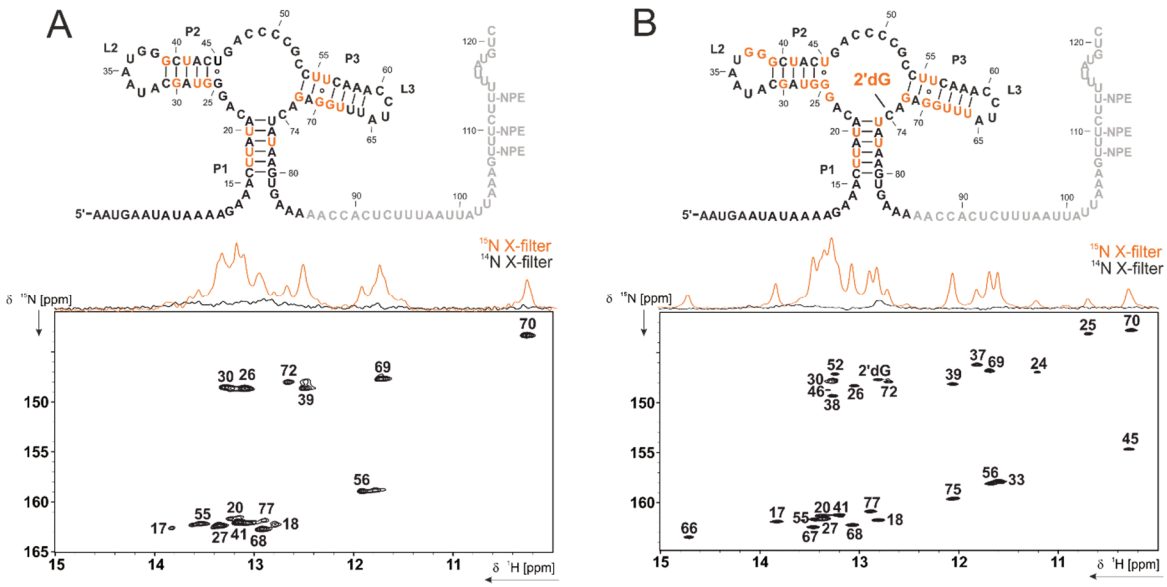


Figure 33: Secondary structure of c-dGsw^{121} in absence (A) and presence (B) of ligand including assigned ^{15}N-TROSY and ^{1}H NMR (^{15}N-X-filtered in orange and ^{14}N-X-filtered in black) spectra. The unlabeled RNA fragment containing nts 86-121 is highlighted in grey. Imino protons of residues highlighted in orange can be detected in 1D and 2D NMR spectra. Ligand-bound spectra were recorded at 600 MHz, 298 K on 100 µM c-dGsw^{121} in presence of 2 eq. of 2'dG and 6 eq. of Mg^{2+}. Ligand-free spectra were recorded at 800 MHz, 298 K on 100 µM c-dGsw^{121} in presence of 6 eq. of Mg^{2+}.

The ^{1}H,^{15}N correlation spectrum of c-dGsw^{121} recorded in the absence of ligand (Figure 33A) contains all reporter signals for the aptamer domain (PA). No imino proton signals characteristic for antiterminator conformation P^{A}T can be detected. Further, the aptamer domain retains its ability to bind 2'dG with high affinity as shown in the ^{1}H,^{15}N correlation spectrum in the presence of ligand (Figure 33B). All critical reporter signals for ligand binding, such as G24, G25, U45, U66, G37 and G38, can be detected. Since the 36 nt fragment containing the photocages (grey segment in Figure 33) was not isotope labeled, intrinsic structure formation within this segment cannot be detected in the spectrum. Imino

protons within this segment were filtered by recording ^{1}H-^{14}N-X-filtered 1D NMR spectra (projected in black above the ^{1}H,^{15}N-correlation spectrum). According to this spectrum the segment containing nts 85-121 does not adopt a homogeneous stable fold. Only broad signals with low signal intensity can be detected, suggesting that the photocaged segment is primarily unstructured and accessible to strand T to form the antiterminator helix PT upon cleavage of the photolabile protection groups.

Subsequently the sample was irradiated at 355 nm to investigate if photocleavage of the NPE protection groups induces formation of a homogeneous antiterminator fold. ^{1}H,^{15}N correlation spectra recorded after laser irradiation including a comparison to the imino proton pattern before the irradiation in a 1D projection are shown in Figure 34. In the absence of ligand, laser irradiation leads to the formation of a homogeneous antiterminator fold and no reporter signals for the aptamer domain can be detected in the final spectrum. With respect to kinetic measurements, antiterminator folding can be monitored by the appearance of the three isolated and characteristic PT helix imino proton reporter signals G12, U108, and U109. Further, the spectrum contains one isolated signal characteristic for the interior loop conformation, namely G25. In the presence of ligand, the riboswitch does not fold completely to the antiterminator conformation after laser irradiation, but the ligand-bound aptamer domain fold is partially retained. The ^{1}H,^{15}N correlation spectrum depicted in Figure 34B shows a mixture of signals corresponding to both the ligand-bound aptamer domain (orange) and the antiterminator conformation (green). This result is in line with the NMR analysis of ligand binding presented in section 4.3.1, according to which dGsw[122] can be switched from the antiterminator conformation $P^{A}T$ to the aptamer (PA) by ~20% in the presence of Mg^{2+}. Since imino protons in the aptamer domain are protected from solvent exchange to a greater extent due to ligand-induced secondary structure stabilizations, relative populations on the basis of exchangeable imino protons cannot be quantified. Despite the incomplete structural rearrangement from PA to PT in the presence of ligand, refolding kinetics can be measured based on several isolated reporter signals for both the antiterminator conformation (U108, U109, U68, G39, G12, U56) and the aptamer domain (U66, G24, G25, G39, U56, G69). Due to the structural identity of helices P2 and P3 in the ligand-free aptamer domain and the antiterminator conformation, the number of reporter signals to monitor the refolding event is by far greater in the ligand-bound state, where helices P2 and P3 are differently aligned with respect to their tertiary structure.[126]

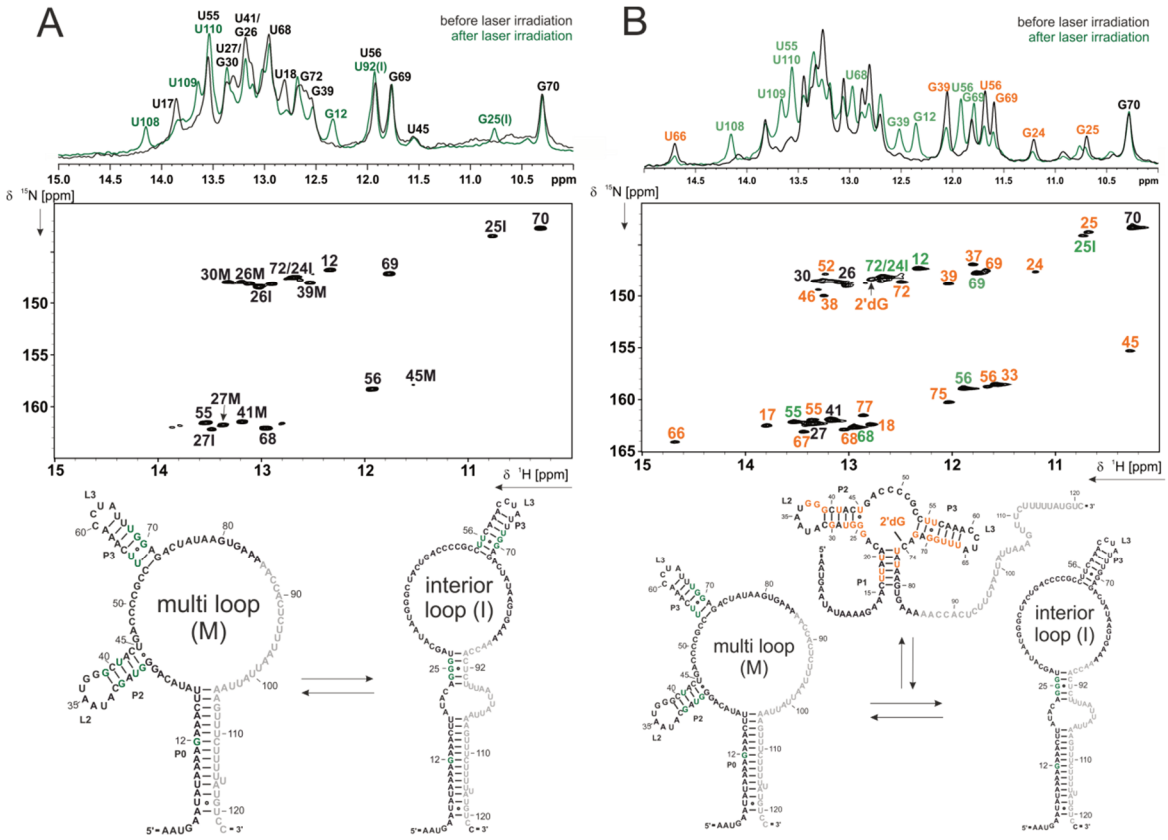


Figure 34: Comparison of NMR spectra of c-dGsw[121] after laser irradiation in absence (A) and presence (B) of ligand. A) ^{1}H NMR spectra recorded before (black) and after (green) laser irradiation at 355 nm including ^{15}N-TROSY spectrum recorded after laser irradiation. Detectable reporter imino proton signals for the antiterminator conformation are highlighted in green in the secondary structure depicted below. The spectra were recorded at 800 MHz, 298 K on 100 µM RNA samples in the presence of 6 eq. of Mg^{2+}. B) ^{1}H NMR spectra recorded before (black) and after (green) laser irradiation at 355 nm including ^{15}N-TROSY spectrum recorded directly after laser irradiation in the presence of ligand. Signals highlighted in orange correspond to the ligand-bound aptamer domain as illustrated in the secondary structure depiction below. Signals highlighted in green correspond to the antiterminator conformation as highlighted in the secondary structure accordingly. Signals labeled in black represent imino protons signals of either helix P2 or P3 that exhibit an identical resonance frequency in both the ligand-bound aptamer domain fold and in the antiterminator conformation. The spectra were recorded at 800 MHz, 298 K on 100 µM RNA samples in the presence of 2eq. 2'dG and 6 eq. of Mg^{2+}.

4.4.4 Kinetics of antiterminator folding

Optimization of laser irradiation

First, the time of laser irradiation was optimized to minimize the dead time of kinetic experiments using c-dGsw[86-121]. Figure 35 shows ^{1}H NMR spectra of c-dGsw[86-121] recorded before and after laser irradiation. The photocleavage reaction can be monitored directly by the change in chemical shift of the NPE methyl group protons, which shift from ~1.5 ppm to ~2.6 ppm upon cleavage. Since c-dGsw[86-121] contains three NPE protection groups, three

signals should be observable. One methyl group resonates at 1.46 ppm and the remaining two resonate at 1.55 ppm. After laser irradiation, they merge into one signal in the final photocleavage product at 2.6 ppm. Signal integration before and after laser irradiation suggests that during 1 s of irradiation, approximately 90% of all photolabile protection groups are cleaved. To increase the time resolution, the duration of laser irradiation can be reduced to 500 ms at the expense of signal sensitivity caused by a decrease in photocleavage yield to ~75%. However, analysis of the methyl group intensity change upon irradiation suggests that the three photocages are cleaved to an equal extent and a cleavage preference for one of the three NPE protection groups cannot be observed.

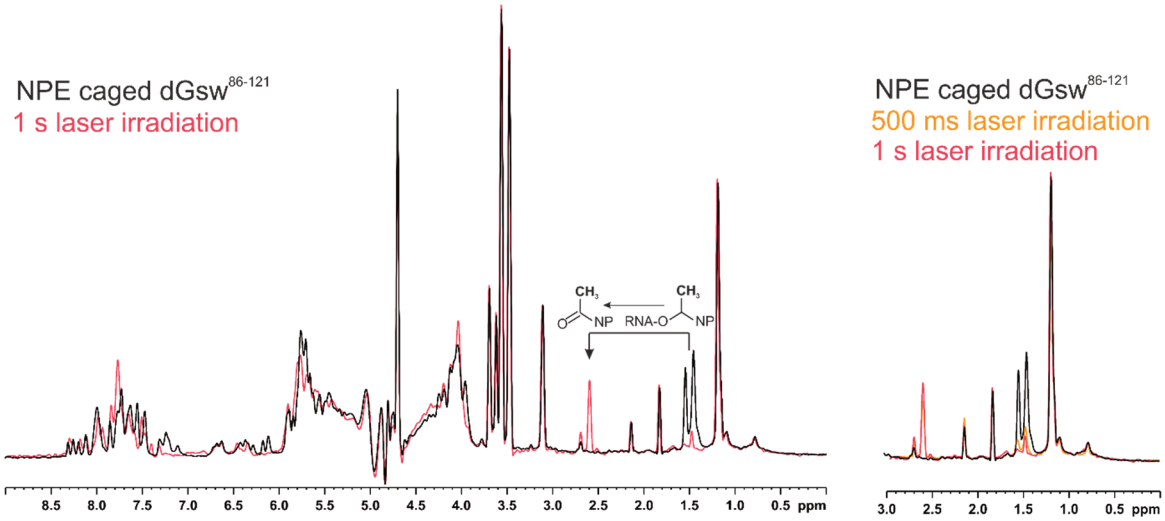


Figure 35: ^{1}H NMR spectra of dGsw$^{86\text{-}121}$ recorded before (black) and after 1s of laser irradiation (red). The chemical shift perturbation of the NPE methyl group caused by the photocleavage reaction from 1.5 ppm to 2.6 ppm is indicated by the depicted chemical structure with NP:2-nitrophenyl. A comparison to a spectrum recorded after 500 ms of laser irradiation (orange) is shown on the right. Spectra were recorded on 100 µM dGsw$^{86\text{-}121}$ at 800 MHz und 298 K.

Kinetics of antiterminator folding in the absence of ligand

Since antiterminator folding in the absence of ligand proved to be relatively fast, kinetics were recorded by irradiating the sample for 500 ms at 355 nm to achieve optimal time resolution. Further, spectra were recorded without implementing any X-filters to achieve optimal sensitivity despite partial isotope labeling of the constructs. The final dead time of the kinetic experiment including the duration of the first experiment is ~900 ms. After this time interval, folding to the antiterminator helix is already completed to ~70%. Figure 36 shows kinetic traces for antiterminator formation averaged over 10 transients. The most

pronounced and isolated reporter signals for antiterminator formation, G12 and U108, cannot clearly resolve this non-equilibration in the first few experiments due to a relatively low S/N. It requires averaging over all detectable signals within P0 (G12, U108, U109, U110) to clearly illustrate that folding from aptamer domain (PA) to the antiterminator helix PT has not yet reached equilibrium during these experiments. The final rate for PT formation in the multi-loop antiterminator conformation $P^{A}T(M)$ is fast with approximately 1.3 s^{-1}. In contrast, the interior loop antiterminator conformation $P^{A}T(I)$, monitored by the isolated reporter signal G25I, folds at a 20-fold slower rate of 6.9 * 10^{-2} s^{-1}.

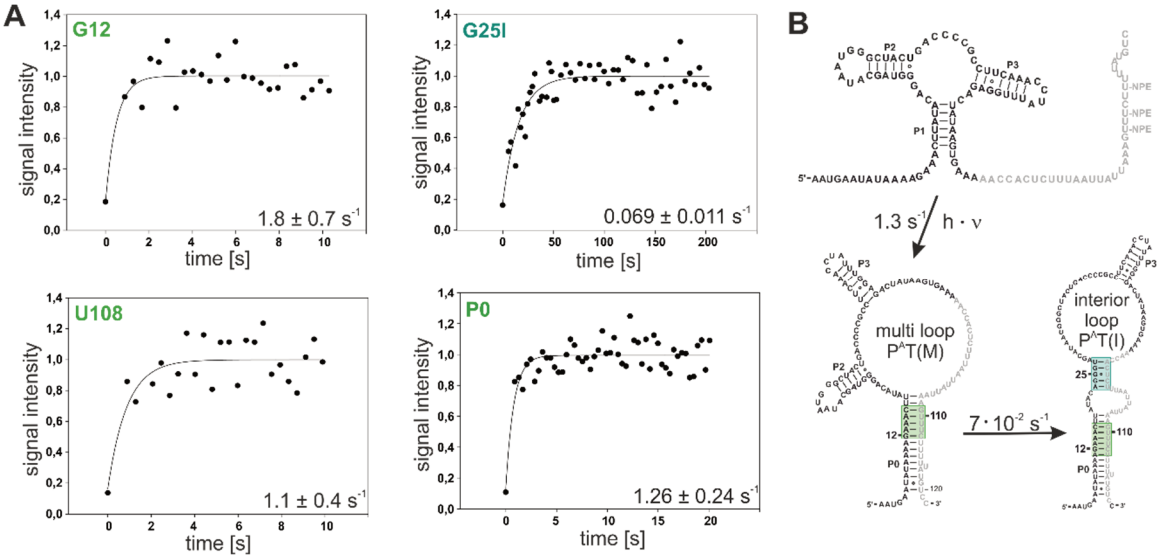


Figure 36: Kinetics of ON-state folding in the absence of ligand. A) Kinetic traces for the isolated PT signals G12, U108 including a kinetic trace averaged over all detectable PT signals (green), and kinetic trace for $P^{A}T(I)$ reporter signal G25I (cyan). B) Schematic representation of secondary structures involved in the refolding event from photocaged dGsw[121] to antiterminator conformations $P^{A}T(M)$ and $P^{A}T(I)$. Helical segments monitored by real-time NMR are highlighted in green (PT) and cyan (elongated PT).

Due to the structural identity of helices P2 and P3 in both aptamer domain and $P^{A}T(M)$, NMR experiments cannot resolve if $P^{A}T(I)$ is adopted directly from the aptamer domain or by equilibration with $P^{A}T(M)$. However, since $P^{A}T(M)$ is formed within seconds and the reporter signal for P1 (U18) disappears equally fast, $P^{A}T(I)$ inevitably has to be adopted from $P^{A}T(M)$. This folding pathway was already predicted in section 4.2.2 and appears reasonable, since direct folding to $P^{A}T(I)$ from the aptamer domain requires dissociation of two helices (P2 and P1) opposed to one (P2) and would involve a much higher energy barrier compared to equilibration from $P^{A}T(M)$. Due to the rapid formation of the PT helix,

fitting of the kinetic trace for G25I with a function including the preceding kinetic ($y(t)=a_0*(1-e^{-1.26t})*(1-e^{-kt})$) leads to an insignificant change in the determined rate.

Compared to rates calculated in section 4.2.2, experiments reveal a ~2.5-fold faster rate for P^AT(M) formation and a ~10-fold slower rate for P^AT(I) formation. The folding rate for P^AT(M) was well estimated by the calculations and the 2.5-fold deviation from experiments likely derives from an overestimation of helix P1 stability in the calculations, which does not distinguish between G-C and A-U base pairs. Similarly, calculated rates to P^AT(I) were purely based on secondary structure, while the presence of reporter signals G39, U45 and G25 suggest weak tertiary interactions between helices P2 and P3 in P^AT(M), which might account for the slower rates in the experiments. Therefore, both differences in helical stability based on the extent of G-C and A-U base pairs and tertiary interactions were not taken into account in the calculations. However, the predictions already suggested that at least *in vitro* and within the time frame of transcription, P^AT(I) cannot quantitatively be formed. This prediction is confirmed by the experiments and reveals P^AT(M) as the single regulatory relevant conformation leading to gene expression. To substantiate this hypothesis, co-transcriptional folding is further simulated based on experimentally determined refolding rates in section 4.4.5.

Kinetics of antiterminator folding in the presence of ligand

In the presence of ligand, samples were irradiated for 1 s and kinetics were monitored after a dead time of approximately 1.4 s. Further, as described in section 4.4.3, helices P2 and P3 can be distinguished in the ligand-bound aptamer and P^AT(M), allowing a more quantitative and accurate analysis of the refolding event. In addition to the formation of the antiterminator helix PT, the dissociation of the ligand-bound aptamer domain complex can be monitored. Individual kinetic traces of both helix P3 and PT formation (Figure 37A) and helix P2 and P3 dissociation (Figure 37B) yield virtually identical rate constants (Figure 37D). This result cross validates signal averaging over single residue transients, as performed for the PT helix in the absence of ligand, to obtain a more accurate final rate for folding of the entire helical segment. The averaged rate for PT helix formation in the presence of ligand corresponds to ~$9 * 10^{-2}$ s^{-1}, and is approximately 10-fold slower compared to the rate determined for the ligand-free aptamer domain. Dissociation of 2'dG and U18 (P1) appears to occur slightly faster with a rate of $1.5 * 10^{-1}$ s^{-1} (Figure 37C) and

precedes the conformational rearrangement. Due to the signal overlap of 2'dG and U18, this preceding kinetics cannot be assigned to either helix P1 dissociation, 2'dG dissociation of if the two dissociation events appear cooperatively.

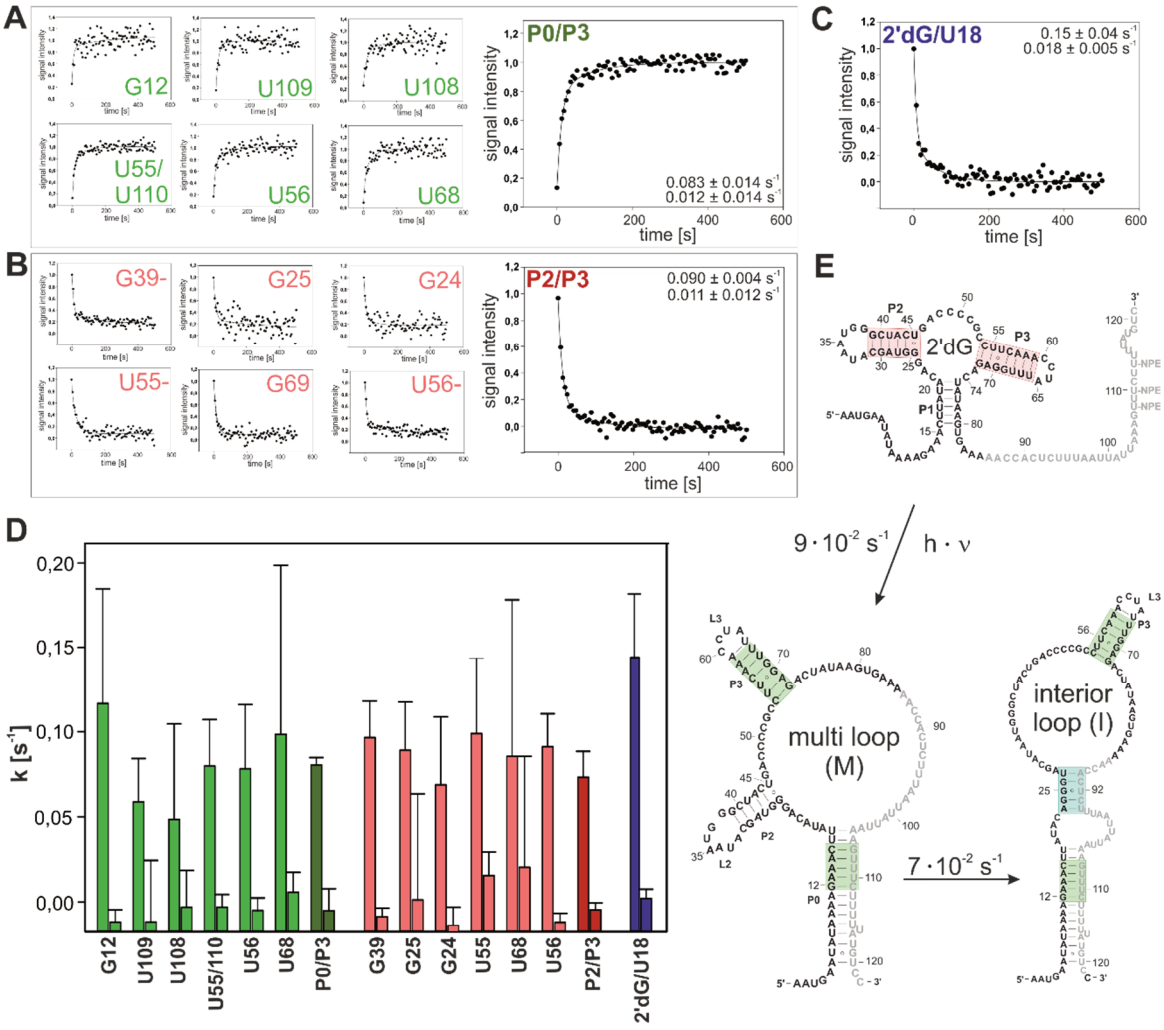


Figure 37: Kinetics of ON-state folding in the presence of ligand. A) Individual kinetic traces of reporter signals for helix P2 and P0 formation in both the multi loop and interior loop formation including averaged kinetic trace over several P2 and P0 signals. B) Individual kinetic traces of reporter signals for helix P2 and P3 dissociation of the ligand-bound aptamer domain including averaged kinetic traces over several P2 and P3 signals. C) Kinetic trace of ligand dissociation including U18 (P1) dissociation. D) Bar plot of refolding rates obtained from individual reporter signals for P0 and P2 formation (light green), averaged P0 and P2 formation (dark green), individual reporter signals for P2 and P3 dissociation (light red), averaged P2 and P3 dissociation (dark red), and ligand including P1 dissociation (blue). E) Secondary structures of the 2'dG-bound aptamer domain in photocaged dGsw[121] and the interior loop and multi loop conformations after cleavage of photolabile protection groups. Secondary structure elements monitored by real-time NMR are color coded according to the bar plot depicted in (D). All kinetic traces were fitted with bi-exponential functions $y=a_0+(1-e^{-k1t})+(1-e^{-k2t})$ (exponential rise) or $y=a_0+e^{-k1t}+e^{-k2t}$ (exponential decay).

Due to incomplete folding from PA to PT in the presence of ligand, the reporter signal for the interior loop antiterminator conformation G25I $P^{A}T(I)$ overlaps partially with G25 in the ligand-bound aptamer domain. Therefore, the refolding rate to $P^{A}T(I)$ cannot reliably be determined. However, if $P^{A}T(I)$ is formed from $P^{A}T(M)$, folding to $P^{A}T(I)$ should not depend on ligand binding and should be identical to the rate determined in the absence of ligand.

Formation of helix PT as well as dissociation of the ligand-bound aptamer complex is best described with bi-exponential functions including an additional slower rate in the order of $\sim 10^{-2}$. The slower rate may resemble re-equilibration induced by the delayed formation of the antiterminator conformation $P^{A}T(I)$. Provided that folding to $P^{A}T(I)$ from $P^{A}T(M)$ is equally fast in the presence of ligand with $7 * 10^{-2}$ s^{-1}, the second rate of $\sim 10^{-2}$ would be on an adequate time-scale for re-equilibration.

Ligand-induced stabilizations of the PA interaction within the aptamer decelerate folding from PA to PT by a factor of 10. This attenuation must be sufficient for the RNA polymerase to transcribe past nts 113-137 at a rate that exceeds folding to PT. Folding to $P^{A}T(M)$ in the presence of ligand occurs on the same time scale as folding to $P^{A}T(I)$ in the absence of ligand. This result further confirms the assumption that $P^{A}T(I)$ cannot form during transcription, otherwise ligand binding would also not lead to effective termination of transcription. In the following section, simulations of co-transcriptional folding in presence and absence of ligand demonstrate the strict fine-tuning between rates of transcription and refolding rates to modulate the regulatory outcome.

4.4.5 Simulations of co-transcriptional folding

Simulations of co-transcriptional folding according to experimentally determined refolding rates were performed in both presence and absence of ligand using a simplified model compared to simulations described in section 4.2.2. The simplified model contains only rates relevant in relation to the time scale of transcription (Figure 38A). For kinetically controlled riboswitches, this approximation is reasonable, since the regulatory outcome is dominated by fast kinetics, while nascent transcripts are never at equilibrium.[2] First, since $P^{A}T(I)$ only reaches a population <10% at very slow transcription rates, refolding from $P^{A}T(I)$ to $P^{A}T(M)$ to reach an equilibrium between the two states was not included in the

simulations. Further, folding to the terminator conformation PA-TH, once the antiterminator conformation $P^{A}T$ is adopted, is assumed to be negligibly slow. This approximation may neglect partial equilibration back to PA-TH during transcription of nts 139-144. However, the refolding window of 5 nts is extremely small for a folding event that was predicted to occur at a rate of ~10^{-5} s^{-1}. Therefore, the simplified model assumes the conformation adopted at a transcript length of 137 nt to represent the regulatory decisive state and only simulates folding during transcription of nts 113 to 137. Folding in the absence of ligand is described in 6 states representing the three conformations: aptamer (PA) (state 1), $P^{A}T$ (M) (state 2) and $P^{A}T$(I) (state 3) at transcription lengths 113 to 137.
In the ligand-free state, PA folds to $P^{A}T$(M) at a rate of 1.26 s^{-1}, followed by $P^{A}T$(I) folding at a rate of 0.07 s^{-1}. In the ligand-bound state, simulations were performed accordingly with a refolding rate from PA to $P^{A}T$(M) of 0.83 s^{-1}. This simulation assumes that all transcripts have bound 2'dG by the time nt 113 is synthesized (6-state model). However, NMR screening of transcriptional intermediates in the presence of ligand (4.3.1) show a continuous decrease in the ligand-bound population during transcription of nts 93-113 from 100% to 70%. Since ligand binding is described by second order kinetics, accurate quantification of the ligand-bound population at nt 113 would require substantial knowledge on physiological RNA concentrations and fluctuating 2'dG concentrations in *Mesoplasma florum*. Therefore, a second simulation was performed, which assumes a ligand-bound population of 70% at a transcript length of 113 nt (8-state model). The two models describe the two extreme cases of rapid ligand binding (within 10-13 nts) followed by kinetic trapping of the ligand bound-state from nts 93-113 opposed to slower ligand binding and/or partial equilibration to 70% by the time nt 113 is transcribed. In both cases, it is assumed that ligand binding is fast compared to the synthesis of nts 80-113. In addition, ligand association rates may vary depending on fluctuations in the cellular 2'dG concentration. Provided the critical cellular concentration of 2'dG required to trigger riboswitch function (2 µM)[1] is exceeded, the accurate fraction of terminated transcripts likely represents an intermediate value between the two models.
Figure 38B shows the fraction of ON-state obtained at a transcript length of 137 depending on the rate of transcription. In Figure 38C, simulations are shown for the three models at transcription rates of 10 nt/s and 40 nt/s. To improve visualization, states representing the identical conformation but a different transcript length were summarized. The simulations show that $P^{A}T$(I) folds only in the ligand-free simulation at slow transcription rates of

10 nt/s with final population of less than 10%, while the fraction of ON-state formed at the regulatory decision point is dominated by $P^{A}T(M)$.

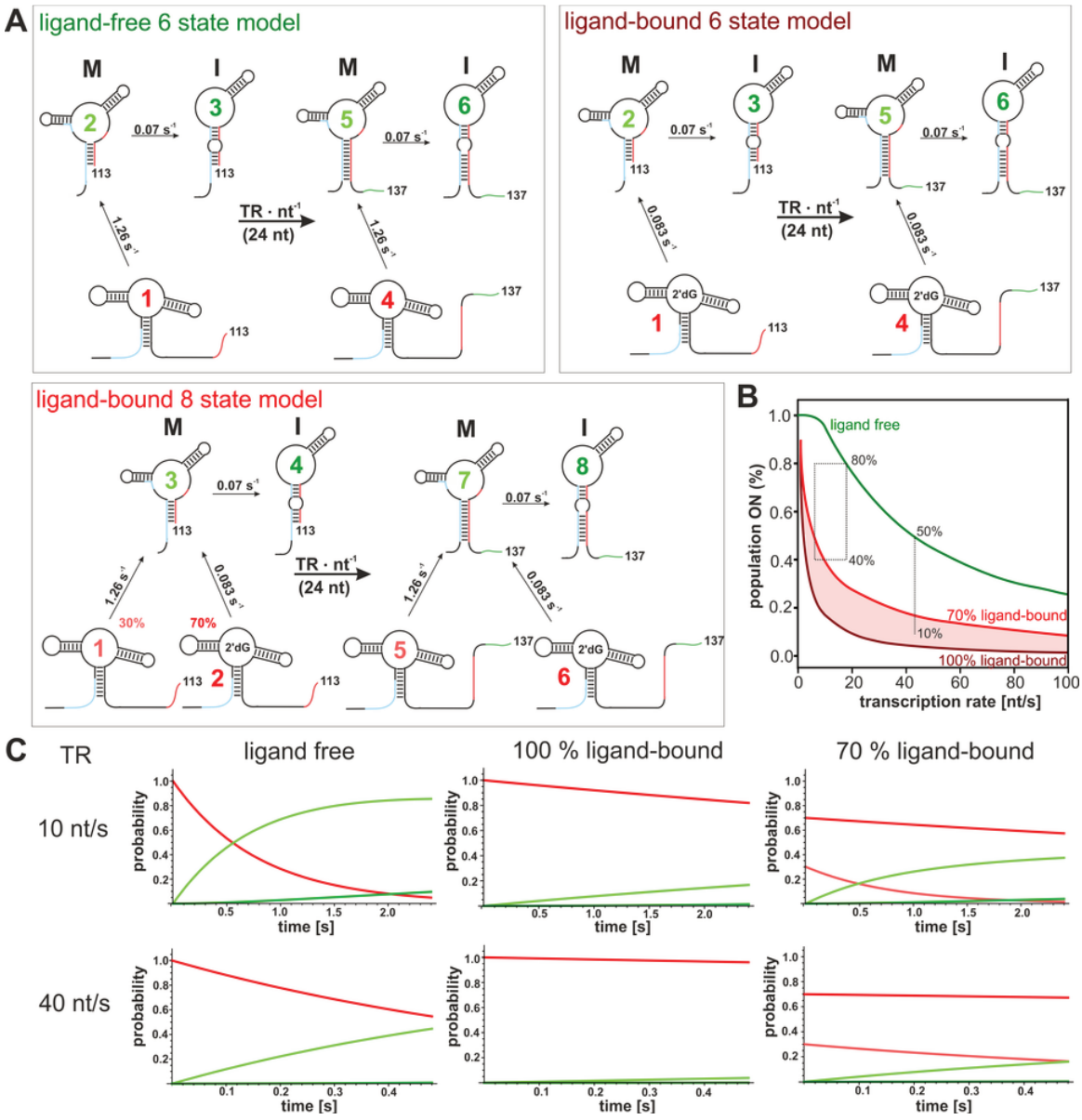


Figure 38:Markovian simulations of co-transcriptional folding. A) Schematic representation of the three models used to simulate co-transcriptional folding during synthesis of nts 113 to 137. The 6 state ligand-free and ligand-bound model describes folding from PA to $P^{A}T(I)$ and $P^{A}T(M)$ based on experimental rates derived from real-time NMR experiments. The 8 state ligand-bound model includes a pre-equilibrium of the aptamer-ligand complex with a final population of the ligand bound state of 70% at transcript length 113 derived from data presented in section 4.3.3. C) Population of the genetic ON-state at the point of regulatory decision depending on the rate of transcription for the three theoretical models: 6 state ligand-free model in green, 6-state ligand-bound model in dark red and 8-state ligand-bound model in light red. Dashed lines represent the riboswitch regulation efficiency determined by *in vitro* transcription assays for the 2'dG riboswitch (80%:40%)[1] and after consideration of the effects of the transcription elongation factor NusA according to reported data (50:10%)[2]. C) Exemplary simulations of co-transcriptional folding for transcription rates of 10 nt/s and 40 nt/s.

According to Figure 38B, the ligand-dependent modulation of gene expression does not exceed 20% at very fast (> 90 nt/s) and very slow (< 5 nt/s) transcription rates. The largest regulation efficiency is achieved at transcription rates of 10-20 nt/s with a ligand-dependent modulation of gene expression of 45%-60%. *In vitro* transcription termination assays propose a 40% modulation of gene expression with a fraction of 80% of ON-state in the absence of ligand and 40% of ON-state in the presence of ligand.[1] To achieve this degree of regulation in both presence and absence of ligand according to our model, transcription rates would be between 10-20 nt/s. Compared to simulations performed in section 4.2.2, transcription rates are shifted to slightly faster rates, in line with the more rapid folding of the antiterminator helix according to real-time NMR experiments.

In vivo, it has been shown that the transcription elongation factor NusA extends the time that polymerase spends on pause sites.[187] *In vitro* transcription termination assays on the FMN riboswitch have shown that addition of NusA causes a ~30% increase in terminated transcripts both in absence and presence of ligand.[2] This increase in the fraction of terminated transcripts is caused by a pause site directly located after the aptamer domain, which extends the window of opportunity for ligand binding. Provided NusA has similar effects in the bacterium *Mesoplasma florum*, transcription rates would be around 40 nt/s. At this rate of transcription, the ligand-dependent modulation of gene expression of 40% is better described by our model (Figure 38B).

4.5. Conclusion and Summary

The current chapter addresses the precise programming in time regarding both ligand-binding and structural rearrangements within the expression platform during transcription. These two folding events need to be tightly coupled to the rate of transcription to allow regulation on the level of transcription as introduced in chapter I (1.2.2).

Mapping of transcriptional intermediates by NMR reveals that both ligand binding and antiterminator folding are precisely fine-tuned in time (Figure 39). In transcriptional OFF-riboswitches, the window of opportunity for ligand binding always precedes the window of opportunity for antiterminator folding. In this way, once the competing antiterminator strand is transcribed, it can directly assess if the ligand is bound and dictate the regulatory outcome. However, the results presented in this chapter suggest that ribowitches act far from the simple model of a binary on-off switch. Small differences in kinetic rates dictate

the regulatory outcome and modulate population ratios with every added nucleotide, such that the allosteric modulation of gene expression does not exceed 50%. This relatively low regulation efficiency may even reflect a protective mechanism by bacteria to prevent overexpression or complete inhibition of gene expression of the downstream gene, and to cause only a small but significant fluctuation in the downstream gene expression.

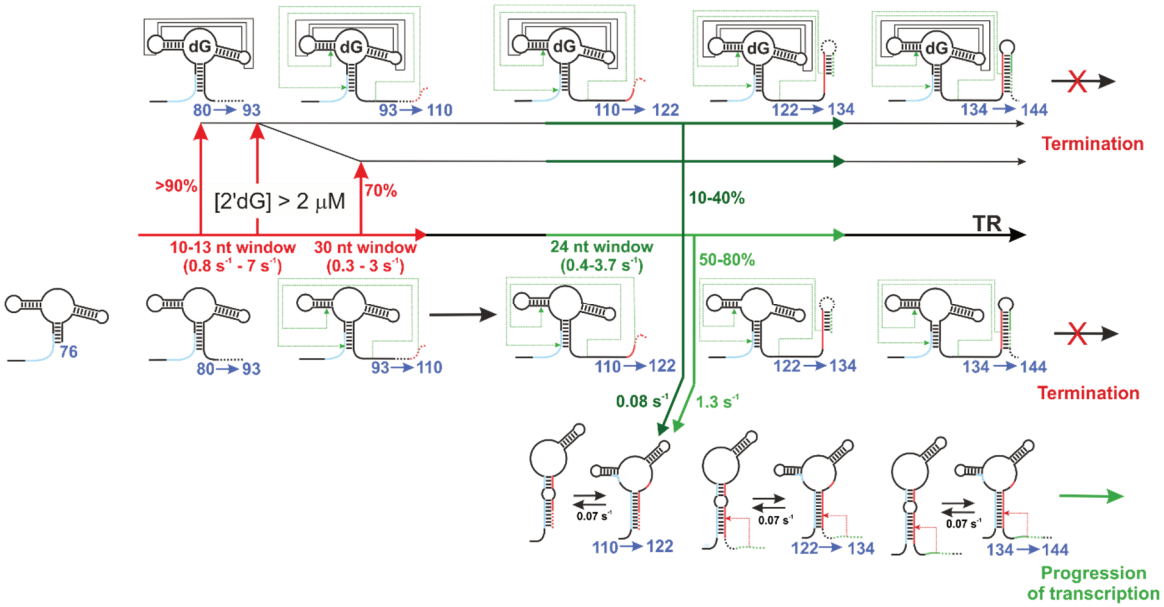


Figure 39: Schematic representation of co-transcriptional folding for the I-A 2'dG-sensing riboswitch. Termination of transcription requires ligand binding directly after the aptamer domain (80 nt) is transcribed. To terminate >90% of all transcripts, ligand binding must occur during the transcription of 10-13 nts. If ligand binding occurs within the transcription of 30 nts, ligand-bound state will only be populated by 70%. After transcription of nt 113, ON-state folding is initiated and time limited to the transcription of 24 nts. According to reported data, ON-state folding completes to 50-80% in the absence of ligand and to 10-40% in the presence of ligand.[1,2] Determined rates by real-time NMR experiments for ON-state folding are indicated accordingly. Transcription rates for respective folding windows are estimated based on reported rates for bacterial polymerases.[9–11] Tertiary interactions within the aptamer domain are indicated by solid black lines. Green and red arrows represent base pairing interactions that compete with displayed secondary structures.

Figure 39B summarizes the distinct folding events in the context of transcription derived from results presented in this chapter. As the mRNA is elongated by the polymerase, the aptamer domain always folds first, and once a stable P1 helix is formed, transcripts can adopt a homogeneous ligand-bound fold. However, ligand binding needs to occur within the transcription of 10-13 nts to enable formation of a stable complex for >90% of all transcripts. If folding to the ligand-bound state cannot be completed within this time frame, the ligand-binding competent population continuously decreases to 70% within the transcription of the following 20 nts. While ligand binding kinetics determined by stopped-flow spectroscopy do not change as the RNA is extended by single nucleotides, the subtle

population changes may allow the riboswitch to react to minor fluctuations in the ligand concentration. The cellular ligand concentration determines the time required for ligand binding. Therefore, a decrease in the ligand concentrations is directly coupled to a lower population in the ligand-bound state. According to this model, the critical ligand concentration required to trigger transcription termination corresponds to the concentration, at which the rate of ligand association is equal to the rate of transcription for nts 80-113. Once the polymerase proceeds past nucleotide 113, refolding to the antiterminator conformation is initiated regardless of whether the ligand is bound. In analogy to the event of ligand binding, where cellular ligand concentrations would always allow equilibrium binding to the aptamer domain, differences in folding kinetics determine if the structural rearrangement can compete with the rate of transcription. Since ON-state folding is described by first order kinetics, it does not depend on cellular concentrations and therefore kinetic rates could be directly derived *in vitro* from real-time NMR experiements. In the absence of ligand, the antiterminator folds at a rate of 1.3 s^{-1}, which is comparable to the rate of transcription to synthesize the 24 nucleotide ON-state folding window (0.4 - 3.7 s^{-1}). In the presence of ligand, the rate of refolding is attenuated slightly below the rate of transcription to 0.08 s^{-1}, allowing the polymerase to transcribe past nts 113-137, while keeping the majority of aptamer-ligand complexes kinetically trapped.

Simulations of co-transcriptional folding visualize how small differences in the transcription rate significantly change population ratios between ON- and OFF-states at the regulatory decision point, and show that the I-A 2'dG riboswitch can only attain >40% regulation efficiency at transcription rates of 10-50 nt/s. Within this dynamic range of transcription rates, the population ratios within both ON-state and OFF-state at the regulatory decision point shifts by ~40%.

References

1. Kim, J. N., Roth, A. & Breaker, R. R. Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16092–16097 (2007).
2. 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. *Mol. Cell* **18,** 49–60 (2005).
3. Frieda, K. L. & Block, S. M. Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* **338,** 397–400 (2012).
4. Lemay, J.-F. *et al.* Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet.* **7,** e1001278 (2011).
5. Mahen, E. M., Watson, P. Y., Cottrell, J. W. & Fedor, M. J. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol.* **8,** e1000307 (2010).
6. Brehm, S. L. & Cech, T. R. The fate of an intervening sequence RNA: Excision and cyclization of the Tetrahymena ribosomal RNA intervening sequence in vivo. *Biochemistry* **22,** 2390–2397 (1983).
7. Kao, C., Zheng, M. & Rudisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5,** 1268–1272 (1999).
8. Helmling, C. *et al.* Rapid NMR screening of RNA secondary structure and binding. *J. Biomol. NMR* **63,** 67–76 (2015).
9. Bremer, H. & Yuan, D. RNA chain growth-rate in Escherichia coli. *J. Mol. Biol.* **38,** 163–180 (1968).
10. Manor, H., Goodman, D. & Stent, G. S. RNA chain growth rates in Escherichia coli. *J. Mol. Biol.* **39,** 1–27 (1969).
11. Adelman, K. *et al.* Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 13538–43 (2002).
12. Crick, F. On protein synthesis. *Symp Soc Exp Biol.* **12,** 138–63 (1958).
13. Crick, F. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).
14. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424,** 147–51 (2003).
15. Gilbert, W. Origin of life: The RNA world. *Nature* **319,** 618–618 (1986).
16. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 3171–5 (1977).
17. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12,** 1–8 (1977).
18. Mattick, J. S. Introns: evolution and function. *Curr. Opin. Genet. Dev.* **4,** 823–31 (1994).
19. Kruger, K. *et al.* Self-Splicing RNA: Autoexcision and Autocyclization of the

Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **31,** 147–57 (1982).

20. Guerrier-Takada, C. *et al.* The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–57 (1983).
21. Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38,** 289–304 (1968).
22. Butcher, S. E. & Brow, D. A. Towards understanding the catalytic core structure of the spliceosome. *Biochem. Soc. Trans.* **33,** 447–9 (2005).
23. Maxwell, E. S. & Fournier, M. J. The Small Nucleolar RNAs. *Annu. Rev. Biochem.* **64,** 897–934 (1995).
24. Henras, A. K., Dez, C. & Henry, Y. RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.* **14,** 335–343 (2004).
25. Meier, U. T. The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114,** 1–14 (2005).
26. Lee, R. C. *et al.* The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75,** 843–54 (1993).
27. Ruvkun, G. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403,** 901–906 (2000).
28. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13959–64 (1998).
29. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).
30. Wassarman, K. M. *et al.* Small RNAs in Escherichia coli. *Trends Microbiol.* **7,** 37–45 (1999).
31. Gottesman, S. *et al.* Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.* **21,** 399–404 (2005).
32. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111,** 747–56 (2002).
33. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 15908–13 (2002).
34. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
35. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **113,** 577–86 (2003).
36. Nahvi, A. *et al.* Genetic Control by a Metabolite Binding mRNA. *Chem. Biol.* **9,** 1043–1049 (2002).
37. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18,** 1433–1445 (2008).
38. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 716–21 (2008).

39. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–23 (2007).

40. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–6 (2008).

41. Mojica, F. J. M., Diez-Villasenor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* **36,** 244–246 (2000).

42. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **60,** 174–182 (2005).

43. 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 Lett.* **555,** 516–20 (2003).

44. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5,** 451–63 (2004).

45. d'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* **216,** 835–58 (1990).

46. Rosenberg, M. & Court, D. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13,** 319–53 (1979).

47. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8,** R239 (2007).

48. Gama-Castro, S. *et al.* RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.* **39,** D98-105 (2011).

49. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* **412,** 793–813 (2011).

50. Farnham, P. J. & Platt, T. Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Res.* **9,** 563–77 (1981).

51. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* **412,** 793–813 (2011).

52. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 14699–704 (1998).

53. Korzheva, N. *et al.* A structural model of transcription elongation. *Science* **289,** 619–25 (2000).

54. Komissarova, N., Becker, J., Solter, S., Kireeva, M. & Kashlev, M. Shortening of RNA:DNA Hybrid in the Elongation Complex of RNA Polymerase Is a Prerequisite for Transcription Termination. *Mol. Cell* **10,** 1151–1162 (2002).

55. Gusarov, I. *et al.* The mechanism of intrinsic transcription termination. *Mol. Cell* **3,** 495–504 (1999).

56. Antao, V. P., Tinoco, I. & Jr. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* **20,** 819–24 (1992).

57. Antao, V. P., Lai, S. Y., Tinoco, I. & Jr. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* **19,** 5901–5 (1991).

58. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4,** (2012).

59. Nelson, J. W. *et al.* Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol.* **9,** 834–9 (2013).

60. Kim, P. B., Nelson, J. W. & Breaker, R. R. An Ancient Riboswitch Class in Bacteria Regulates Purine Biosynthesis and One-Carbon Metabolism. *Mol. Cell* **57,** 317–328 (2015).

61. Dambach, M. *et al.* The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. *Mol. Cell* **57,** 1099–109 (2015).

62. Li, S., Hwang, X. Y., Stav, S. & Breaker, R. R. The yjdF riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds. *RNA* **22,** 530–41 (2016).

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

64. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441,** 1167–1171 (2006).

65. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441,** 1172–5 (2006).

66. Price, I. R. *et al.* Mn2+-Sensing Mechanisms of yybP-ykoY Orphan Riboswitches. *Mol. Cell* **57,** 1110–1123 (2015).

67. Serganov, A., Huang, L. & Patel, D. J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458,** 233–7 (2009).

68. Huang, L., Serganov, A. & Patel, D. J. Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Mol. Cell* **40,** 774–86 (2010).

69. Kang, M., Peterson, R. & Feigon, J. Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell* **33,** 784–90 (2009).

70. Garst, A. D., Héroux, A., Rambo, R. P. & Batey, R. T. Crystal structure of the lysine riboswitch regulatory mRNA element. *J. Biol. Chem.* **283,** 22347–51 (2008).

71. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11,** 1729–41 (2004).

72. Ren, A., Rajashankar, K. R. & Patel, D. J. Global RNA Fold and Molecular Recognition for a pfl Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. *Structure* **23,** 1375–81 (2015).

73. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9,** 644–7 (2003).

74. Nahvi, A., Barrick, J. E. & Breaker, R. R. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32,** 143–50 (2004).

75. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA

structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* **10,** 701–7 (2003).

76. 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 Dev.* **17,** 2688–97 (2003).
77. Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306,** 275–9 (2004).
78. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **11,** 29–35 (2004).
79. Sudarsan, N. *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321,** 411–3 (2008).
80. Watson, P. Y. & Fedor, M. J. The ydaO motif is an ATP-sensing riboswitch in Bacillus subtilis. *Nat. Chem. Biol.* **8,** 963–5 (2012).
81. 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 (2004).
82. Collins, J. A., Irnov, I., Baker, S. & Winkler, W. C. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev.* **21,** 3356–68 (2007).
83. Welz, R. & Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA* **13,** 573–82 (2007).
84. Butler, E. B., Xiong, Y., Wang, J. & Strobel, S. A. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem. Biol.* **18,** 293–8 (2011).
85. Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314,** 300–4 (2006).
86. Sherman, E. M., Esquiaqui, J., Elsayed, G. & Ye, J.-D. An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. *RNA* **18,** 496–507 (2012).
87. Ruff, K. M. & Strobel, S. A. Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. *RNA* **20,** 1775–1788 (2014).
88. Poiata, E., Meyer, M. M., Ames, T. D. & Breaker, R. R. A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* **15,** 2046–56 (2009).
89. Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43,** 867–79 (2011).
90. Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* **30,** 737–49 (1998).
91. Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15,** 439–42 (1999).
92. Roth, A. *et al.* A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14,** 308–17 (2007).
93. Meyer, M. M., Roth, A., Chervin, S. M., Garcia, G. A. & Breaker, R. R.

Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA* **14,** 685–95 (2008).

94. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329,** 845–8 (2010).

95. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **15,** 177–82 (2008).

96. Klein, D. J., Edwards, T. E. & Ferré-D'Amaré, A. R. Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **16,** 343–4 (2009).

97. Jones, C. P. & Ferré-D'Amaré, A. R. Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat. Struct. Mol. Biol.* **22,** 679–85 (2015).

98. Montange, R. K. & Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37,** 117–33 (2008).

99. Edwards, T. E., Klein, D. J. & Ferré-D'Amaré, A. R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr. Opin. Struct. Biol.* **17,** 273–9 (2007).

100. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* **113,** 577–586 (2003).

101. Noeske, J. *et al.* Interplay of 'induced fit' and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res.* **35,** 572–83 (2007).

102. Noeske, J. *et al.* An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1372–7 (2005).

103. Thore, S., Leibundgut, M. & Ban, N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312,** 1208–11 (2006).

104. Smith, K. D. *et al.* Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat. Struct. Mol. Biol.* **16,** 1218–23 (2009).

105. McDaniel, B. A. M., Grundy, F. J., Artsimovitch, I. & Henkin, T. M. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3083–8 (2003).

106. Wickiser, J. K., Cheah, M. T., Breaker, R. R. & Crothers, D. M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44,** 13404–14 (2005).

107. Pan, T., Artsimovitch, I., Fang, X. W., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9545–50 (1999).

108. Lutz, B., Faber, M., Verma, A., Klumpp, S. & Schug, A. Differences between cotranscriptional and free riboswitch folding. *Nucleic Acids Res.* **42,** 2687–96 (2014).

109. Reining, A. *et al.* Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499,** 355–9 (2013).

110. Cole, P. E., Yang, S. K. & Crothers, D. M. Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry* **11,** 4358–4368 (1972).

111. Abbondanzieri, E. A., Shaevitz, J. W. & Block, S. M. Picocalorimetry of transcription by RNA polymerase. *Biophys. J.* **89,** L61-3 (2005).

112. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).

113. Christiansen, L. C., Schou, S., Nygaard, P. & Saxild, H. H. Xanthine metabolism in Bacillus subtilis: characterization of the xpt-pbuX operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *J. Bacteriol.* **179,** 2540–50 (1997).

114. Gilbert, S. D., Love, C. E., Edwards, A. L. & Batey, R. T. Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry* **46,** 13297–309 (2007).

115. Mulhbacher, J. & Lafontaine, D. A. Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Res.* **35,** 5568–80 (2007).

116. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the adenine riboswitch. *Chem. Biol.* **13,** 857–68 (2006).

117. de la Peña, M., Dufour, D. & Gallego, J. Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold. *RNA* **15,** 1949–64 (2009).

118. Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *J. Mol. Biol.* **359,** 754–768 (2006).

119. Haller, A., Soulière, M. F. & Micura, R. The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.* **44,** 1339–48 (2011).

120. Buck, J., Noeske, J., Wöhnert, J. & Schwalbe, H. Dissecting the influence of Mg2+ on 3D architecture and ligand-binding of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **38,** 4143–4153 (2010).

121. Noeske, J., Schwalbe, H. & Wöhnert, J. Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res.* **35,** 5262–73 (2007).

122. Razin, S., Yogev, D. & Naot, Y. Molecular biology and pathogenicity of mycoplasmas. *Microbiol. Mol. Biol. Rev.* **62,** 1094–156 (1998).

123. Hutchison, C. A. & Montague, M. G. in *Molecular Biology and Pathogenicity of Mycoplasmas* 221–253 (Springer US, 2002). doi:10.1007/0-306-47606-1_10

124. Arraes, F. B. M. *et al.* Differential metabolism of Mycoplasma species as revealed by their genomes. *Genet. Mol. Biol.* **30,** 182–189 (2007).

125. Edwards, A. L. & Batey, R. T. A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. *J. Mol. Biol.* **385,** 938–48 (2009).

126. Wacker, A. *et al.* Structure and dynamics of the deoxyguanosine-sensing riboswitch studied by NMR-spectroscopy. *Nucleic Acids Res.* **39,** 6802–12 (2011).

127. Pikovskaya, O., Polonskaia, A., Patel, D. J. & Serganov, A. Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat. Chem. Biol.* **7,** 748–55 (2011).

128. Pikovskaya, O., Polonskaia, A., Patel, D. J. & Serganov, A. Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat. Chem. Biol.* **7,** 748–55 (2011).

129. Wacker, A., Buck, J., Richter, C., Schwalbe, H. & Wöhnert, J. Mechanisms for differentiation between cognate and near-cognate ligands by purine riboswitches. *RNA Biol.* **9,** 672–80 (2012).

130. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5,** 593–9 (2009).

131. Uhlin, U. & Eklund, H. Structure of ribonucleotide reductase protein R1. *Nature* **370,** 533–9 (1994).

132. Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5958–63 (2005).

133. Nandakumar, J., Ho, C. K., Lima, C. D. & Shuman, S. RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J. Biol. Chem.* **279,** 31337–47 (2004).

134. Hengesbach, M. *et al.* RNA intramolecular dynamics by single-molecule FRET. *Curr. Protoc. Nucleic Acid Chem.* **Chapter 11,** Unit 11.12 (2008).

135. Fürtig, B. *et al.* Conformational dynamics of bistable RNAs studied by time-resolved NMR spectroscopy. *J. Am. Chem. Soc.* **129,** 16222–9 (2007).

136. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4,** 936–62 (2003).

137. Latham, M. P., Brown, D. J., McCallum, S. A. & Pardi, A. NMR methods for studying the structure and dynamics of RNA. *Chembiochem* **6,** 1492–505 (2005).

138. Hall, K. B. RNA in motion. *Curr. Opin. Chem. Biol.* **12,** 612–8 (2008).

139. Cromsigt, J., van Buuren, B., Schleucher, J. & Wijmenga, S. Resonance assignment and structure determination for RNA. *Methods Enzymol.* **338,** 371–99 (2001).

140. Wüthrich, K. *NMR of proteins and nucleic acids*. (Wiley, 1986).

141. Flinders, J. & Dieckmann, T. NMR spectroscopy of ribonucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **48,** 137–159 (2006).

142. Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32,** 287–387 (1998).

143. Sklenář, V. & Bax, A. Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* **74,** 469–479 (1987).

144. Woergoetter, E., Wagner, G. & Wuethrich, K. Simplification of two-dimensional proton NMR spectra using an X-filter. *J. Am. Chem. Soc.* **108,** 6162–6167 (1986).

145. Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **101,** 760–762 (1979).

146. Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2,** 661–665 (1992).

147. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds.

J. Biomol. NMR **33,** 199–211 (2005).

148. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127,** 8014–5 (2005).

149. Ernst, R. R., Bodenhausen, G. & Wokaun, A. Principles of Nuclear Magnetic Resonance in One And Two Dimensions. **1987,** (1987).

150. Pervushin, K., Vögeli, B. & Eletsky, A. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *J. Am. Chem. Soc.* **124,** 12898–902 (2002).

151. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 12366–71 (1997).

152. Lescop, E., Kern, T. & Brutscher, B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *J. Magn. Reson.* **203,** 190–8 (2010).

153. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187,** 163–9 (2007).

154. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **128,** 9042–9043 (2006).

155. Dingley, A. J. & Grzesiek, S. Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide 2 J NN Couplings. *J. Am. Chem. Soc.* **120,** 8293–8297 (1998).

156. Wöhnert, J. *et al.* Direct identification of NH...N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res.* **27,** 3104–10 (1999).

157. Shaka, A. J., Barker, P. B. & Freeman, R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.* **64,** 547–552 (1985).

158. Goddard, T. D. & Kneller, D. G. Sparky 3. *University of California, San Francisco*

159. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the adenine riboswitch. *Chem. Biol.* **13,** 857–68 (2006).

160. Lemay, J.-F. *et al.* Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet.* **7,** e1001278 (2011).

161. Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods in Enzymology* **180,** 51–62 (1989).

162. Pokrovskaya, I. D. & Gurevich, V. V. In vitro transcription: preparative RNA yields in analytical scale reactions. *Anal. Biochem.* **220,** 420–3 (1994).

163. Beckert, B. & Masquida, B. Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* **703,** 29–41 (2011).

164. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA

templates. *Nucleic Acids Res.* **15,** 8783–98 (1987).

165. Draper, D. E., White, S. A. & Kean, J. M. Preparation of specific ribosomal RNA fragments. *Methods Enzymol.* **164,** 221–37 (1988).
166. Krupp, G. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* **72,** 75–89 (1988).
167. Been, M. D., Perrotta, A. T. & Rosenstein, S. P. Secondary structure of the self-cleaving RNA of hepatitis delta virus: applications to catalytic RNA design. *Biochemistry* **31,** 11843–11852 (1992).
168. Chowrira, B. M., Pavco, P. A. & McSwiggen, J. A. In vitro and in vivo comparison of hammerhead, hairpin, and hepatitis delta virus self-processing ribozyme cassettes. *J. Biol. Chem.* **269,** 25856–64 (1994).
169. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249,** 398–408 (1995).
170. Ferre-D'Amare, A. R. & Doudna, J. A. Use of Cis- and Trans-Ribozymes to Remove 5' and 3' Heterogeneities From Milligrams of In Vitro Transcribed RNA. *Nucleic Acids Res.* **24,** 977–978 (1996).
171. Birikh, K. R., Heaton, P. A. & Eckstein, F. The structure, function and application of the hammerhead ribozyme. *Eur. J. Biochem.* **245,** 1–16 (1997).
172. Azarani, A. & Hecker, K. H. RNA analysis by ion-pair reversed-phase high performance liquid chromatography. *Nucleic Acids Res.* **29,** E7 (2001).
173. Petrov, A., Wu, T., Puglisi, E. V. & Puglisi, J. D. RNA purification by preparative polyacrylamide gel electrophoresis. *Methods Enzymol.* **530,** 315–330 (2013).
174. Cazenave, C. & Uhlenbeck, O. C. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6972–6 (1994).
175. Nacheva, G. A. & Berzal-Herranz, A. Preventing nondesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur. J. Biochem.* **270,** 1458–1465 (2003).
176. Tabor, C. W. & Tabor, H. Polyamines. *Annu. Rev. Biochem.* **53,** 749–790 (1984).
177. Fürtig, B. *et al.* Conformational dynamics of bistable RNAs studied by time-resolved NMR spectroscopy. *J. Am. Chem. Soc.* **129,** 16222–9 (2007).
178. Lai, D., Proctor, J. R., Zhu, J. Y. A. & Meyer, I. M. R-CHIE: a web server and R package for visualizing RNA secondary structures. *Nucleic Acids Res.* **40,** e95 (2012).
179. Ward, D. C., Reich, E. & Stryer, L. Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J. Biol. Chem.* **244,** 1228–37 (1969).
180. Stivers, J. T. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.* **26,** 3837–44 (1998).
181. Jean, J. M. & Hall, K. B. 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 37–41 (2001).
182. Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. & Pitsch, S. Kinetics of Photoinduced RNA Refolding by Real-Time NMR Spectroscopy. *Angew. Chemie*

Int. Ed. **44,** 2600–2603 (2005).

183. Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. & Pitsch, S. A Caged Uridine for the Selective Preparation of an RNA Fold and Determination of its Refolding Kinetics by Real-Time NMR. *ChemBioChem* **7,** 417–420 (2006).

184. Fürtig, B. *et al.* Time-resolved NMR studies of RNA folding. *Biopolymers* **86,** 360–383 (2007).

185. Wirmer, J., Kühn, T. & Schwalbe, H. Millisecond Time Resolved Photo-CIDNP NMR Reveals a Non-Native Folding Intermediate on the Ion-Induced Refolding Pathway of Bovineα-Lactalbumin. *Angew. Chemie* **113,** 4378–4381 (2001).

186. Kühn, T. & Schwalbe*, H. Monitoring the Kinetics of Ion-Dependent Protein Folding by Time-Resolved NMR Spectroscopy at Atomic Resolution. *J. Am. Chem. Soc.* **122,** 6169–6174 (2000).

187. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7090–5 (2000).

188. Smith, D. J., Query, C. C. & Konarska, M. M. 'Nought may endure but mutability': spliceosome dynamics and the regulation of splicing. *Mol. Cell* **30,** 657–66 (2008).

189. Pan, J. & Woodson, S. A. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **280,** 597–609 (1998).

190. Fong, N. *et al.* Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28,** 2663–76 (2014).

191. Diegelman-Parente, A. & Bevilacqua, P. C. A mechanistic framework for co-transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J. Mol. Biol.* **324,** 1–16 (2002).

192. Lewicki, B. T., Margus, T., Remme, J. & Nierhaus, K. H. Coupling of rRNA transcription and ribosomal assembly in vivo. Formation of active ribosomal subunits in Escherichia coli requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **231,** 581–93 (1993).

193. Perdrizet, G. A., Artsimovitch, I., Furman, R., Sosnick, T. R. & Pan, T. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 3323–8 (2012).

194. Heilman-Miller, S. L. & Woodson, S. A. Effect of transcription on folding of the Tetrahymena ribozyme. *RNA* **9,** 722–33 (2003).

195. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 161–75 (2006).

196. Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19,** 1461–73 (2013).

Appendix

A1: Plasmid DNA

I-A 2'dG riboswitch

GAATTCTAATACGACTCACTATAGGGAGAATTCATTCTGATGAGAGCGAAAG
CTCGAAACAGCTGTGAAGCTGTCAATGAATATAAAAGAAACTTATACAGGGT
AGCATAATGGGCTACTGACCCCGCCTTCAAACCTATTTGGAGACTATAAGTG
AAAAACCACTCTTTAATTATTAAAGTTTCTTTTTATGTCCAAAAGACAAGAAG
AAACTTTCTGCAG

I-A 2'dA riboswitch

GAATTCTAATACGACTCACTATAGGGAGAATTCATTCTGATGAGAGCGAAAG
CTCGAAACAGCTGTGAAGCTGTCAATGAATATAAAAGAAACTTATACAGGGT
AGCATAATGGGCTACTGACCCCGCCTTCAAACCTATTTGGAGATTATAAGTGA
AAAACCACTCTTTAATTATTAAAGTTTCTTTTTATGTCCAAAAGACAAGAAGA
AACTTTCTGCAG

Primers used for site directed mutagenesis:

Forward Primer: CTT ATT TGG AGA TTA TAA GTG AAA AAC

Reverse Primer: TTT GAA GGC GGG GTC AGC AGT AG

I-A 2'dG riboswitch *E.coli* promotor

GAATTCTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTAC
AGCCATGAATATAAAAGAAACTTATACAGGGTAGCATAATGGGCTACTGACC
CCGCCTTCAAACCTATTTGGAGACTATAAGTGAAAAACCACTCTTTAATTATT
AAAGTTTCTTTTTATGTCCAAAAGACAAGAAGAAACTTTTTTATTTAGTTGAA
TTTATAATAAGAGAAAAAGAAAGGATATTAT**ATG**GCAAAAATAAAAAACCA
ATATTACAACGAGTCTGTTTCGCCAATTGCTGCAG

I-B 2'dG riboswitch

GAATTCTAATACGACTCACTATAGGGAGATTCAGTTCTGATGAGAGCGAAAG
CTCGAAACAGCTGTGAAGCTGTCAACTGAATATTAAGAAACTTATACAGGGT

AGCATAATGGGCTACTGACCCCGCCATGAAACCTATTTCATGACTATAGGTCT
TTATAAACACTTTAATTAAAGTTTCTTTCAACATCAAAACTTGAAAGAAACTT
TTTATTTAGTTGAATTTTTAATAAAAAGAAAGGAAAAACATA**ATG**AAAAAAT
TAATAGCAATTTTAGGGGCAGTTGGATTAACTGCTACAGGTTCAGCAAGTTTA
TTAGTTTCTGCAG

III-B G riboswitch

GAATTCTAATACGACTCACTATAGGGAGATAATTATTCTGATGAGAGCGAAA
GCTCGAAACAGCTGTGAAGCTGTCAATAATTAAATATAAAAAACTTATACAT
GACAACATATTGGGTTGTCGACCTGCCTCTGGACCTATCCTTAGACTATAAGC
GTGAGGTTTTTTTACACAAAAAAGCCTCTTTTTTATTTAGAAGGAGATTAAA**A**
TGAAAAATACTCAAATCTTGATTTTAGATTTTGGAAGTCAATATACACAGTTA
TTAGCCTGCAG

Legend:

Restriction sites

Promotor sequence

Ribozyme sequence

RNA sequence

Translation start codon

A2: RNA constructs prepared by *in vitro* transcription

Sequence and primers

Construct	Sequence	Forward Primer	Reverse Primer
dGsw^{FL}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUAU GUCCAAAAGACAAGAAGAAACUUUUUC	T7	AAA AAA GTT TCT TCT TGT CTT TTG
dGsw^{ON}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCC	T7/T7GG	GGA CAT AAA AAG AAA CTT TAA TAA T
dGsw^{APT}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAA	T7/T7GG	TTT CAC TTA TAG TCT CCA AAT AG
dGsw^{75} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACU	T7GG	AGT CTC CAA ATA GGT TTG AAG
dGsw^{76} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUA	T7GG	UAG TCT CCA AAT AGG TTT GAA G
dGsw^{77} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAU	T7GG	AUA GTC TCC AAA TAG GTT TGA A
dGsw^{78} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUA	T7GG	UAT AGT CTC CAA ATA GGT TTG A
dGsw^{79} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAA	T7GG	UUA TAG TCT CCA AAT AGG TTT G
dGsw^{80} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAG	T7GG	CUT ATA GTC TCC AAA TAG GTT T
dGsw^{81} (Sara)	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGU	T7GG	AC TTA TAG TCT CCA AAT AGGTTT
dGsw^{90}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCA	T7GG	TGG TTT TTC ACT TAT AGT CTC C
dGsw^{94}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCU	T7GG	AGA GTG GTT TTT CAC TTA TAG T
dGsw^{96}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU	T7GG	AAA GAG TGG TTT

	CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUU		TTC ACT TAT AG
dGsw^{100}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUU	T7GG	AAT TAA AGA GTG GTT TTT CAC TTA
dGsw^{104}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUA	T7	TAA TAA TTA AAG AGT GGT TTT TCA C
dGsw^{105}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAA	T7GG	TTA ATA ATT AAA GAG TGG TTT TTC A
dGsw^{106}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAA	T7GG	TTT AAT AAT TAA AGA GTG GTT TTT C
dGsw^{107}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAG	T7GG	CTT TAA TAA TTA AAG AGT GGT TTT T
dGsw^{108}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGU	T7	ACT TTA ATA ATT AAA GAG TGG TTT T
dGsw^{109}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUU	T7/T7GG	AAC TTT AAT AAT TAA AGA GTG GTT TT
dGsw^{110}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUU	T7/T7GG	AAA CTT TAA TAA TTA AAG AGT GGT TT
dGsw^{111}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUC	T7	GAA ACT TTA ATA ATT AAA GAG TGG
dGsw^{112}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCU	T7/T7GG	AGA AAC TTT AAT AAT TAA AGA GTG G
dGsw^{113}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUU	T7/T7GG	AAG AAA CTT TAA TAA TTA AAG AGT G
dGsw^{114}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUU	T7	AAA GAA ACT TTA ATA ATT AAA GAG TG
dGsw^{117}	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUA	T7	TAA AAA GAA ACT TTA ATA ATT AAA GAG

dGsw^{118} (Viktor)	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUAU	T7	ATAAAAA GAAACTTT AATAATTA AAGAG
dGsw^{119} (Viktor)	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUAU G	T7	CATAAA AAGAAA CTTTAA TAATTA AAGA
dGsw^{120} (Viktor)	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUAU GU	T7	ACATAA AAAGAA ACTTTA ATAATT AAAG
dGsw^{121} (Viktor)	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACCUAUUUGGAGACUAUAAGUGAAAAACC ACUCUUUAAUUAUUAAAGUUUCUUUUUAU GUC	T7	GACATA AAAAGA AACTTT AATAAT TAAA
dGsw^{132}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGA	T7GG	CTT GTC TTT TGG ACA TAA AAA G
dGsw^{134}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAG	T7GG	TTC TTG TCT TTT GGA CAT AAA AAG
dGsw^{135}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAGA	T7GG	CTT CTT GTC TTT TGG ACA TAA A
dGsw^{136}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAGAA	T7GG	TCT TCT TGT CTT TTG GAC ATA A
dGsw^{137}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAGAA	T7GG	TTC TTC TTG TCT TTT GGA CAT A
dGsw^{138}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAGAAAC	T7GG	TTT CTT CTT GTC TTT TGG ACA T
dGsw^{139}	(GG)AAUGAAUAUAAAAGAAACUUAUACAG GGUAGCAUAAUGGGCUACUGACCCCGCCUU CAAACCUAUUUGGAGACUAUAAGUGAAAA ACCACUCUUUAAUUAUUAAAGUUUCUUUU UAUGUCCAAAAGACAAGAAGAAACU	T7GG	GTT TCT TCT TGT CTT TTG GAC
dGsw$^{0\text{-}62}$	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCUUCAA ACC	T7	GGT TTG AAG GCG GGG TC
dGsw$^{62\text{-}122}$	GGUUUGGAGACUAUAAGUGAAAAACCACU CUUUAAUUAUUAAAGUUUCUUUUUAUGUC C	TAA TAC GAC TCA CTA TAG	GGA CAT AAA AAG

		GTT TGG AGA CTA TAA GTG A	AAA CTT TAA TAA T
dGsw$^{0-35}$	AAUGAAUAUAAAAGAAACUUAUACAGGGU AGCAUAAU	T7	ATT ATG CTA CCC TGT ATA AGT T
dGsw$^{35-122}$	GGGCUACUGACCCCGCCUUCAAACCUAUUU GGAGACUAUAAGUGAAAAACCACUCUUUA AUUAUUAAAGUUUCUUUUUAUGUCC	TAA TAC GAC TCA CTA TAG GGC TAC TGA CCC CGC C	GGA CAT AAA AAG AAA CTT TAA TAA T
dGsw$^{90-122}$	GGACUCUUUAAUUAUUAAAGUUUCUUUUU AUGUCC	TAA TAC GAC TCA CTA TAG GACUCUUU AAUUAUU	GGA CAT AAA AAG AAA CTT TAA TAA T
dGsw$^{C74U-78}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUA	T7GG	TAT AAT CTC CAA ATA GGT TTG A
dGsw$^{C74U-79}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAA	T7GG	TTA TAA TCT CCA AAT AGG TTT G
dGsw$^{C74U-80}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAG	T7GG	CTT ATA ATC TCC AAA TAG GTT T
dGsw$^{C74U-85}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAGUGAAA	T7GG	TTT CAC TTA TAA TCT CCA AAT AG
dGsw$^{C74U-90}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAGUGAAAAA CCA	T7GG	TGG TTT TTC ACT TAT AAT CTC C
dGsw$^{C74U-92}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAGUGAAAAA CCACU	T7GG	AGT GGT TTT TCA CTT ATA ATC T
dGsw$^{C74U-94}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAGUGAAAAA CCACUCU	T7GG	AGA GTG GTT TTT CAC TTA TAA T
dGsw$^{C74U-96}$	GGAAUGAAUAUAAAAGAAACUUAUACAGG GUAGCAUAAUGGGCUACUGACCCCGCCUUC AAACCUAUUUGGAGAUUAUAAGUGAAAAA CCACUCUUU	T7GG	AAA GAG TGG TTT TTC ACT TAT AA
III-B^{FL}	AAUAAUUAAAUAUAAAAAACUUAUACAUG ACAACAUAUUGGGUUGUCGACCUGCCUCUG GACCUAUCCUUAGACUAUAAGCGUGAGGU UUUUUUACACAAAAAAGCCUCUUUUUU	TAA TAC GAC TCA CTA TAG GAA TAA TTA AAT ATA AAA AAC TTA TAC ATG	AAA AAA GAG GCT TTT TTG TGT AAA
III-B^{ON}	AAUAAUUAAAUAUAAAAAACUUAUACAUG ACAACAUAUUGGGUUGUCGACCUGCCUCUG	TAA TAC GAC TCA	TAA AAA AAC CTC

	GACCUAUCCUUAGACUAUAAGCGUGAGGU UUUUUUA	CTA TAG GAA TAA TTA AAT ATA AAA AAC TTA TAC ATG	ACG CTT ATA G
III-B^{APT}	AAUAAUUAAAUAUAAAAACUUAUACAUGA CAACAUAUUGGGUUGUCGACCUGCCUCUGG ACCUAUCCUUAGACUAUAAGCG	TAA TAC GAC TCA CTA TAG GAA TAA TTA AAT ATA AAA AAC TTA TAC ATG	CGC TTA TAG TCT AAG GAT AG
III-B^{T}	GGGAGGUUUUUUUACACAAAAAAGCCUCU UUUUU	TAA TAC GAC TCA CTA TAG GGA GGT TTT TTT ACA CAA AAA AGC CTC	AAA AAA GAG GCT TTT TTG TGT AAA
I-B^{FL}	AACUGAAUAUUAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCAUGAA ACCUAUUUCAUGACUAUAGGUCUUUAUAA ACACUUUAAUUAAAGUUUCUUUCAACAUC AAAACUUGAAAGAAACUUUUUAUUUAGUU GAAUUUUU	TAA TAC GAC TCA CTA TAG GAA CTG AAT ATT AAG AAA CTT ATA CA	AAA AAG TTT CTT TCA AGT TTT GAT G
I-B^{ON}	AACUGAAUAUUAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCAUGAA ACCUAUUUCAUGACUAUAGGUCUUUAUAA ACACUUUAAUUAAAGUUUCUUUCAACAUC AAAA	TAA TAC GAC TCA CTA TAG GAA CTG AAT ATT AAG AAA CTT ATA CA	TTT TGA TGT TGA AAG AAA CTT TAA T
I-B^{APT}	AACUGAAUAUUAAGAAACUUAUACAGGGU AGCAUAAUGGGCUACUGACCCCGCCAUGAA ACCUAUUUCAUGACUAUAGGUCUUUAUAA AC	TAA TAC GAC TCA CTA TAG GAA CTG AAT ATT AAG AAA CTT ATA CA	GTT TAT AAA GAC CTA TAG TCA TG
I-B^{T}	GGAAGUUUCUUUCAACAUCAAAACUUGAA AGAAACUUUUUAUUUAGUUGAAUUUUU	TAA TAC GAC TCA CTA TAG GAA AGT TTC TTT CAA CAT CAA AAC TTG AAA GAA ACT TT	AAA AAG TTT CTT TCA AGT TTT GAT G

Extinction coefficients and Labeling Schemes

Construct	Labeling Scheme	Extinction coefficient [$LM^{-1}cm^{-1}$]
dGsw^{FL}	unlabeled, ^{15}N G/U	1344200
dGsw^{ON}	unlabeled, ^{15}N-G/U, full ^{15}N, ^{15}N-U/^{13}C+^{15}N-A	1123200
dGsw^{APT}	unlabeled, ^{15}N-U, ^{15}N-G/U	789100
dGsw^{75} (Sara Keyhani)	unlabeled	687700
dGsw^{76} (Sara Keyhani)	unlabeled	698100
dGsw^{77} (Sara Keyhani)	unlabeled	707200
dGsw^{78} (Sara Keyhani)	unlabeled	717600
dGsw^{79} (Sara Keyhani)	unlabeled, ^{15}N-G/U	728000
dGsw^{80} (Sara Keyhani)	unlabeled	738400
dGsw^{81} (Sara Keyhani)	unlabeled	747500
dGsw^{90}	unlabeled	833300
dGsw^{94}	unlabeled	864500
dGsw^{96}	unlabeled	882700
dGsw^{100}	unlabeled	921700
dGsw^{104}	unlabeled	960700
dGsw^{105}	unlabeled	971100
dGsw^{106}	unlabeled	981500
dGsw^{107}	unlabeled	991900
dGsw^{108}	unlabeled	1001000
dGsw^{109}	unlabeled	1010100
dGsw^{110}	unlabeled	1019200
dGsw^{111}	unlabeled	1025700
dGsw^{112}	unlabeled	1034800
dGsw^{113}	unlabeled	1043900
dGsw^{114}	unlabeled	1053000
dGsw^{117}	unlabeled	1081600
dGsw^{118} (Viktor Pfeifer)	unlabeled	1090700
dGsw^{119} (Viktor Pfeifer)	unlabeled	1101100
dGsw^{120} (Viktor Pfeifer)	unlabeled	1110200
dGsw^{121} (Viktor Pfeifer)	unlabeled	1103700
dGsw^{132}	unlabeled	1233700
dGsw^{134}	unlabeled	1254500
dGsw^{135}	unlabeled	1264900
dGsw^{136}	unlabeled	1275300
dGsw^{137}	unlabeled	1285700
dGsw^{138}	unlabeled	1292200
dGsw^{139}	unlabeled	1301300
dGsw$^{0-62}$	unlabeled	573300
dGsw$^{62-122}$	^{15}N-G/U	551200

dGsw$^{0\text{-}35}$	^{15}N-U	361400
dGsw$^{35\text{-}122}$	^{15}N-G	761800
dGsw$^{90\text{-}122}$	unlabeled	321100
dGsw$^{C74U\text{-}78}$	unlabeled	720200
dGsw $^{C74U\text{-}79}$	unlabeled	730600
dGsw $^{C74U\text{-}80}$	unlabeled	740000
dGsw $^{C74U\text{-}85}$	unlabeled	791700
dGsw $^{C74U\text{-}90}$	unlabeled	835900
dGsw $^{C74U\text{-}92}$	unlabeled	851500
dGsw $^{C74U\text{-}94}$	unlabeled	867100
dGsw $^{C74U\text{-}96}$	unlabeled	885300
III-B^{FL} (Strahinja Lucic)	^{15}N-G/U	1084200
III-B^{ON} (Strahinja Lucic)	^{15}N-G/U	904800
III-B^{APT} (Alexander Mazur)	unlabeled, ^{15}N-G/U	748800
III-B^{T} (Strahinja Lucic)	^{15}N-G/U	315900
I-B^{FL} (Strahinja Lucic)	^{15}N-G/U	1301300
I-B^{ON} (Strahinja Lucic)	^{15}N-G/U	1120600
I-B^{APT} (Alexander Mazur)	unlabeled, ^{15}N-G/U	759200
I-B^{T} (Strahinja Lucic)	^{15}N-G/U	384800

A3: Synthetic RNA constructs

Construct	**Sequence**
Terminator model	AAAGUUUCUUUUUAUGUCCAAAAGACAAGAAGAAACUUUG
Antiterminator model	AUAUAAAAGAAACUUUUCGAAGUUUCUUUUUAUGUC
dGsw$^{88\text{-}193}$ (model for ON-state)	AGUGAAAAACCACUC
dGsw$^{86\text{-}121}$ (ON-state assembly)	AAU GAA UAU AAA AGA AAC UUA UAC AGG GUC C
dGsw$^{0\text{-}29}$ (ON-state assembly)	AAUGAAUAUAAAAGAAACUUAUACAGGGUCC
dGsw$^{0\text{-}27}$ (interior loop conformation assembly)	AAU GAA UAU AAA AGA AAC UUA UAC AGG GUC C
N3-methylated dGsw$^{86\text{-}121}$	AAC CAC UCU UUA AUU AUU AAA G(3-M-U)U (3-M-U)CU U(3-M-U)U UAU GUC
NPE-caged dGsw$^{86\text{-}121}$	5'-pAAC CAC UCU UUA AUU AUU AAA G(^{C}U)U (^{C}U)CU (^{C}U)UU UAU GUC-3'
III-B Gsw P2 model	AUGACAACAUAUUGGGUUGUCGA
III-B Gsw AT model	AAAAAACUUAUACUUCGGUGAGGUUUUUUG

A4: Mathematica script for Markovian simulations

Script for the 9-state simulation in the absence of ligand (5.2.2)

Synthesis_folding_rev.nb (by Boris Fürtig)

```
In[1]:=Clear[k12, k13, k14, k15, k16, k17, k18, k19, k21, k23, k24, k25,
k26, k27, k28, k29, k31, k32, k34, k35, k36, k37, k38, k39, k41, k42,
k43, k45, k46, k47, k48, k49, k51, k52, k53, k54, k56, k57, k58, k59,
k61, k62, k63, k64, k65, k67, k68, k69, k71, k72, k73, k74, k75, k76,
k78, k79, k81, k82, k83, k84, k85, k86, k87, k89, k91, k92, k93, k94,
k95, k96, k97, k98]
n = 9
kMatrix = {{0, k12, k13, k14, k15, k16, k17, k18, k19}, {k21, 0, k23,
k24, k25, k26, k27, k28, k29}, {k31, k32, 0, k34, k35, k36, k37, k38,
k39},{k41, k42, k43, 0, k45, k46, k47, k48, k49},{k51, k52, k53, k54, 0,
k56, k57, k58, k59},{k61, k62, k63, k64, k65, 0, k67, k68, k69},{k71,
k72, k73, k74, k75, k76, 0, k78, k79},{k81, k82, k83, k84, k85, k86, k87,
0, k89},{k91, k92, k93, k94, k95, k96, k97, k98, 0}};
kMatrix //MatrixForm
K = Transpose[kMatrix]; For[i = 1, i <= n, i++, K[[i, i]] = -
Sum[kMatrix[[i, j]], {j, 1, n}]];
K // MatrixForm
Clear[p];
p[t_, p0_] := MatrixExp[t K] . p0
p0 = {1,0,0,0,0,0,0,0,0};
synth=40
k12=0.43;
k13=0.03;
k14=synth*1/24;
k15=0;
k16=0;
k17=0;
k18=0;
k19=0;
k21=0;
k23=0.43;
k24=0;
k25=synth*1/24;
k26=0;
k27=0;
k28=0;
k29=0;
```

```
k31=0;
k32=0.071;
k34=0;
k35=0;
k36=synth*1/24;
k37=0;
k38=0;
k39=0;
k41=0;
k42=0;
k43=0;
k45=0.43;
k46=0.03;
k47=synth*1/8;
k48=0;
k49=0;
k51=0;
k52=0;
k53=0;
k54=0;
k56=0.43;
k57=0;
k58=synth*1/8;
k59=0;
k61=0;
k62=0;
k63=0;
k64=0;
k65=0.071;
k67=0;
k68=0;
k69=synth*1/8;
k71=0;
k72=0;
k73=0;
k74=0;
k75=0;
k76=0;
k78=0;
k79=0;
k81=0;
```

```
k82=0;
k83=0;
k84=0;
k85=0;
k86=0;
k87=0;
k89=0.43;
k91=0;
k92=0;
k93=0;
k94=0;
k95=0;
k96=0;
k97=0;
k98=0.071;

K // MatrixForm
p[t, p0]
plotend=4

p1=Plot[{p[t, p0][[1]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Lighter[Orange],Thick},AspectRatio->.5]
p2=Plot[{p[t, p0][[2]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Lighter[Blue],Thick},AspectRatio->.5]
p3=Plot[{p[t, p0][[3]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Darker[Blue], Thick},AspectRatio->.5]
p4=Plot[{p[t, p0][[4]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Darker[Orange],Thick},AspectRatio->.5]
p5=Plot[{p[t, p0][[5]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Lighter[Cyan], Thick},AspectRatio->.5]
p6=Plot[{p[t, p0][[6]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Darker[Cyan],Thick},AspectRatio->.5]
p7=Plot[{p[t, p0][[7]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Red,Thick},AspectRatio->.5]
p8=Plot[{p[t, p0][[8]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Lighter[Green],Thick},AspectRatio->.5]
p9=Plot[{p[t, p0][[9]]}, {t, 0, plotend}, PlotRange -> {0,1.1},PlotStyle-
>{Darker[Green],Thick},AspectRatio->.5]

Show[p1,p2,p3,p4,p5,p6,p7,p8,p9]
```

A5: Pulse programs

^{15}N(ω2)-X-filtered NOESY

xfilter1_noe.jb (by Janina Buck)

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"
"d0=3u"
"DELTA=d0+d0+p22"
"d16=100u"
"d22=d2-p16-d16-p1*2-d19*2"
"d23=d2-p16-d16-p21*4-6u"
"p22=p21*2"
"p2=p1*2"
"cnst10=90/90"

1 ze
2 d11 do:f3
3 d12 do:f3
  d1
  10u pl1:f1
  10u pl3:f3
  d12 fq=cnst19(bf ppm):f1
  p1 ph7
  d0
  (p22 ph1):f3
  d0
  (p2 ph1)
  DELTA
  p1 ph8
  d8
  d12 fq=cnst18(bf ppm):f1
  p1 ph1
  d19
  p1 ph2
  d22 UNBLKGRAD
```

```
  p16:gp1
  d16
  p1 ph3
  d19*2
  p1 ph4
  (p21 ph5):f3
  3u
  (p22 ph1):f3
   3u
  (p21 ph6):f3
  p16:gp1
  d16 pl16:f3
  d23 BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2 F1PH(ip7, id0)
  exit


ph1=0
ph2=2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph5=0 2
ph6=0 0 0 0 0 0 0 0
ph7=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph8=0
ph31=0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2
```

^{14}N(ω2)-X-filtered NOESY

xfilter2_noe.jb (by Janina Buck)

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"
"d0=3u"
"DELTA=d0+d0+p22"
"d16=100u"
"d22=d2-p16-d16-p1*2-d19*2"
```

```
"d23=d2-p16-d16-p21*4-6u"
"p22=p21*2"
"p2=p1*2"
"cnst10=90/90"

1 ze
2 d11 do:f3
3 d12 do:f3
  d1
  10u pl1:f1
  10u pl3:f3
  d12 fq=cnst19(bf ppm):f1
  p1 ph7
  d0
  (p22 ph1):f3
  d0
  (p2 ph1)
  DELTA
  p1 ph8
  d8
  d12 fq=cnst18(bf ppm):f1
  p1 ph1
  d19
  p1 ph2
  d22 UNBLKGRAD
  p16:gp1
  d16
  p1 ph3
  d19*2
  p1 ph4
  (p21 ph5):f3
  3u
  (p22 ph1):f3
   3u
  (p21 ph6):f3
  p16:gp1
  d16 pl16:f3
  d23 BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2 F1PH(ip7, id0)
  exit
```

```
ph1=0
ph2=2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph5=0 2
ph6=0 0 0 0 0 0 0 0
ph7=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph8=0
ph31=0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2
```

^{13}C(ω1)-X-filtered NOESY

noesy11echo2013_13Cxfw1.ch

```
#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"d12=20u"
"in0=inf1/2"
"d0=3u"
"DELTA=d8-d12-p16-d16"
"acqt0=-p1*2/3.1416"

1 ze
2 d11 do:f3
  d12 pl16:f3 pl2:f2
3 d12 pl10:f1 pl3:f3
  p17 ph8
  p17*2 ph9
  d1 pl1:f1
  d12 fq=cnst19(bf ppm):f1
  p1 ph1
  d2
  (p2 ph1):f1
  (p3 ph7):f2
  3u
  (p3 ph8):f2
  d2
```

```
  d0
  (p22 ph1):f3 (p4 ph1):f2
  d0
  p1 ph2
  d8:r
  d12 fq=cnst18(bf ppm):f1
  p1 ph3
  d19
  p1 ph4  ;
  50u UNBLKGRAD
  p16:gp1
  d16
  p1 ph5
  d19*2
  p1 ph6
  50u
  p16:gp1
  d16 pl16:f3
  4u  BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2 F1PH(ip1, id0)
exit


ph0=0
ph1=0 2
ph2=(8) 1
ph3=0 0 2 2
ph4=2 2 0 0
ph5=0
ph6=2
ph7=0 2 2 0
ph8=0
ph9=1
ph31=0 0 0 0
```

Real-time NMR experiments

rt_jr_2D (by Boris Fürtig)

```
prosol relations=<triple>


#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"d11=30m"
"d12=20u"
"d16=100u"
"d22=d2-p16-d16-p1-d19-p21"
"d23=d2-p16-d16-p1-d19-p21"
"p22=p21*2"
"p2=p1*2"
"cnst20=cnst21/90"
"p10=p1*cnst20"
"l0=0"
"l23=1"
"l31=cnst10*ns+1"

1 ze
2 d11 do:f3
3 d12
 3m iu0
 if "l0 == l31"
     {
  10u setnmr3|14
  d27
  10u setnmr3^14
  d28
     }
  else
     {
  d1
     }
  10u pl1:f1
  10u pl3:f3
```

```
  p10 ph1
  d19
  p10 ph2
  p16:gp1
  d16

  (p1 ph3 d19*2 p1 ph4):f1

  p16:gp1
  d16 pl16:f3  pl12:f2


  go=2 ph31 cpd3:f3
  d11 do:f3  mc #0 to 2

     F1QF(iu23)
exit

ph1=0
ph2=2
ph3=0
ph4=2
ph5=0
ph6=0
ph7=2
ph31=0
```

German Summary

Riboschalter sind wichtige regulatorische RNA Elemente, die durch selektive Binding von Metaboliten (Liganden) die Genexpression in Bakterien regulieren können. Riboschalter, die den Metaboliten 2‘-desoxyguanosin (2’dG) binden, wurden bislang nur in dem Bakterium *Mesoplasma florum* gefunden und gehören zur Klasse der purin-bindenden Riboschalter.[1] Im Falle des Typ I-A 2’dG-bindenden Riboschalters führen hohe zelluläre Konzentrationen von 2’dG zu einer vorzeitigen Terminierung der Transkription. Dadurch wird der nachfolgende codierende Bereich für das Enzym Ribonukleotid Reduktase nicht transkribiert und somit die Genexpression verhindert.

Studien an Riboschaltern, bei denen die Genexpression auf der Ebene der Transkription reguliert wird, haben gezeigt, dass die Regulation primär durch Faltungskinetiken und nicht durch thermodynamische Gleichgewichte bestimmt wird. Dabei spielt co-transkriptionelle Faltung eine wichtige Rolle.[2–4] Wird die RNA während der Transkription am 3‘-Ende durch einzelne Nukleotide verlängert, führt dies zu einer kontinuierlichen Erweiterung möglicher Strukturen, die angenommen werden können. Dabei konkurrieren Umfaltungsraten zur thermodynamisch stabilsten Struktur mit der Transkriptionsrate.[5,6] Daher können während der Transkription über einen gewissen Zeitraum Strukturen angenommen werden, die nicht der thermodynamisch stabilsten Struktur entsprechen, sondern kinetisch stabilisiert sind. Solche Strukturen haben sich für viele RNA-basierte regulatorische Prozesse als relevant erwiesen.[2,188–196] In Riboschaltern sind sowohl die Ligandenbindung als auch die Umfaltung der Expressionsplattform zeitlich direkt mit der Transkription koordiniert (Abbildung 1).

Während der Transkription faltet zunächst die Ligandenbindungsdomäne (Aptamer). Direkt im Anschluss bestimmt die zelluläre Ligandenkonzentration, ob die RNA den OFF-Faltungspfad oder den ON-Faltungspfad einschlägt. Um die Genexpression auszuschalten, muss die Ligandenbindung zeitlich vor der Synthese des 3‘-Antiterminatorstranges (rot) erfolgen, damit die Terminatorhelix (rot:grün) ausgebildet werden kann. Wird die Aptamerdomäne (orange:blau) nicht vor der Synthese des roten Segments durch Ligandbindung stabilisiert, bildet sich die Antiterminatorhelix (blau:rot) und der Riboschalter befindet sich im ON-Faltungspfad. Gleichzeitig muss die Faltung der Antiterminatorhelix (blau:rot) zeitlich vor der Synthese des 3‘-Terminatorstranges (grün) erfolgen um die Bildung der Terminatorhelix (rot:grün) in Abwesenheit des Liganden zu verhindern. In der vorliegenden Arbeit wurde die zeitliche Koordination dieser

Faltungsereignisse in Relation zur Transkriptionsrate anhand des kinetisch kontrollierten 2'dG-bindenden Riboschalters aus *M.florum* detailliert untersucht.

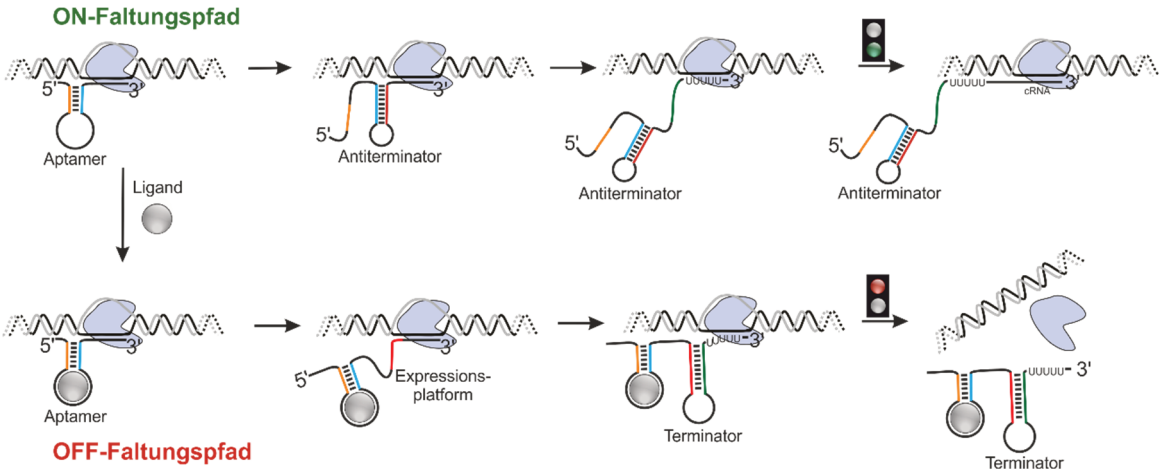


Abbildung 1: Schematische Darstellung des kinetischen Modells der Genregulation auf transkriptioneller Ebene durch Riboschalter.

Zunächst wurden ON- und OFF-Zustand des 2'dG-bindenden Riboschalters mittels NMR-Spektroskopie strukturell charakterisiert. Dabei zeigte sich, dass der Riboschalter voller Länge sowohl im liganden-gebundenen und im liganden-freien Zustand die Terminatorhelix ausbildet und in beiden Fällen den regulatorischen OFF-Zustand annimmt. Der ON-Zustand konnte nur durch Verkürzung der RNA um das grüne Segment in Abbildung 1 in einem Transkriptionsintermediat (122 nt) stabilisiert und identifiziert werden. Mittels struktureller NMR Analyse konnten zwei Konformationen im regulatorischen ON-Zustand nachgewiesen werden, die auf einer Zeitskala von >1 s austauschen. Weiterhin unterschied sich der 2'dG-bindende Riboschalter bezüglich der liganden-abhängigen allosterischen Modulation von den verwandten Guanin- und Adenin-bindenden Riboschaltern (Abbildung 2). Um herauszufinden, ob dieses Umfaltungsmuster charakteristisch für 2'dG-bindende Riboschalter oder spezifisch für das Bakterium *M.florum* ist, wurden zwei weitere Riboschalter aus dem Bakterium *M.florum* untersucht: der I-B 2'dG-bindende Riboschalter und der III-B Guanin-bindende Riboschalter. Strukturelle Untersuchungen mittels NMR-spektroskopie ergaben, dass die allosterische Modulation in allen *M.florum* Riboschaltern identisch ist und daher nicht der Ligandenklasse zugeordnet werden kann. Weiterhin wiesen die NMR-spektroskopischen Untersuchungen darauf hin, dass der I-B 2'dG-bindendende Riboschalter analog zu dem I-A 2'dG-bindenden Riboschalter zwei Konformationen im ON-Zustand annimmt.

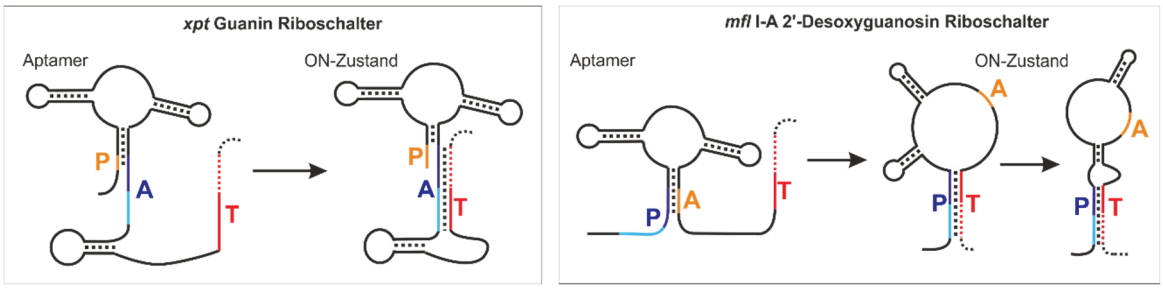


Abbildung 2: Schematische Darstellung der strukturellen Umordnung zwischen den Strängen P, A und T bei der Umfaltung von der Aptamerdomäne in den ON-Zustand beim *xpt* Guanin-bindenden Riboschalter aus *B.subtilis* und dem 2'dG-bindenden Riboschalter aus *M. florum*.

Im Anschluss an die strukturelle Charakterisierung des I-A 2'dG-bindenden Riboschalters wurde eine Methode entwickelt, die es ermöglicht, eine Vielzahl an Transkriptionsintermediaten parallel und schnell zu synthetisieren. Mit dieser Methode sollte schrittweise verfolgt werden, wie sich Strukturen durch Verlängerung der RNA durch einzelne Nukleotide verändern können, und somit der Transkriptionsprozess rekonstruiert werden. Der konventionelle Prozess der RNA Herstellung durch *in vitro* Transkription konnte dabei in zwei Schritten verbessert und beschleunigt werden. Zunächst werden bei run-off-Transkriptionen mit T7 Polymerase 1-2 zufällige Nukleotide an das 3'-Ende angehängt. Die Addition zufälliger Nukleotide wird typischerweise durch die Verwendung von Ribozymen verhindert, die sich während der Transkription selbst abschneiden. Die Verwendung von Ribozymen hingegen verlangsamt und kompliziert den Prozess der RNA Synthese. Weiterhin sind standardisierte Aufreinigungsverfahren für RNA, wie HPLC und PAGE, zeitaufwändig und können nicht parallel durchgeführt werden.

Die 3'-Homogenität von run-off-Transkriptionen konnte durch die Verwendung von DMSO als Zusatzlösemittel signifikant verbessert werden. Wurde DMSO in Kombination mit DNA-Templaten, die 2'-Methoxy Modifikationen an den letzten zwei 5'-Nukleotiden enthielten,[7] eingesetzt, wurde die zufällige Addition von Nukleotiden an das 3'-Ende der RNA komplett unterbunden. Diese Transkriptionsbedingungen lieferten sehr reine und homogene RNA, so dass die Transkriptionsmischung direkt in Zentrifugalkonzentratoren gewaschen und umgepuffert werden konnten. Die resultierenden NMR-Spektren wiesen eine vergleichbare Qualität zu Spektren auf, die von HPLC oder PAGE aufgereinigten Proben aufgenommen wurden.

Der beschleunigte Prozess der RNA Synthese ermöglichte es, eine Vielzahl an Transkriptionsintermediaten innerhalb von zwei Tagen herzustellen und kann in drei Schritte gegliedert werden (Abbildung 3): Zunächst werden DNA Template durch PCR unter Variation der Rückwärts-Primer innerhalb von zwei Stunden hergestellt.

Anschließend werden Transkriptionen auf Basis der PCR-Produkte über Nacht durchgeführt. Im letzten Schritt wird die Transkriptionsmischung in Zentrifugalkonzentratoren für ca. 8-10 Stunden mit Wasser gewaschen und anschließend umgepuffert.[8]

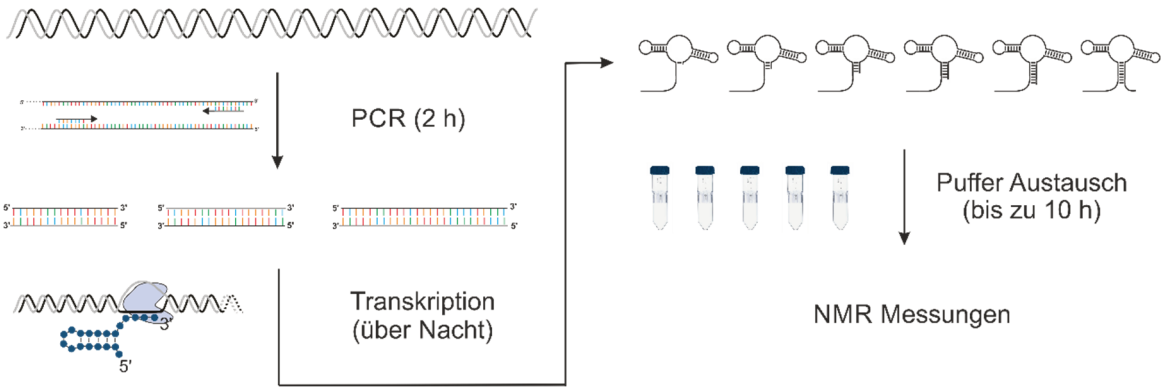


Abbildung 3: Schematische Darstellung der Methode zur schnellen und parallelen Synthese multipler Transkriptionsintermediate. Die DNA Template für die Transkription werden parallel innerhalb von zwei Stunden durch Variation der Rückwärtsprimer hergestellt. Transkriptionen erfolgen über Nacht. Zum Schluss wird die Transkriptionsmischung mit Zentrifugalkonzentratoren innerhalb von 10 Stunden gewaschen und umgepuffert.

Mit der entwickelten Methode wurden 29 Transkriptionsintermediate parallel synthetisiert und biophysikalisch hinsichtlich Struktur und Ligandenbindungseigenschaften untersucht. Die Ergebnisse dieser Untersuchungen sind in Abbildung 4 zusammengefasst.

Zunächst wurde die Sekundärstruktur der Transkriptionsintermediate in Abwesenheit von 2'dG mittels NMR-Spektroskopie bestimmt. Dabei konnte gezeigt werden, dass sich der ON-Zustand über eine Spanne von 24 Nukleotiden zwischen den Transkriptionslängen von 113 und 137 Nukleotiden ausbildet. Demzufolge muss die Umfaltung in den ON-Zustand auf der gleichen Zeitskala erfolgen, wie die Transkription dieser 24 Nukleotide um die Genregulation in Abwesenheit des Liganden anzuschalten. Ab einer Transkriptionslänge von 139 Nukleotiden verschiebt sich das thermodynamische Gleichgewicht in den OFF-Zustand. Da die Entscheidung über den Transkriptionsverlauf 5 Nukleotide später erfolgt, muss der metastabile ON-Zustand nur über diesen begrenzten Zeitraum kinetisch stabilisiert werden.

Anschließend wurde überprüft, über welchen sequentiellen Bereich der Ligand stabil an die Aptämerdomäne binden kann. NMR-spektroskopische Untersuchungen zeigten, dass der liganden-gebundene Zustand nur in der einzelnen Aptamerdomäne zu >90% über einen Sequenzbereich von 10-13 Nukleotiden (Transkriptionslängen 80-93) angenommen wird.

Eine Verlängerung der RNA führt zu einer kontinuierlichen Abnahme (bis zu 70%) des liganden-gebundenen Zustands über die folgenden 20 Nukleotide bis zu einer Transkriptionslänge von 113 Nukleotiden. Der Riboschalter voller Länge bindet den Liganden ebenso nur zu 70%. Die NMR Ergebnisse konnten durch Bestimmung der Dissoziationskonstante mittels ITC für einige selektierte Transkriptionsintermediate bestätigt werden. Auch wurden Ligandenbindungskinetiken mittels Stopped-Flow Fluoreszenz Spektroskopie gemessen. Diese Experimente zeigten, dass sich die Rate der Ligandenbindung mit einer Verlängerung der RNA nicht verändert, insofern die Aptamer Domäne als stabilste Struktur vorliegt. Daraus folgt, dass die Rate der Ligandenbindung ausschließlich durch Fluktuationen in der zellulären 2'dG Konzentration reguliert wird. Da die Ligandenbindung eine Kinetik zweiter Ordnung ist, wird durch eine Erhöhung der Ligandenkonzentration direkt die Bindungsrate beschleunigt. Demensprechend wird bei höheren 2'dG Konzentrationen der liganden-gebundene Zustand durch schnellere Bindung stärker populiert. Der sinkende Populationsverlauf des liganden-gebundenen Zustands mit zunehmender Transkriptionslänge reflektiert somit direkt eine Feinabstimmung der liganden-abhängigen Regulation. Weiterhin zeigen die Ergebnisse, dass die kritische zelluläre Ligandenkonzentration und die daraus resultierende Bindungsrate, die benötigt wird um die Genregulation auszuschalten, der Syntheserate von ca. 30 Nukleotiden entsprechen muss. Ausgehend von Transkriptionsraten für bakterielle RNA Polymerasen, liegt diese Rate im Bereich von 0.3 - 3 s^{-1}.[9–11]

Zuletzt wurde die Umfaltungsrate von der Aptamerdomäne zum ON-Zustand experimentell bestimmt. Dazu wurde das verkürzte, ON-Zustand stabilisierte Transkriptionsintermediat (122 Nukleotide), hergestellt und durch photolabile Schutzgruppen in der Expressionplattform dazu gebracht, die Aptamer Konformation anzunehmen. Zur Herstellung des Konstruktes wurde ein modifiziertes Fragment (36 nt), hergestellt durch Festphasensynthese, an ein isotopenmarkiertes Fragment (85 nt), hergestellt durch *in vitro* Transkription, enzymatisch ligiert. Licht-induzierte Spaltung der photolabilen Schutzgruppen initiiert somit eine Umfaltung von der Apatmerdomäne in den ON-Zustand. Die kinetischen Experimente zeigten, dass die liganden-freie Aptamerdomäne mit einer Rate von 1.3 s^{-1} und die liganden-gebundene Aptamerdomäne mit einer Rate von 0.08 s^{-1} in den ON-Zustand faltet. Während die liganden-freie Aptamerdomäne auf der gleichen Zeitskala in den ON-Zustand faltet, auf der die RNA Polymerase die Nukleotide 113-137 synthetisiert (0.4 - 3.7 s^{-1}), ist die Umfaltungsrate im liganden-gebundenen Zustand im Vergleich zur Transkriptionsrate zu langsam.

Simulationen zur co-transkriptionellen Faltung auf Basis dieser Ergebnisse zeigten, dass der I-A 2'dG-bindende Riboschalter die Genexpression nur in einem Bereich von ~50% regulieren kann und dabei auf charakteristische bakterielle Transkriptionsraten angewiesen ist (10-90 nt/s[9–11]). Folglich ist das traditionelle Modell eines binären AN/AUS Schalters stark vereinfacht. Im Gegensatz dazu werden Populationsverhältnisse durch kontinuierliche Verlängerung der RNA während der Transkription angepasst und durch subtile Fluktuation in der Ligandenkonzentration abgestimmt.

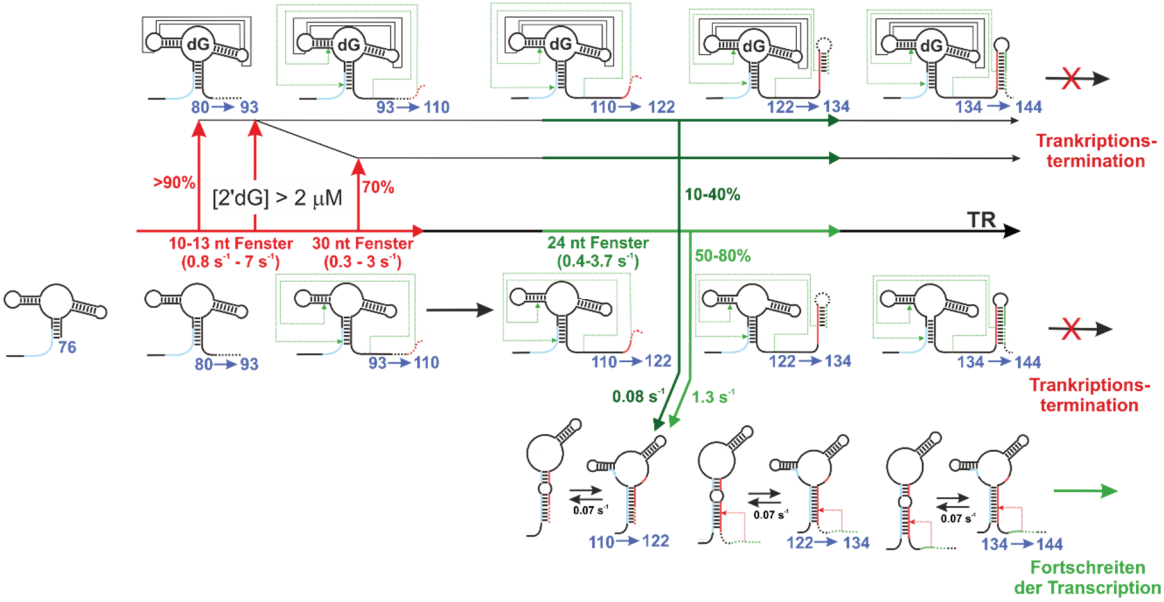


Abbildung 4: Zusammenfassung der co-transkriptionellen Faltung des I-A 2'dG-bindenden Riboschalters. Um die Transkription auszuschalten, muss die Ligandenbindung direkt nach der Synthese der Aptamerdomäne (80 nt) erfolgen. Um 90% aller Transkripte zu terminieren, muss die Ligandenbindung weiterhin während der Transkription von 13 Nukleotiden erfolgen. Sollte der Ligand während der Transkription von 30 Nukleotiden binden, so wird der liganden-gebundene Zustand nur zu 70% ausgebildet. Nach der Transkription von Nukleotid 113, faltet die RNA in den regulatorischen ON-Zustand. Diese Faltung ist auf die Synthese von 24 Nukleotiden beschränkt. Experimentell bestimmte Umfaltungsraten, sowie abgeschätzte Transkriptionsraten für die jeweiligen Faltungsereignisse, sind entsprechend angegeben. Tertiäre Interaktionen sind mit durchgezogenen Linien gekennzeichnet. Gestrichelte grüne und rote Pfeile beschreiben Interaktionen, die mit der angezeigten Sekundärstruktur konkurrieren.

Acknowledgements

First and foremost, I would like to thank Harald Schwalbe for the opportunity to join his group and work on this interesting project in this excellent work environment. Thank you for your consistent support and advice, especially given your limited time, but also for the freedom to work independently and follow my own ideas.

My sincerest gratitude belongs to Anna Wacker for her continuous support on my project. I always found your advice insightful and felt inspired by your enthusiasm and deep knowledge on RNA. Thank you so much! I would also like to express my deepest gratitude to Martin Hengesbach for his helpfulness regarding both project related discussions or difficulties in the lab. Without all your inspiring ideas the project would definitely not have been as successful. Boris Fürtig also deserves a special thanks for all the insightful discussions and always helping out, in particular for the extensive help with the laser setup. I am also grateful to Irene Bessi, who trained me with so much patience when I arrived in the group and introduced me to world of RNA.

My sincerest thanks go to Christian Richter for his indispensable assistance at the NMR-spectrometer, as well as always having an answer to all NMR related questions. Thank you for the many times you saved me in time of spectrometer crisis!

I would like to thank my collaboration partners Dean Klötzner and Alexander Heckel for the persistence on synthesizing and providing the photocaged RNA. For the reliable purification of my many RNA constructs, I would express my gratitude to Elke Stirnal, in particular for the tedious purification of the photocaged ligation product.

Dominic Wagner deserves a special thanks for being so helpful with my PRE project and for all the assistance and training you have done regarding NMR theory to not only me but also many others. Thank you for challenging me and showing me that there is so much more to know about NMR... and for the pile of books on quantum mechanics that now sits on Erhans desk.

I also wish to thank Anna Paulus and Kerstin Dathe for the organizational support. In particular, thanks to Kerstin for always providing cookies and sweets.

I thank the changing members of the ‘boysroom’ within the past years: Melanie Ziegeler, Boris Fürtig, Kai Schnorr, Fabian Hiller, Laurie Lannes, Martin Hähnke, Sara Keyhani, Marie Hutchison, Erhan Cetiner, and Alix Tröster for the cheerful and creative working atmosphere. Thank you so much for all the fun we had in the last years! In particular, I would like to thank Fabian Hiller and Martin Hähnke for the uncomplicated and immediate help with all IT associated questions and issues. Thanks to Fabian for assisting me through my difficult time of learning python. Thanks to Martin for your cheerful and helpful attitude, and for taking partial custody of ‘Fluffy’. Thanks also to Nusrat for always coming to visit at 9 am in search for either food or balcony company.

I thank the lunch group Irene Bessi, Florian Lehner, Laurie Lannes, Hannah Steinert, Marie Hutchison for their company during lunch hours in the social room and the many interesting discussions. Thanks also to all people involved in the “Cocktail evenings”, in particular Sara Keyhani, Diana Müller, Nusrat Qureshi, Gerd Hanspach and Alix Tröster. Thank you for your friendship and all the fun times we had.

I would like to thank all my friends who have followed my path through undergraduate studies and the PhD in the past 8 years. Thanks to Sven Warhaut, Franziska Kappert, Sara Keyhani, Erhan Cetiner and Alix Tröster for your friendship and support.

To all group members including those not mentioned: Thank you for generating this helpful and friendly work environment and for all the pizza and barbecue evenings.

Finally, I would like to thank my family. Without your moral and financial support I would not be where I am today.

Publications

Helmling C, Bessi I, Wacker A, Schnorr KA, Jonker HRA, Richter C, Wagner D, Kreibich M, Schwalbe H (2014). Noncovalent spin labeling of riboswitch RNAs to obtain long-range structural NMR restraints. *ACS Chem Biol*, 9(6):1330-9.

Neugebauer P, Krummenacker JG, Denysenkov VP, **Helmling C**, Luchinat C, Parigi G, Prisner TF (2014). High-field liquid state NMR hyperpolarization: a combined DNP/NMRD approach. *Phys Chem Chem Phys*, 16(35):18781-7.

Helmling C, Keyhani S, Sochor F, Fürtig B, Hengesbach M, Schwalbe H (2015). Rapid NMR screening of RNA secondary structure and binding. *J Biomol NMR*, 63(1):67-76.

Fürtig B, Schnieders R, Richter C, Zetzsche H, Keyhani S, **Helmling C**, Kovacs H, Schwalbe H (2016). Direct (13)C-detected NMR experiments for mapping and characterization of hydrogen bonds in RNA. *J Biomol NMR*, 64(3):207-21.

Helmling C, Wacker A, Wolfinger MT, Hofacker IL, Hengesbach M, Fürtig B, Schwalbe H. NMR structural profiling of transcriptional intermediates reveals riboswitch regulation by metastable RNA conformations. *Revision submitted.*