Cotranscriptional Folding and Fragment-Based Drug Discovery of Regulatory mRNA

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

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

von
Albrecht Eduard Völklein
aus Bloemfontein

Frankfurt 2023

(D30)

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

Dekan:	Prof. Dr. Clemens Glaubitz
Gutachter:	Prof. Dr. Harald Schwalbe Prof. Dr. Michael Göbel
Datum der Disputation:	11th of March 2024

This might be the moment,

the moment you realize that you will die within the next instant of the universe

and

you might realize that a marginally better control of

the biochemistry of your body might cut you time

to find better biochemistry

to perpetually run from the Grim Reaper

and be there

at the end of it all.

But it probably won't be the moment for you – It was for me.

For

Bridget, Karl and Gabi

Table of Contents

Summary 11

Zusammenfassung in deutscher Sprache 17

Chapter 1 Introduction 25

1.1 RNA 25

1.1.1 Chemical composition 25

1.1.2 RNA structure 26

1.1.3 Biological activity of *cis*-acting RNAs 29

1.1.4 Transcription 29

1.1.5 Cyclic-di-nucleotides 31

1.2 Riboswitches and aspects of riboswitch-based gene regulation 32

1.2.1 Kinetic and thermodynamic control of transcriptional riboswitches 36

1.2.2 Pausing and riboswitch function 39

1.2.3 Global k_{on} as a descriptor of ligand binding under kinetic control 39

1.2.4 k_{on} and k_{off} rates of different riboswitches 41

1.3 Nuclear Magnetic Resonance 43

1.3.1 NMR of RNA 43

1.3.2 Challenges of NMR analysis 44

1.3.3 Signal acquisition and assignment 46

1.4 NMR-based methods for drug discovery 48

1.4.1 Comparison of FBS and high throughput screening 51

1.4.2 Experimental challenges of FBS 52

1.4.3 Targeting RNA with fragment-based screening 53

1.5 Affinity and kinetic measurements using ITC 55

1.5.1 History of ITC 56

1.5.2 NMR and ITC 57

1.5.3 Kinetics studies using ITC 57

1.6 Simulation of riboswitch RNA folding 59

1.6.1 The cotranscriptional folding problem 59

1.6.2 Computational modeling of specific riboswitches 61

Chapter 2 Cotranscriptional Folding Landscape of CDN-Riboswitches 65

2.1 Introduction 65

2.1.1 CDN-binding riboswitches 65

2.1.2 Transcriptional riboswitches 67

2.1.3 Cotranscriptional folding of riboswitches 67

2.2 Results 68

2.2.1 NMR chemical shift assignment of the pilM riboswitch 68

2.2.2 Affinity analyses of different transcript lengths 70

2.2.3 Modeling of the cotranscriptional conformation landscape of CDN riboswitches 72

2.3 Discussion 77

2.3.1 Determination of key transcriptional intermediates of the pilM riboswitch 78

2.3.2 The importance of the P1 helix (P1A) 79

2.3.3 Cotranscriptional modeling parameters and regulatory limits of kinetic and thermodynamic control 80

2.3.4 Pausing reduces ligand dependence of Cd1 and indicates thermodynamic control with a regulatory window at lower concentrations. 81

2.3.5 Median response concentrations and regulatory limits of kinetic and thermodynamic control 81

2.3.6 Base pair closing rate as a major driver of regulatory transitions 84

2.3.7 Key conformations and time intervals of ON and OFF regulation 84

2.3.8 Erroneous gene expression is minimized at optimal transcription speeds and by large dynamic range of signaling for OFF switches 85

2.3.9 Riboswitches minimize gene leakiness under thermodynamic control for OFF switches and maximize ligand responsiveness under kinetic control for ON switches 87

2.3.10 Theoretical analysis from single nucleotide extension 88

Chapter 3 ^{19}F NMR-Based Fragment Screening of RNA 89

3.1 Introduction 89

3.1.1 Fragment-based screening for drug discovery 89

3.1.2 ^{19}F-CPMG-based screening for FBS 90

3.1.3 Diversity of RNA structural elements – Targeting RNA 92

3.2 Results 94

3.2.1 ^{19}F-CPMG screening of RNA and other macromolecules 94
3.2.2 Cheminformatic analysis of hit data 99
3.2.3 Follow-up chemistry 102
3.3 Discussion 103
3.3.1 Feasabitly ^{19}F-CPMG screening of RNA and other macromolecules 104
3.3.2 Correlational analysis and limits of molecular descriptors 105
3.3.3 Feasibility of drugs discovery follow up 106
3.3.4 ^{19}F-CPMG NMR as a powerful tool for fragments-based drug discovery 106
Chapter 4 Concluding remarks 109
4.1.1 Outlook on targeting kinetically controlled transcriptional riboswitches with FBS. 112
References 113
Chapter 5 Appendix 133
5.1 Figure and table captions 133
5.2 Methods and supplementary information 142
5.2.1 DNA templates 142
5.2.2 PCR primers 143
5.2.3 RNA and CDN preparation 144
5.2.4 PAGE analyses 145
5.2.5 NMR spectroscopy 145
5.2.6 NMR data 146
5.2.7 ITC 156
5.2.8 ITC data 156
5.2.9 Kinetic data 162
5.2.10 Simulation of cotranscriptional folding 163
5.2.11 Source code of Markov model 164
5.2.12 Fragment-based screening of the iNEXT library 171
5.2.13 Biological context and available structural data of screened targets 173
5.2.14 Fragment library 175
5.2.15 Hit confirmation 185
Acknowledgments 187

Publications and recent conference contributions ....**Error! Bookmark not defined.**

Curriculum vitae ....**Error! Bookmark not defined.**

Summary

This work investigated cyclic-di-nucleotide (CDN) riboswitches, gaining a detailed understanding of their folding during transcription and the feasibility of ^{19}F-NMR fragment-based screening for RNA binding lead structures for riboswitch aptamers.

A recently discovered class of riboswitches has been found to regulate genes for the environment, membranes, and motility (GEMM) in bacteria by binding cyclic-di-nucleotides (CDN), a bacterial second messenger. Environmental stimuli cause bacteria to adjust their internal CDN concentration through increased formation or degradation, driving lifestyle changes like colony formation. At the center of this regulatory network, riboswitches sense the *in vivo* concentration and regulate genes required for these lifestyle changes, acquiring a pivotal role in bacterial physiology and virulence. This work investigated the pilM riboswitch, an ON switch that enables pili expression in *Geobacter metallireducens* (Figure 1A). To further understand the transcriptional folding processes of riboswitches, the equivalent experimental results of Tom Landgraf on the Cd1 (*Clostridium difficile)* riboswitch, an OFF switch limiting flagella expression, were combined into a Markov simulation of cotranscriptional riboswitch function. The simulation showed the expected regulatory behavior in ligand concentrations associated with lifestyle changes and at typical transcription speeds.

Due to their high relevance to bacterial metabolism, novel strategies to regulate riboswitches have been a sought-after endeavor. The looming threat of multi-drug-resistant germs further drives this effort. Drugs that interact with CDN riboswitches could reduce the disease burden caused by *Vibrio cholera* and *Clostridium difficile*. In contrast to classical drug discovery approaches utilizing large compound libraries to screen for lead structures, a recent approach utilizing small molecular fragments to find weak but high-quality interactions has yielded FDA-approved medications. This work outlines how nuclear magnetic resonance (NMR) can be used for fragment-based drug discovery targeting RNA. To assess the drugability of CDN riboswitches, the pilM RNA was one of 14 RNA targets of different secondary and tertiary structures screened using ^{19}F-NMR for binding with five pools of 20 small molecule fragments and compared to five DNAs and five proteins. It was shown that RNA could be specifically targeted. Key parameters analyzed were the targetability of the different RNAs, the library's capability to target RNA over proteins and DNA, the quality of the detected interaction, and the ability to synthetically innovate hits in tighter binding lead compounds through linking.

Project 1: Cotranscriptional Folding Landscape of CDN-Riboswitches

This project utilized the rapid RNA preparation procedures developed by C. Helmling to produce RNA constructs of different lengths. An initial optimization phase allowed high-yield preparation of binding competent RNA with homogenous length in unmodified wildtype pilM sequence. Starting at the aptamer sequence, constructs were continuously extended or shortened until binding was persistently

no longer observed to establish a binding competent window. All constructs were titrated with cognate ligand to assess binding-associated changes in their 1D-NMR spectra. Constructs of particular relevance were further analyzed by preparing isotopically labeled samples, and comprehensive imino assignments of ligand-free and ligand-bound conformations were accomplished.

Key findings included that the truncated version of the aptamer was still binding competent until the residue C75 was removed. C75 is known to engage in Watson-Crick-base pairing with the ligand. Constructs remained binding competent even when the initial bases of the terminator stem were included in the transcript. Binding was no longer observed when the terminator stem reached a length that included a C75 base pair. The findings were substantiated by observing the P2, and P3 stem for all constructs, while indications of a P1 stem formation required ligand addition. Ligand addition was observed through several changes in the imino spectrum and in line with homolog crystal structures. The binding competent window was independently confirmed in triplicate ITC measurements and completely agreed with the obtained NMR results. The binding competent window for pilM is 25 nt (nucleotides) long. The dissociation constants for pilM within this window varied from 19 µM to 0.25 µM, spanning two orders of magnitude (Figure 1B and C). Line shape analyses of the ITC

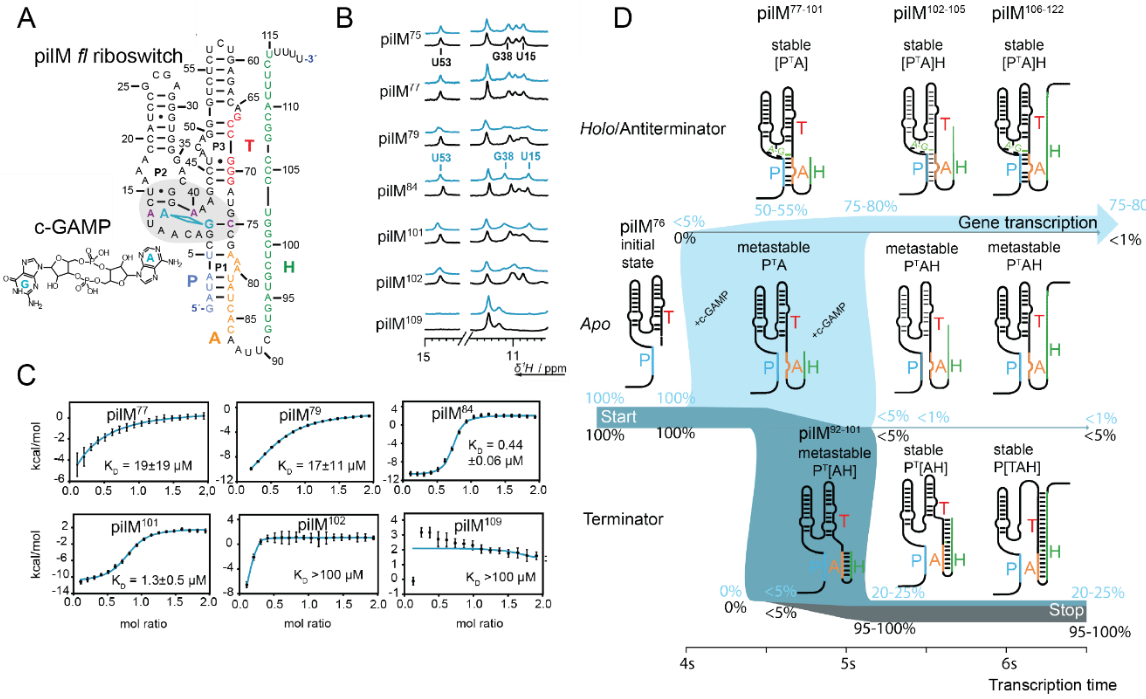

A
pilM fl riboswitch
c-GAMP
P1
P2
P3
P
A
T
H
B
pilM75
pilM77
pilM79
pilM84
pilM101
pilM102
pilM109
U53
G38 U15
15
11
δ¹H / ppm
C
kcal/mol
pilM77
K_D = 19±19 µM
pilM79
K_D = 17±11 µM
pilM84
K_D = 0.44 ±0.06 µM
pilM101
K_D = 1.3±0.5 µM
pilM102
K_D >100 µM
pilM109
K_D >100 µM
mol ratio
D
pilM77-101
stable [PTA]
pilM102-105
stable [PTA]H
pilM106-122
stable [PTA]H
Holo/Antiterminator
pilM76 initial state
Apo
metastable PTA
metastable PTAH
+c-GAMP
Gene transcription
Start
Stop
pilM92-101 metastable PT[AH]
stable PT[AH]
stable P[TAH]
Terminator
100%
<5%
0%
50-55%
75-80%
<1%
20-25%
95-100%
4s
5s
6s
Transcription time

Figure 1 Project 1 overview: (A) Secondary structures of pilM in ligand-bound conformation. The bound ligands are indicated in blue. The 5´-aptamer strand P pairs with an aptamer-stabilizing strand A to form the binding competent aptamer. The switching strand T and the terminator strand H form the expression platform. (B) and (C) show key constructs in outlining the ligand recognition window through ^{1}H-NMR and ITC titration with c-GAMP. Ligand-containing samples are indicated in blue. (D) Markov model simulations of cotranscriptional folding state distribution over time. State population densities are shown for 100 nM ligand concentration (gray, based on data shown in Figure 13B) and 100 µM ligand concentration (blue, based on data shown in Figure 13C). Areas of the figure where population densities overlap are shown in a gray-blue color. The bar thickness indicates a relative population ranging from 5% to 100%. Transcription intermediates of length from 77-101 nucleotides are binding capable. Figure modified from Landgraf *et al.* (1) and used in Figure 10, Figure 12, and Figure 15.

injection peaks yielded kinetic parameters of ligand binding. While the association rate constant (k_{on}) remained similar for all constructs, the differences in dissociation constant (K_D) were caused by dissociation rate constant (k_{off}) differences. T. Landgraf carried out similar analyses of the Cd1 riboswitch. He observed that the absence of a P1 stem in the Cd1 sequence could be explained by Cd1 reaching the same affinity regimes as pilM through the elongation of the aptamer by a single nucleotide, increasing the aptamer affinity 10-fold. The K_D, k_{on}, and k_{off} rates of the two tightest binding riboswitch aptamers were 0.25±0.04 µM, 17000±3000 $M^{-1}s^{-1}$, and 0.0042±0.0005 s^{-1} for pilM and 0.25±0.04 µM, 21000±300 $M^{-1}s^{-1}$, and 0.0045±0.0005 s^{-1} for Cd1, respectively.

Both riboswitches were investigated in a Markov model of cotranscriptional folding. The Markov model expended B. Fürtig’s previous modeling of purine sensing riboswitches. The model simulated cotranscriptional folding by assigning each transcript length to one of three possible states: *apo, holo*, or aptamer competing fold (Figure 1D). The Markov model demonstrated the capability to predict the regulation of the riboswitch gene and its dependency on the transcription rate, pausing, and concentration of ligand. Transitions between possible states had rates assigned based on that state's specific secondary structure at that transcript length and the determined binding affinity. The comparison of the two riboswitches, one an OFF and the other an ON switch, with the same model, allowed the comparison of the observed changes. It further proved the general applicability of the model for transcriptional ON and OFF switches with different lengths of ligand-binding competent windows. The model reproduced the biologically expected outputs in all high and low cognate ligand concentration conditions. The model was used to plot population densities of *apo, holo*, or aptamer competing folds as contour plots. This analysis allowed the observation of transitions from the initially populated *apo* state towards more stable states as a function of time and a second variable. These contour plots were initially utilized to determine the lower limit of a base pair closing rate required to describe base pairs that can form without breaking base pairs elsewhere in the structure. Despite usually occurring on nanosecond time scales, we found that the model was not influenced when speeds of 2.5 ms were used.

Further analysis using contour plots applied commonly assumed *in vivo* CDN concentrations, transcription speeds, and pausing durations. CDN concentration was assumed to be 100 nM or 100 µM depending on the lifestyle state and corresponding cellular signaling. The transcription speed was assumed to be 20 nt/s, and the Cd1 riboswitch was modeled with and without an additional pausing delay of 10 seconds at a suspected pause site. For each contour plot, one of these parameters was varied. The resulting contour plots allowed the observation of the time-dependent influence of CDN concentration between 100 nM and 100 µM, with transcription speed between 1 and 100 nt/s. The terminator population of pilM decreased from over 95% to 20-25% at increasing CDN concentrations (Figure 1D). The same change in concentration increased the terminator population from under 5% to 90-95% for Cd1. Beyond reproducing their expected biological behavior as ON and OFF switches, the riboswitches showed opposing transient maxima at ligand concentrations above 40 µM.

The Cd1 *apo* state transitioned into and saturated its *holo* state population before populating the antiterminator, and subsequently, a significant part of the *holo* state transitioned to the antiterminator. An opposing trend was observed for pilM. PilM started populating the terminator state shortly after beginning to populate the *holo* state, and even at high ligand concentration, a significant amount of *apo* state transitions into the terminator state. For pilM, a subsequent partial transition of the terminator to the *holo* state was observed. Based on their slow k_{off} rates, both riboswitches will not be influenced by dissociation events and are kinetically controlled. However, given the expected regulatory concentration reaching up to 100 µM ligand, orders of magnitude above the aptamers K_Ds, we argue that the Cd1 approaches saturation and exerts effectively thermodynamic control. This saturation-based assessment in high ligand concentration regimes allows an evaluation of the riboswitches based on the ratio between association time (t_{on}=(k_{on}*c_{Lig})$^{-1}$) and transcription time (with and without pausing delays). The approach of thermodynamic-like control was also observed as a reduced transcription speed dependence of Cd1, compared to pilM, that further decreased when pausing was introduced. Ligand concentration limits for the transition from kinetic to thermodynamic control were determined, and pilM was found to fully utilize the linear range of the ligand response curve associated with kinetic control. The median response concentrations of pilM, Cd1, and Cd1 with pausing of 33 µM, 13 µM, and 3 µM were determined. They indicate that if a pause site is confirmed for Cd1, a lower in vivo concentration than expected is likely to regulate Cd1. An innate difference between ON and OFF regulating riboswitches is when gene leaking can occur. While ligand-induced imperfect full OFF signaling results in residual basal gene expression, imperfect full ON signaling is reduced in strength by the residual false OFF signaling but does not leak gene expression. PilM, not required to minimize gene leaking through a high *holo* population, can utilize the full dynamic range that has a linear ligand dependence and can, as a result, maximize ligand responsiveness. Contrary, Cd1 cotranscriptionally reaches near full *holo* population, likely required to minimize gene leaking.

Project 2: ^{19}F NMR-Based Fragment Screening of RNA

This project investigated the ability to target RNA aptamers with a fragment-based screening (FBS) approach (Figure 2A). It utilized CPMG NMR experiments to observe enhanced T2 relaxation on ^{19}F-labeled small molecule fragment signals due to a binding-associated change in rotational correlation times. The use of ^{19}F detection has several advantages compaired to other NMR screening approaches. ^{19}F is a naturally abundant NMR active isotope with a high gyromagnetic ratio. ^{19}F labels are not prohibitively expensive and allow fast signal acquisition. When decoupled, one signal peak corresponds to one ^{19}F-fragment label. Additionally, ^{19}F signals have a sharp linewidth compared to the broad spectral range they fall on. Mixtures of multiple fragments can be screened in parallel without spectral overlap.

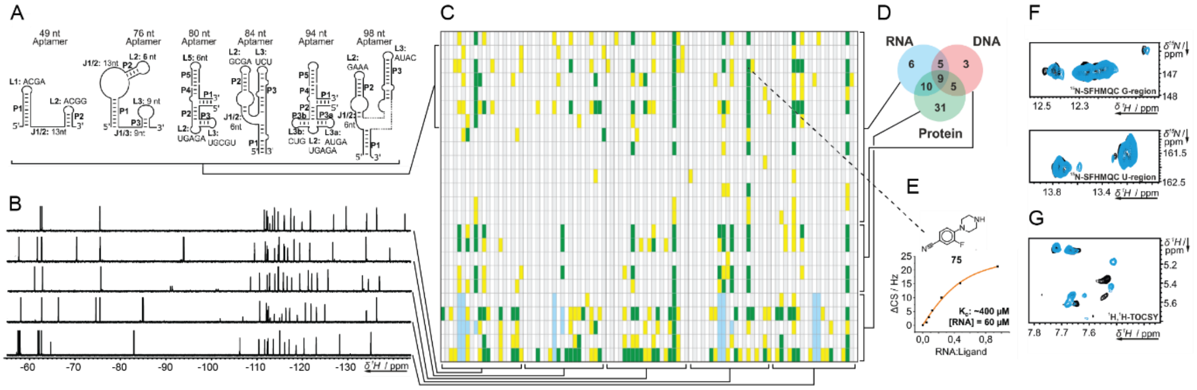


Figure 2 Project 2 overview: (A) Schematic secondary structures of the RNA aptamer targets investigated by 19F-FBS. Stems (P), loops (L), and junctions (J) are annotated, respectively. (B) 19F-1D NMR-spectra of the 19F-library fragment mixtures optimized to avoid signal overlap. (C) Interaction table of all fragments and biological targets screened. Hits were classified as no binding (Qbind > 0.67, alternating gray and white), weak (Qbind=0.66-0.33, yellow), or strong binding (Qbind < 0.32, green). For protein screens, hits for ~5% of the ligands could not be assigned (light blue). (D) Venn diagram of hit overlap between major target classes, RNA, DNA, and Proteins. 1H-1D-NMR-titration (E), 1H, 15N-correlation (F), and 1H,1H-TOCSY (G) of fragment 75 and 76 nt riboswitch aptamer as hit validation. Measurements with fragment are highlighted in blue. Figure modified from Binas et al. (2) and used in Figure 16, Figure 17, Figure 19, Figure 20, and Figure 22.

Having access to the iNEXT fragment library, we collaborated with O. Binas, T. Landgraf, C. Richter, and S. Sreeramulu to optimize a screening methodology using a subset of 102 ^{19}F fragments. The method was optimized for ideal relaxation times across the full ^{19}F spectrum and tested on three riboswitch RNAs. A relaxation delay of 200 ms was a good tradeoff between detecting weak binders and enabling the detection of residual signals for strong binders. The fragments were subdivided into 5 mixtures with minimal spectral overlap (Figure 2B). 14 RNA targets were screened against five pools of 20 small molecule fragments and compared to five DNAs and five proteins to assess the general applicability of the screening and the targetability of riboswitches through FBS (Figure 2C). V. de Jesus provided multiple RNAs to the screen and coordinated the screening of additional macro molecules prepared by other co-authors.

The integrity of the screening data was assured and monitored in collaboration with H. Berg. The observed hits in the fragment pools were corrected using macro molecule free references and classified as weak or strong binders based on an intensity ratio derived Q_{bind} value. In a computational analysis of fragment hit patterns conducted by K. Azzaoui and M. Blommers an initially qualitative increased targetablilty of RNA aptamers could be confirmed. Their correlation analysis allowed a clear differentiation between the hit pattern observed for RNA compared to Protein and DNA targets (Figure 2D). Further, RNAs fell into three clusters: Small, large, and aptamer. Small RNAs yielded a limited amount of hits compared to larger RNAs. Despite being similar in size, RNA aptamers had more hits than other large RNAs lacking a binding pocket. This increased hit rate indicates that riboswitch RNAs expose a larger diversity of structural interaction and have a higher targetability than other RNAs of similar size.

A subset of hits that showed strong binding to riboswitches was confirmed in single compound measurements and further analyzed. All hits could be confirmed in individual measurements, and no false positives were observed. The additional analysis showed clear indications of interaction, for example, a $^1H,^1H$-TOCSY spectrum of ZMP riboswitch exhibited substantial changes through fragment addition (Figure 2E, F, and G). Competition experiments with cognate riboswitch ligands showed that fragments can interact with multiple binding sites on the same target, including the cognate ligand binding pocket. O. Binas also determined the K_D of one strong binder to be submillimolar, 0.4 mM.

To demonstrate the feasibility of follow-up chemistry, a binding fragment was developed into a higher affinity binder by J. Martins under the supervision of A. Tröster. They connected a commercially available structural homolog of a binding fragment (benzamide P2D11) to the intercalator acridine. They tested the affinity of the linked fragment towards two terminator stems and one antiterminator stem. Linking a commercially available homolog of an initial hit to acridine allowed remarkable affinity increases compared to acridine-only measurements. In the best case, a 54-fold stronger affinity of 1.1 µM towards the SAM-Antiterminator from 59 µM (acridine only) was achieved. The synthesis of a novel low micromolar affinity binder through straightforward chemistry using readily available chemicals was possible.

This work showed that NMR is a highly capable tool to investigate the gene regulation of riboswitches and to find novel small molecules that bind to them. A comprehensive understanding of the mechanistic complexity of cotranscriptional riboswitch regulation and the feasibility of fragment-based drug discovery targeting riboswitch RNA were successfully established.

Zusammenfassung in deutscher Sprache

Diese Arbeit untersuchte cyklische Dinukleotid-Riboschalter (CDN) und erlangte ein detailliertes Verständnis ihrer Faltung während der Transkription sowie der Machbarkeit eines ^{19}F-NMR-Fragment-basierten Screenings, um RNA bindende Leitstrukturen für Riboschalteraptamere zu finden.

Eine vor Kurzem entdeckte Klasse von Riboschaltern reguliert Gene für die Umwelt, Membranen und Motilität (GEMM) in Bakterien, indem sie cyklische Dinukleotide (CDN), bakterielle sekundäre Botenstoffe, bindet. Umweltreize auf die Bakterien bewirken eine Anpassung ihrer internen CDN-Konzentration durch erhöhte Bildung oder erhöhten Abbau, was zu Änderungen des Lebensstils wie der Koloniebildung führt. Im Zentrum dieses regulatorischen Netzwerks erfassen Riboschalter die *in-vivo*-Konzentration und regulieren die für diese Lebensstiländerungen erforderlichen Gene. Sie spielen eine zentrale Rolle in der Physiologie und Virulenz von Bakterien. Diese Arbeit untersuchte den pilM-Riboschalter, einen EIN-Schalter, der die Pili-Expression in *Geobacter metallireducens* ermöglicht (Abbildung 1A). Um die kotranskriptionellen Faltungsprozesse von Riboschaltern tiefer zu verstehen, wurde die äquivalente experimentelle Arbeit von Tom Landgraf an dem Cd1-Riboschalter (*Clostridium difficile*), einem Flagellenexpression unterdrückenden AUS-Schalter, in einer Markov-Simulation kotranskriptioneller Riboschalterfunktionen kombiniert. Die Simulation zeigte das erwartete regulatorische Verhalten bei Lebensstil ändernden Ligandenkonzentrationen und bei typischen Transkriptionsgeschwindigkeiten.

Aufgrund ihrer hohen Relevanz für den bakteriellen Stoffwechsel waren und sind neuartige Strategien zur Regulierung von Riboschaltern ein gefragtes Unterfangen. Die zunehmende Bedrohung durch multiresistente Keime treibt diese Bemühungen zusätzlich voran. Medikamente, die mit CDN-Riboschaltern interagieren, könnten die *Vibrio cholera* und *Clostridium difficile* assoziierte Krankheitslast verringern. Im Gegensatz zu klassischen Ansätzen zur Arzneimittelentwicklung, bei denen große Verbindungsbibliotheken zur Suche nach Leitstrukturen genutzt werden, hat ein neuerer Ansatz, der kleine Molekülfragmente verwendet, um schwache, aber qualitativ hochwertige Wechselwirkungen zu finden, bereits zugelassene Medikamente ermöglicht. In dieser Arbeit wird dargelegt, wie die Kernspinresonanz (NMR) für die fragmentbasierte Arzneimittelentwicklung an RNA genutzt werden kann. Um die Eigenschaften von CDN-Riboschaltern als Angriffsziel zu beurteilen, war die pilM-RNA eine von 14 RNA-Zielen unterschiedlicher Sekundär- und Tertiärstrukturen, die mittels ^{19}F-NMR auf Bindung untersucht wurden. Hierbei wurden fünf Pools von 20 kleinen Molekülfragmenten verwendet und die Ergebnisse wurden mit fünf DNAs und fünf Proteinen verglichen. Es wurde gezeigt, dass RNA gezielt angegriffen werden kann. Zu den wichtigsten analysierten Parametern gehörten Ansprechverhalten verschiedener RNAs, die Fähigkeit der Bibliothek, RNA-Ziele anstelle von Proteinen und DNA anzusprechen, die Qualität der nachgewiesenen Wechselwirkung und die Fähigkeit, Treffer in fester bindende Leitverbindungen durch synthetische Verknüpfung zu entwickeln.

Projekt 1: Die kotranskriptionelle Faltungslandschaft cyklischer Dinukleotid-Riboschalter

Dieses Projekt nutzte die von C. Helmling entwickelten schnellen RNA-Präparationsverfahren, um RNA-Konstrukte unterschiedlicher Länge herzustellen. Eine erste Optimierungsphase ermöglichte die Herstellung bindungsfähiger unmodifizierter Wildtyp-pilM-RNA mit homogener Länge und hoher Ausbeute. Die Konstrukte wurden ausgehend von der Aptamersequenz kontinuierlich verlängert oder verkürzt, bis persistent keine Bindung mehr beobachtet wurde, um ein bindungsfähiges Fenster zu etablieren. Alle Konstrukte wurden mit nativen Liganden titriert, um bindungsassoziierte Veränderungen in ihren 1D-NMR-Spektren zu beobachten. Konstrukte von besonderer Relevanz wurden durch die Herstellung isotopenmarkierter Proben weiter analysiert und umfassende Imino-Zuordnungen von ligandenfreien und ligandengebundenen Konformationen waren möglich.

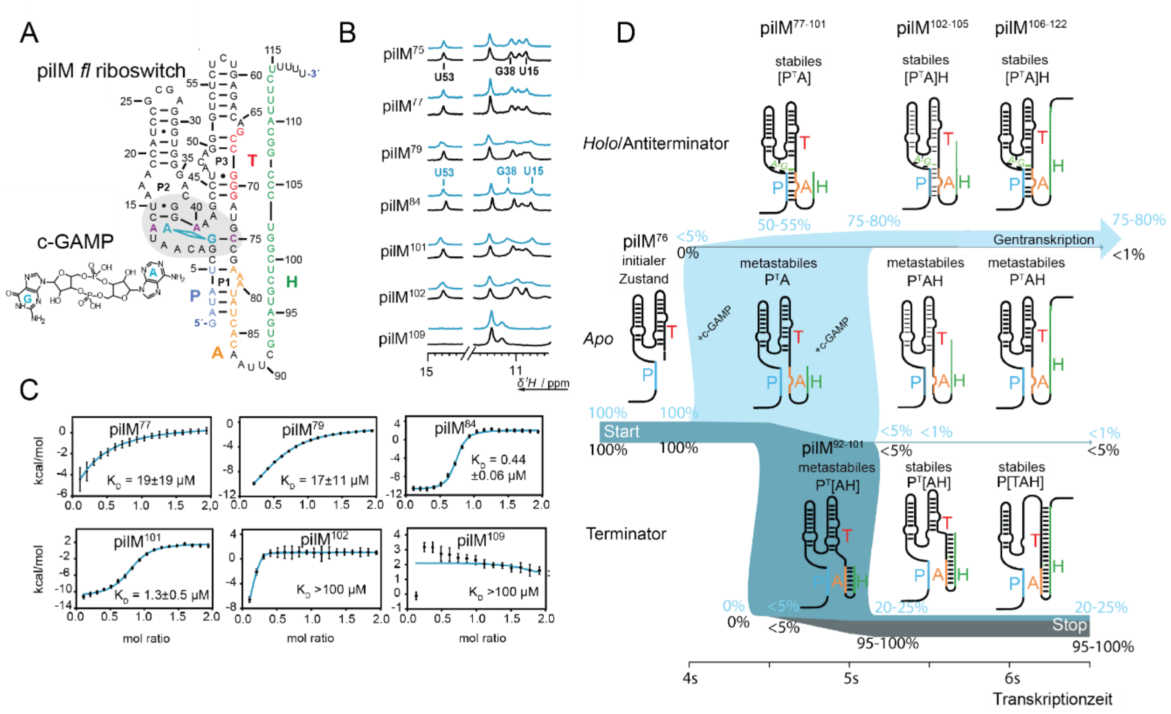

A
pilM fl riboswitch
c-GAMP
P1
P2
P3
P
A
T
H
5'-G
U-3'
B
pilM75
pilM77
pilM79
pilM84
pilM101
pilM102
pilM109
U53
G38 U15
15
11
δ 1H / ppm
C
kcal/mol
pilM77
KD = 19±19 µM
pilM79
KD = 17±11 µM
pilM84
KD = 0.44 ±0.06 µM
pilM101
KD = 1.3±0.5 µM
pilM102
KD >100 µM
pilM109
KD >100 µM
mol ratio
D
pilM77-101
pilM102-105
pilM106-122
stabiles [PTA]
stabiles [PTA]H
stabiles [PTA]H
Holo/Antiterminator
50-55%
75-80%
75-80%
pilM76 initialer Zustand
<5%
0%
Gentranskription
<1%
metastabiles PTA
metastabiles PTAH
metastabiles PTAH
Apo
+c-GAMP
100%
100%
Start
100%
100%
<5%
<1%
<1%
<5%
pilM92-101
metastabiles PT[AH]
stabiles PT[AH]
stabiles P[TAH]
Terminator
0%
<5%
20-25%
20-25%
0%
<5%
95-100%
Stop
95-100%
4s
5s
6s
Transkriptionszeit

Abbildung 1 Überblick über Projekt 1: (A) Sekundärstrukturen von pilM in ligandengebundener Konformation. Der gebundene Ligand ist blau markiert. Der 5´-Aptamer-Strang P basenpaart mit einem Aptamer-stabilisierenden Strang A, um das bindungsfähige Aptamer zu bilden. Der Schaltstrang T und der Terminatorstrang H bilden die Expressionsplattform. (B) und (C) zeigen Schlüsselkonstrukte bei der Festellung des Ligandenerkennungsfensters durch ^{1}H-NMR und ITC-Titration mit c-GAMP. Liganden enthaltende Proben sind blau gekennzeichnet. (D) Markov-Modellsimulationen der zeitabhängigen Besetzung kotranskriptioneller Faltungszustände. Die Populationsdichten der Zustände werden für eine Ligandenkonzentration von 100 nM (grau, basierend auf den in Figure 13B gezeigten Daten) und eine Ligandenkonzentration von 100 µM (blau, basierend auf den in Figure 13C gezeigten Daten) angezeigt. Bereiche der Abbildung, in denen sich die Populationsdichte überschneidet, werden graublau dargestellt. Die Balkendicke gibt eine relative Population im Bereich von 5% bis 100% an. Transkriptionszwischenprodukte mit einer Länge von 77–101 Nukleotiden sind bindungsfähig. Abbildung modifiziert aus Landgraf et al. (1) und in Figure 10, Figure 12 und Figure 15 verwendet.

Zu den wichtigsten Erkenntnissen gehörte, dass die verkürzten Versionen des Aptamers bindungsfähig waren, bis Nucleotid C75 entfernt wurde. Es ist beschrieben, dass C75 eine Watson-Crick-Basenpaarung mit dem Liganden eingeht. Die Konstrukte blieben ebenfalls bindungskompetent, selbst wenn die initialen Basen des Terminatorstamms in das Transkript integriert wurden. Bindung wurde ab Terminatorstammlängen, die ein C75-Basenpaar enthalten, nicht mehr beobachtet. Die Ergebnisse wurden durch die Beobachtung von P2- und P3-Helices für alle Konstrukte gestützt, während Hinweise auf eine P1-Helixbildung eine Ligandenzugabe erforderten. Die Ligandenaddition wurde durch mehrere Veränderungen im Iminospektrum und im Einklang mit homologen Kristallstrukturen beobachtet. Das bindungsfähige Fenster wurde zusätzlich in ITC-Messungstriplikaten bestätigt und stimmte vollständig mit den NMR-Ergebnissen überein. Das bindungsfähige Fenster für pilM ist 25 nt (Nukleotide) lang. Die Dissoziationskonstanten für pilM innerhalb dieses Fensters variierten von 19 µM bis 0,25 µM und erstreckten sich über zwei Größenordnungen (Abbildung 1B und C). Linienformanalysen der ITC-Injektionspeaks ergaben kinetische Ligandenbindungsparameter.

Während die Assoziationsratenkonstante (k_{on}) für alle Konstrukte annähernd gleich blieb, wurden Unterschiede in der Dissoziationskonstante (K_D) durch Unterschiede in der Dissoziationsratenkonstante (k_{off}) verursacht. T. Landgraf führte äquivalente Analysen des Cd1-Riboschalters durch. Er beobachtete, dass ein Fehlen einer P1-Helix in der Cd1-Sequenz dadurch erklärt werden konnte, dass Cd1 durch die Verlängerung des Aptamers um ein einzelnes Nukleotid die Aptamer-Affinität um das Zehnfache erhöht und die gleichen Affinitätsregime wie pilM erreicht. Die K_D, k_{on}- und k_{off}-Raten der beiden am stärksten bindenden Riboschaltertranskripte betrugen 0,25±0,04 µM, 17000±3000 $M^{-1}s^{-1}$ und 0,0042±0,0005 s^{-1} für pilM und 0,25±0,04 µM, 21000±300 $M^{-1}s^{-1}$ bzw. 0,0045±0,0005 s^{-1} für Cd1.

Beide Riboschalter wurden in einem Markov-Modell kotranskriptionaler Faltung untersucht. Das Markov-Modell basierte auf B. Fürtigs früherer Modellierung purinbindender Riboschalter. Das Modell simulierte die kotranskriptionale Faltung, indem es jede Transkriptlänge einem von drei möglichen Zuständen zuordnete: *apo*, *holo* oder Aptamer-konkurrierende Faltung (Abbildung 1D). Das Markov-Modell zeigte die Fähigkeit die Genregulation des Riboschalters und seine Abhängigkeit von der Transkriptionsrate, der Transkriptionspausierung und der Konzentration des Liganden vorherzusagen. Den Übergängen zwischen besetzbaren Zuständen wurden Raten zugewiesen, die auf der spezifischen Sekundärstruktur dieses Zustandes mit dieser Transkriptlänge und der gemessenen Bindungsaffinität basierten. Der Vergleich der beiden Riboschalter, einem AUS- und einem EIN-Schalter, mit demselben Modell ermöglichte den Vergleich der beobachteten Änderungen. Er bewies außerdem die allgemeine Anwendbarkeit des Modells für transkriptionelle EIN- und AUS-Schalter mit unterschiedlich langen ligandenbindungskompeteten Fenstern. Das Modell reproduzierte die erwartete biologische Regulation unter hohen und niedrigen Konzentrationen des Liganden. Die durch das Modell ermittelten Populationsdichten von *apo*, *holo* oder Aptamer-konkurrierende Faltung wurden als Konturdiagramme dargestellt. Diese Analyse ermöglichte die Beobachtung von Übergängen vom initialen *apo*-Zustand zu stabileren Zuständen in Abhängigkeit von Zeit und einer zweiten Variablen. Diese Konturdiagramme wurden zunächst verwendet, um die Untergrenze einer Basenpaar-Schließungsrate zu bestimmen. Diese Schließungsrate ist zur Beschreibung von Basenpaaren erforderlich, die sich bilden können, ohne Basenpaare an anderer Stelle in der Struktur zu brechen. Obwohl die Geschwindigkeit normalerweise auf Nanosekunden-Zeitskalen liegt, stellten wir fest, dass das Modell nicht beeinflusst wurde, wenn Geschwindigkeiten von 2,5 ms verwendet wurden.

Die Folgeanalyse verwendete Konturdiagramme mit beschriebenen *in vivo* CDN-Konzentrationen, Transkriptionsgeschwindigkeiten und Pausierungszeiten. Es wurde angenommen, dass die CDN-Konzentration je nach Lebensstil und entsprechender zellulärer Signalisierung 100 nM oder 100 µM beträgt. Die Transkriptionsgeschwindigkeit wurde als 20 nt/s angenommen und der Cd1-Riboschalter wurde mit und ohne zusätzliche Pausierung von 10 Sekunden an einer vermuteten Pausierungsstelle modelliert. Für jedes Konturdiagramm wurde einer dieser Parameter variiert. Die resultierenden Konturdiagramme ermöglichten die Beobachtung des zeitabhängigen Einflusses der CDN-Konzentration zwischen 100 nM und 100 µM mit einer Transkriptionsgeschwindigkeit zwischen 1 und 100 nt/s. Die Terminatorpopulation von pilM verringerte sich bei steigenden CDN-Konzentrationen von über 95% auf 20–25% (Abbildung 1D). Die gleiche Konzentrationsänderung erhöhte die Terminatorpopulation von unter 5% auf 90–95% für Cd1. Die Riboschalter reproduzierten nicht nur ihr erwartetes biologisches Verhalten als EIN- und AUS-Schalter, sondern zeigten auch gegensätzliche zwischenzeitliche Maxima bei Ligandenkonzentrationen über 40 µM.

Der Cd1-*apo*-Zustand ging in seine *holo*-Zustandspopulation über und sättigte sie ab, bevor er den Antiterminator-Zustand besetzen konnte, und anschließend ging ein signifikanter Teil des *holo*-Zustandes in den Antiterminator über. Bei pilM war ein gegenteiliger Trend zu beobachten. PilM begann kurz nach Beginn der Besetzung des *holo*-Zustandes mit der Besetzung des Terminator-Zustandes, und selbst bei hoher Ligandenkonzentration ging ein signifikanter Anteil des *apo*-Zustandes in den Terminator-Zustand über. Für pilM wurde ein anschließender teilweiser Übergang des Terminator- in den *holo*-Zustand beobachtet. Aufgrund ihrer langsamen k_{off}-Raten werden beide Riboschalter nicht durch Dissoziationsereignisse beeinflusst und sind kinetisch kontrolliert. Angesichts der erwarteten regulatorischen Konzentration von bis zu 100 µM Ligand liegen sie jedoch um Größenordnungen über den K_D der Aptamere. Wir argumentieren, dass sich Cd1 einer *holo*-Sättigung annähert und effektiv eine thermodynamische Kontrolle ausübt.

Dieses sättigungsbasierte Kontrollkriterium in Regimen mit hoher Ligandenkonzentration ermöglicht eine Charakterisierung der Riboschalter basierend auf dem Verhältnis zwischen Assoziationszeit ($t_{on}=(k_{on}*c_{Lig})^{-1}$) und Transkriptionszeit (mit und ohne Pausierung). Der Ansatz einer effektiven thermodynamischen Steuerung wurde auch in einer verringerten Abhängigkeit der Transkriptionsgeschwindigkeit von Cd1 im Vergleich zu pilM beobachtet, die bei Pausierung weiter abnahm. Es wurden Ligandenkonzentrationsgrenzen für den Übergang von der kinetischen zur thermodynamischen Kontrolle bestimmt, und es wurde festgestellt, dass pilM den linearen Bereich der mit der kinetischen Kontrolle verbundenen Ligandenreaktionskurve vollständig nutzt. Die Halbreaktionskonzentrationen von pilM, Cd1 und Cd1 mit Pausierung wurden als 33 µM, 13 µM und 3 µM bestimmt. Diese Werte wiesen darauf hin, dass die Bestätigung einer Pausierungsstelle für Cd1 eine niedrigere regulierende *in vivo*-Konzentration als erwartet wahrscheinlich macht. Ein inhärenter Unterschied zwischen AN- und AUS-schaltenden Riboschaltern besteht darin, wann eine residuale basale Genexpression auftreten kann. Während die ligandeninduzierte imperfekte AUS-Signalisierung zu einer verbleibenden basalen Genexpression führt, wird die Intensität der imperfekten AN-Signalisierung durch die residuale falsche AUS-Signalisierung lediglich verringert, resultiert jedoch nicht in fälschlicher basaler Genexpression. PilM kann, ohne die Notwendigkeit unbeabsichtigte Genexpression durch eine starke *holo*-Population zu minimieren, den gesamten dynamischen Bereich nutzen, in dem eine lineare Ligandenabhängigkeit besteht, und dadurch die Ligandensensitivität maximieren. Im Gegensatz dazu erreicht Cd1 kotranskriptionell nahezu die vollständige Populierung des *holo*-Zustandes, potentiell um falsch positive Genexpression zu minimieren.

Projekt 2: 19F-NMR-basiertes Fragment-Screening von RNA

Dieses Projekt untersuchte die Machbarkeit gezielter fragmentbasierter Screenings (FBS) an RNA-Aptameren (Abbildung 2A). CPMG-NMR-Experimente wurden genutzt, um eine verstärkte T2-Relaxation an ^{19}F-markierten Signalen kleiner Molekülfragmente durch eine bindungsbedingte Änderung ihrer Rotationskorrelationszeiten zu beobachten. Die Verwendung der ^{19}F-Detektion bietet im Vergleich zu anderen NMR-Screening-Ansätzen mehrere Vorteile. ^{19}F ist ein natürlich vorkommendes NMR-aktives Isotop mit einem hohen gyromagnetischen Verhältnis. ^{19}F-Markierungen sind kommerziell erschwinglich und ermöglichen eine schnelle Signalerfassung. Nach Entkopplung entspricht ein Signalpeak einer einzelnen ^{19}F-Fragmentmarkierung. Darüber hinaus weisen ^{19}F-Signale eine geringe Linienbreite im Vergleich zu ihrem breiten Spektralbereich auf. Mischungen mehrerer Fragmente können ohne spektrale Überlappung parallel gescreent werden.

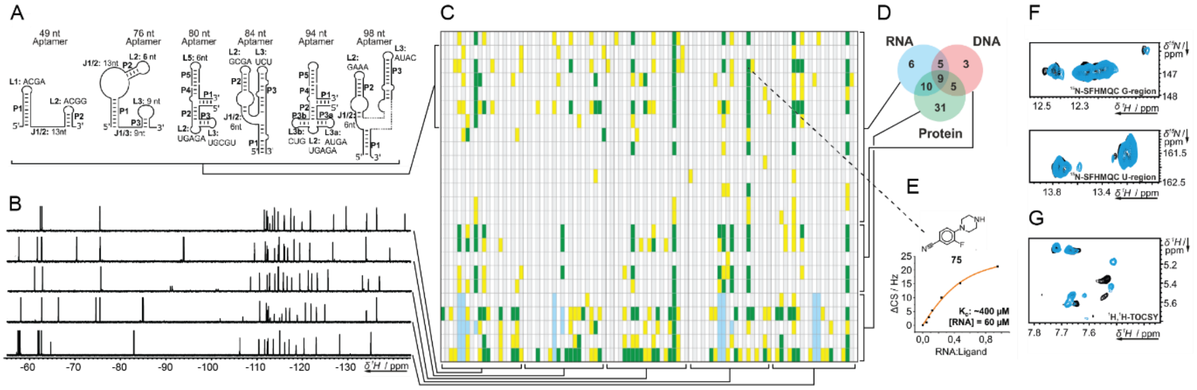


Abbildung 2 Überblick über Projekt 2: (A) Schematische Sekundärstrukturen der von ^{19}F-FBS untersuchten RNA-Aptamer-Ziele. Helices (P), Schleifen (L) und Knotenpunkte (J) sind jeweils mit Anmerkungen versehen. (B) ^{19}F-1D-NMR-Spektren der ^{19}F-Bibliotheksfragmentmischungen, optimiert zur Vermeidung von Signalüberlapp. (C) Interaktionstabelle aller untersuchten Fragmente und biologischen Ziele. Treffer wurden als keine Bindung (Q_{bind} > 0,67, abwechselnd grau und weiß), schwache (Q_{bind}=0,66–0,33, gelb) oder starke Bindung (Q_{bind} < 0,32, grün) klassifiziert. Bei Protein-Screenings konnten Treffer für ca. 5% der Liganden nicht zugeordnet werden (hellblau). (D) Venn-Diagramm des Trefferüberlapps zwischen den Hauptzielklassen, RNA, DNA und Proteinen. ^{1}H-1D-NMR-Titration (E), ^{1}H, ^{15}N-Korrelation (F) und ^{1}H,^{1}H-TOCSY (G) von Fragment 75 und 76 nt Riboschalter-Aptamer als Treffervalidierung. Messungen mit Fragment werden blau hervorgehoben. Abbildung modifiziert aus Binas *et al.* (2) und in Figure 16, Figure 17, Figure 19, Figure 20 und Figure 22 verwendet.

Nachdem uns ein Zugriff auf die iNEXT-Fragmentbibliothek ermöglicht wurde, arbeiteten wir mit O. Binas, T. Landgraf, C. Richter und S. Sreeramulu zusammen, um eine Screening-Methode unter Verwendung der Teilmenge von 102 ^{19}F-Fragmenten zu entwickeln. Die Methode wurde für ideale Relaxationszeiten im gesamten ^{19}F-Spektrum optimiert und an drei Riboschalter-RNAs getestet. Eine Relaxationsverzögerung von 200 ms zeigte eine gute Balance zwischen der Detektion schwacher Binder und der Detektion von Restsignalen für starke Binder. Fragmente wurden in fünf Mischungen mit minimalem spektralen Überlapp unterteilt (Abbildung 2B). 14 RNA-Ziele wurden gegen fünf Mischungen von 20 kleinen Molekülfragmenten gescreent und mit fünf DNAs und fünf Proteinen verglichen, um die allgemeine Machbarkeit des Screenings und eines gezielten Ansprechens von Riboschaltern durch FBS zu bewerten (Abbildung 2C). V. de Jesus stellte dem Screening mehrere RNAs zur Verfügung und koordinierte das Screening zusätzlicher Makromoleküle, die von anderen Koautoren bereit gestellt wurden.

Die Integrität der Screening-Daten wurde in Zusammenarbeit mit H. Berg sichergestellt und überwacht. Die beobachteten Treffer in den Fragmentmischungen wurden mithilfe makromolekülfreier Referenzmessungen korrigiert und auf Basis eines vom Intensitätsverhältnis abgeleiteten Q_{bind}-Werts als schwache oder starke Binder klassifiziert. In einer von K. Azzaoui und M. Blommers durchgeführten Computeranalyse von Fragmenttreffermustern konnte eine zunächst qualitativ erhöhte Ansprechbarkeit/Targetability von RNA-Aptameren bestätigt werden. Eine nachfolgende Korrelationsanalyse ermöglichte eine klare Unterscheidung zwischen den für RNAs beobachteten Treffermustern von den für Protein- und DNA-Zielen beobachteten (Abbildung 2D). Darüber hinaus ließen sich RNAs in drei Cluster einteilen: klein, groß und Aptamer. Kleine RNAs lieferten im Vergleich zu größeren RNAs eine limitierte Anzahl an Treffern. Trotz ihrer ähnlichen Größe hatten RNA-Aptamere mit Bindetasche mehr Treffer als andere große RNAs. Diese erhöhte Trefferquote weist darauf hin, dass Riboschalter-RNAs eine größere Vielfalt an strukturellen Wechselwirkungen aufweisen und eine höhere Ansprechbarkeit/Targetability aufweisen als andere RNAs ähnlicher Größe.

Eine Untergruppe von Treffern, die eine starke Bindung an Riboschalter zeigten, wurde in Einzelverbindungsmessungen bestätigt und weiter analysiert. Alle Treffer konnten in Einzelmessungen bestätigt werden und es wurden keine falsch positiven Ergebnisse beobachtet. Die zusätzliche Analyse ergab klare Hinweise auf eine Wechselwirkung, beispielsweise zeigte ein ^{1}H,^{1}H-TOCSY-Spektrum des ZMP-Riboschalters erhebliche Veränderungen durch Fragmentzugabe (Abbildung 2E, F und G). Verdrängungsexperimente mit nativen Riboschalter-Liganden zeigten, dass Fragmente an mehreren Bindungsstellen mit den Riboschaltern interagieren können, auch mit der Bindungstasche des nativen Liganden. O. Binas bestimmte außerdem, dass die K_D eines starken Binders submillimolar war, 0,4 mM.

Um die Machbarkeit von Folgechemie zu demonstrieren, wurde von J. Martins unter der Aufsicht von A. Tröster ein bindendes Fragment zu einem Binder mit höherer Affinität modifiziert. Ein kommerziell erhältliches Strukturhomolog eines bindenden Fragments (Benzamid P2D11) wurde mit dem

Interkalator Acridin verbunden. Die Affinität des Acridin verknüpften Fragments zu zwei Terminatorhelices und einer Antiterminatorhelix wurde getestet. Die Verknüpfung ermöglichte bemerkenswerte Affinitätssteigerungen im Vergleich zu reinen Acridin Messungen. Im besten Fall wurde eine 54-fach stärkere Affinität für den SAM-Antiterminator von ursprünglich 59 µM (nur Acridin) zu 1,1 µM erreicht. Die Synthese eines neuen niedrig mikromolaren Binders war durch einfache Chemie unter Verwendung breit verfügbarer Chemikalien möglich.

Diese Arbeit zeigte, dass NMR ein äußerst vielseitiges Werkzeug ist, um die Genregulation von Riboschaltern zu untersuchen und neue kleine Moleküle zu finden, die an Riboschalter binden. Es wurde erfolgreich ein umfassendes Verständnis der mechanistischen Komplexität der kotranskriptionellen Riboschalter-Regulation und der Machbarkeit fragmentbasierter Arzneimittelentwicklung an Riboschalter-RNA erlangt.

Chapter 1 Introduction

1.1 RNA

RNA (ribonucleic acid) is possibly how it all began (3–5). RNA is a biological polymer comprising unbranched nucleotide monomer chains. It can retain and process genetic information, and for that and many other reasons, RNA is hypothesized to be the first self-replicating molecule, the start of life. RNA is involved in coding, decoding, regulating, and expressing genes. RNA is used in rare cases to store genetic information, for example, in influenza and HIV viruses. However, RNA serves as the messenger RNA (mRNA) that acts as a bridge between the protein-synthesis machinery and the gene. Transfer RNA (tRNA) is an adapter between mRNA codons and amino acids. As with the RNA parts of the ribosome (rRNA), RNA may also have a structural function. RNA also serves as a regulatory molecule that binds to and prevents the transcription or translation of genes through switches encoded in the sequence of mRNAs (Riboswitches 1.2). RNAs can also act as enzymes that catalyze crucial biological processes. When these RNAs are not part of the ribosome machinery, they are called ribozymes.

1.1.1 Chemical composition

The RNA polymer consists of an alternating backbone of phosphates and ribose sugars (6). Each of the ribose sugars has an additional residue attached. This residue is named the nucleobase, a purine or pyrimidine base. There are two purine bases found in RNA, adenine (A) and guanine (G), and two pyrimidine bases, cytosine (C) and uracil (U). D-ribose is a pentose, and after cyclization into a five-membered ring, four hydroxy residues remain for new bond formations. In RNA, the first hydroxy group (1′) following the counting order for ring systems is bound to the nucleobase, while the backbone is formed by the 3′- and 5′-hydroxy groups binding phosphates. The 2′-hydroxy group of the ribose remains free. Its absence in the related molecule DNA is responsible for the letter D (Deoxyribose). RNA is elongated from nucleoside triphosphates monomers in a process called transcription. RNA is almost exclusively observed as a single-stranded molecule. However, RNA displays considerable double-helical character and can fold into various structures through intra-strand base pairing.

RNA has a broader range of base pairing options than DNA. In addition to the classical Watson-Crick pairings A:U and C:G, non-Watson-Crick pairings like U pairing with G are possible. The G:U base pair contains hydrogen bonds between the carbonyls on C2 of uracil and N1 of guanine and between N3 of uracil and the carbonyl on C6 of guanine. RNA may form tertiary structures containing additional interactions between nucleobases and the sugar-phosphate backbone. For some of these structures, enzymatic activities have been described, formerly considered a capability reserved exclusively for proteins.

1.1.2 RNA structure

In the formation of RNA structure, the secondary structure is generated depending on the primary sequence, and the interactions of preformed secondary structures eventually produce a tertiary structure (7). Since each base pair adds 1-3 kcal/mol of free energy to the final fold, the formation of RNA secondary structure is the major contributor to the total free energy of folding (8, 9). For example, a tRNA's primary sequence determines the construction of a secondary structure known as a ‘clover leaf’. The connection between two hairpin loops required for the final tertiary structure of tRNAs provides around 1.5 kcal/mol of free energy, the equivalent of one additional base pair (10).

Secondary structure research found great success in predicting functional RNA structures and tracking homologs in different organisms. Further, due to the large contribution of secondary structures and the computational accessibility of secondary structure prediction, a lot of predictive power can be found in comparatively simple models (7). Drug discovery focus can be found for both levels of structure. Secondary structure elements are better predictors of general targetability since multiple primary sequences can result in the same secondary structure, and the secondary structures can be separated into accessible (loops and junctions) and inaccessible (stems). Contrary, the dynamic nature of tertiary structure tends to be the cause of regulatory mechanism or catalyst activity because the tertiary structure can be influenced with little energetic effort, and tertiary refolding barriers can be passed at biological temperatures. A detailed understanding of the tertiary structure is required to gain insight into the underlying interactions and possible targetable domains. Both secondary structure and tertiary structure can be targeted.

Structure elements – secondary structure

Sections of the RNA where base pairing has occurred are called stems or duplexes. While these are commonly depicted as bands, these structures always represent double helices with opposing 5′-3′-directions. The ability to form additional noncanonical base pairs in RNA contributes to its propensity to fold double-helical structures. In RNA, non-Watson-Crick base pairs can be found in all possible combinations. rRNA has a high concentration of G:A and G:U. RNA chains have a greater probability for self-complementarity folding than DNA since such noncanonical base pairs and the two traditional Watson-Crick base pairs can exist.

RNA typically displays domains of base pairing but not DNA-like long-range helicity. RNA cannot assume a B-form helix commonly found in DNA because of the 2′-hydroxy groups in its backbone. Instead, RNA helices are more comparable to double-helical A-form structure. The helix contains two grooves. The major groove provides better sequence-specific interactions but is deeper and narrower than the minor groove. The RNA double helix is less suitable for sequence-specific interactions with proteins and small molecules than DNA (11). RNA interactions through sequence-specificity are more common in loop regions, bulges, and distortions caused by noncanonical base pairing.

Despite RNA molecules being single-stranded, they frequently contain double-stranded domains. RNA strands regularly fold back on to themselves to create base-paired regions in between sections of complementary sequences. The structures formed in this process can be broken down into basic structure elements. Some of these have multiple names and are used interchangeably in literature (6, 12). When complementary sequences are separated by short noncomplementary sections, stem-loop shapes, called hairpins, are formed (Figure 3). In stem-loop, the noncomplementary sequence forms a loop out from one end of the double-helical complementary stem resulting in the characteristic hairpin shape. For loop sequences, the unique characteristics of the loop aid in increasing the stability of stem-loop structures (13, 14). Stem-loops with a four nucleotide loop are called tetraloops, and specific loop sequences like UUCG are remarkably stable due to an additional G:U base pair, a cytosine phosphate contact, and base stacking interactions (15, 16). Other structural elements include internal or interior loops, a section of unpaired sequence on opposing strands connecting two stems. A bulge is a connection of two stems where one strand contains one nucleotide of a noncomplementary sequence. In the case of multiple nucleotides, it is called a bulge loop. When multi stems are connected, this is called a junction, a point where multiple stems meet. Commonly junctions contain short stretches of unpaired sequences that connect to the basis of the junction-forming stems. These sequences can be drawn as a stem-interrupted loop and are called the junction loop.

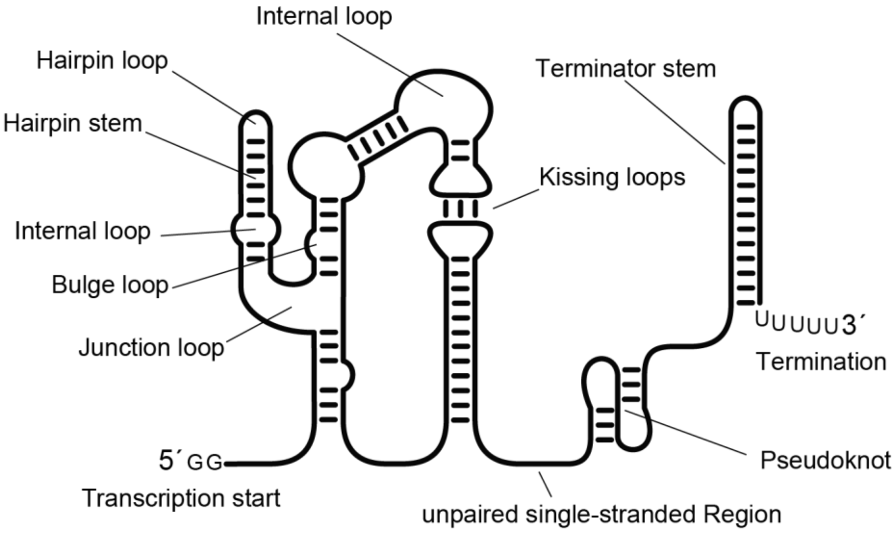


Figure 3 RNA secondary structure elements (12, 17).

The initial folding of these secondary structure elements and the higher rate of neighboring sequences contribution to hairpin formation can be explained through the higher frequency of collisions between complementary bases (9). This rate is based on the effective relative volume available to two bases, which depends on the effective concentration of one base in relation to the other. Small hairpin loops form faster because the effective relative volume correlates with the distance between the bases. Subsequently, longer effective distances will result in the formation of internal loops and bulges as well as connections of multiple stems. Secondary structural interactions usually arise before tertiary interactions, and tertiary interactions involve two or more of these secondary structure elements (9). The concept of decoupled processes for secondary structure formation and the formation of tertiary structures has been well-described for decades (18).

Tertiary structure

The exact line that divides secondary and tertiary structure has been debated in the past (9). Conceptually, the RNA molecule forms very energetically favored secondary structures. When these structures form and consequently bring their negatively charged backbones near each other, the resulting charge repulsion destabilizes them. Further interaction can be formed to reduce the total energy of the system. The interaction can be additional RNA-RNA interactions like base pairs over longer ranges that require the formation of secondary structure elements or other hydrogen bridging contacts, such as ribose phosphate interactions. Alternatively, other contributions, like interactions with ions or other molecules, can stabilize the tertiary structure. The possibility for forming these tertiary interactions arises from the primary sequence but is far more challenging to predict than the secondary structure (19). Extensive research has been conducted to cluster tertiary structures into shared interactions and function classes. Riboswitches, as an example of this, have been investigated for decades, and their specific structures, dynamics, and regulatory mechanisms are elusive to this day (20, 21). Newfound examples of riboswitches have frequently required single nucleotide variant analyses and high-resolution structures to understand unique aspects of the corresponding systems.

Another aspect of tertiary structures' relatively low energy landscape is that alternative tertiary structures are accessible at ambient temperatures and are often the basis of biological function (8). RNA tertiary structures with Watson-Crick base-paired regions of non-contiguous sequences are called pseudoknots (Figure 3). The minimal motif of a pseudoknot is an RNA structure of two helical segments connected by single-stranded regions or loops (17). The pseudoknot involves nucleotides from a larger loop region of a stem-loop binding a complementary sequence further up or downstream on the RNA strand. Further distances are associated with slower formation times. If the interactions are formed by two base pairing loops, they are called ‘kissing’ loops. Because each phosphate group of the RNA backbone holds a negative charge, positively charged ions significantly impact how RNA folds (9). Divalent ions, most commonly Mg^{2+}, have a significantly greater impact on tertiary folding than secondary structure. Since the initial formation of secondary structures results in an outward orientation of the phosphate groups, which causes charge repulsion between multiple secondary structures, these high negative charge densities contribute to forming metal binding sites and stabilize structures in high Mg^{2+} concentrations. Monitoring UV absorbance as a function of temperature at different Mg^{2+} concentrations is a standard method to distinguish between secondary and tertiary structures (8). Large RNA structures, like the ribosome, form tertiary structures with proteins. The negative charges of the RNA backbone are neutralized by proteins (22).

1.1.3 Biological activity of *cis*-acting RNAs

RNA structures can perform significant biological tasks and influence other regions on the same RNA strand in the absence of protein cofactors (23). The two primary examples of this are enzymatic activity and genetic regulation. The enzymatic activity without the need for protein cofactors is found in ribozymes and described below. Genetic regulation is accomplished by combining an RNA-based sensor motif with a structural motif that regulates transcription or translation. The sensor can be temperature-sensitive (Thermometer RNAs) or sensitive to the presence of other molecules. Regulatory motifs are terminator RNA stems or ribosome binding sites (RBS) that act as the starting point of transcription. In both cases, the sensor and the regulator interact and change their tertiary structure depending on the state of the sensor. The combination of sensor and regulator sequence is referred to as riboswitch, and sequences of both motifs can overlap. The amino acid lysine, the nucleobase guanine, the enzyme co-factor coenzyme B12, the metabolite glucosamine-6-phosphate, and bacterial second messengers are some examples of metabolites recognized by riboswitches. There is also an example of a temperature-sensitive riboswitch (24). Due to their great relevance, aspects of riboswitches are introduced in further detail in Chapter 1.2 and discussed throughout this work.

Ribozymes and RNA hydrolysis

RNA enzyme-like properties are called ribozymes (6). They have many characteristics of a traditional enzyme, including an active site, a binding site for a substrate, and a binding site for a co-factor, such as a metal ion. If ribozymes process themselves, they lack the property of a catalyst to reemerge unchanged at the end of the catalyzed reaction. For this self-processing reactivity, the ribozyme structure creates conditions favoring an RNA hydrolysis mechanism that also occurs naturally. In an alkaline environment, the 2′-hydroxy group of the ribose in the RNA backbone is deprotonated. The resulting negatively charged oxygen can attack the adjacent phosphate. A 2′-3′-cyclic-phosphate is formed, and an RNA strand-free 5′-hydroxy group is released. The RNA molecule can be fully broken down into nucleotides in this process, which can occur at each ribose of an RNA chain. Ribozymes like the hammerhead ribozyme use an Mg^{2+} ion in their active site to activate the 2′-hydroxyl of the RNA. This catalytic center drastically increases the speed of the transesterification reaction while operating at the physiological pH. During RNA splicing, other ribozymes carry out similar trans-esterification reactions to remove introns from the precursors of mRNAs, tRNAs, and rRNAs.

1.1.4 Transcription

Transcription is the process of RNA formation, copying the sequence from a DNA template (25). This process is carried out by RNA polymerases and uses ATP, CTP, GTP, and UTP as building blocks. This reaction is powered by the release and hydrolysis of the pyrophosphate at every elongation step. While eucrites contain four of five types of RNA polymerases, bacteria only contain one type of RNA polymerase, which produces all RNA except for RNA primers required for DNA replication. The transcribed DNA strand bound by the RNA polymerase is called the antisense strand. Transcription can be described as a four-step process of template binding, RNA chain initiation, chain elongation, and chain termination.

Template binding

RNA synthesis starts from specific sites. The sequences of these sites are called promoters. They are positioned upstream (5′) of the transcription start site. While the initiating nucleotide where transcription starts is called +1, promotors are typically positioned between -35 and -10. For example, the *e. coli* promotor consists of a pair of 6 nucleotide sequences that are highly conserved and separated by a stretch of 16-19 bases where the sequence is unimportant. The second sequence is most conserved and is called the Pribnow box. The mutations in the promotor region can influence the rate at which over three orders of magnitude transcribe a gene. Mutations that increase transcription rates are called ‘up mutations’, and decreased rates correspondingly ‘down mutations’. Increased GC content in the -10 region is generally associated with decreased promotor efficiency.

Initiation

After binding, the promotor domain transcription is initiated. In the process, the first nucleotide attacks the second and retains its triphosphate group. Due to the high-affinity interaction with the promotor region and the RNA polymerase, the initial steps of transcription frequently fail to ‘escape’ the promotor. This results in the release of short 10mer RNAs and is called abortive initiation. The RNA polymerase remains bound and reinitiates elongation.

Elongation and pausing

After ‘escaping’ the promotor, the RNA polymerase proceeds with a speed of 20-70 nucleotides per second (nt/s) (26–29). Pause sites can substantially reduce the average transcription speed by stalling the RNA polymerase, enabling specific folding pathways to occur (30). This pausing has been shown to be a necessary delay for gene regulation in bacteria (31). The RNA polymerase is highly processive and transcripts long stretches of RNA despite lacking a clamp-like structure. The transcription of the additional copies of RNA can occur before the first transcript is fully formed. For some RNAs, transcription is as rapidly initiated as sterically possible, which is equivalent to once per second. Since the polymerase moves at around 20-70 nt/s, several RNAs form simultaneously. This parallel transcription can be observed in electron micrographs as ‘arrowhead’ structures of continuously longer RNAs observed alongside a DNA strand template. RNA structures can form while the RNA is transcribed (32, 33). This process is called cotranscriptional folding (34).

Termination

Like initiation, RNA termination occurs at specific sites (35). Intrinsic and Rho-dependent termination are the two processes in bacteria that control appropriate transcript termination. Rho-dependent termination relies on the ATP-dependent RNA translocase Rho, which binds nascent RNA and dissociates the elongation complex. Intrinsic termination does not require protein interactions. This termination is directly encoded into the DNA sequence of the transcribed RNA. The intrinsic terminator sequence consists of a 7-8 nt long GC-rich terminator stem directly followed by a U-rich track that pauses transcription. The formation of the terminator stem inside the RNA polymerase exit tunnel is followed by the inactivation and dissociation of the elongation complex. The cotranscriptional formation of antiterminator structures can prevent the formation of the terminator stem if they persist throughout the pausing at the U-rich sequence of the terminator. When the formation of these antiterminator structures is influenced by the binding of cognate ligands to the RNA, this regulatory sequence is called a riboswitch.

1.1.5 Cyclic-di-nucleotides

Cyclic di-nucleotides (CDN) are tiny circular RNAs comprised of nucleoside monophosphates (NMPs). They are formed from different NTPs by an array of cyclases and broken down by phosphodiesterases. Cyclic di-guanosine monophosphate (c-di-GMP) and other CDNs have become one of the most prevalent and significant second messengers in bacteria since their initial discovery as an allosteric activator of bacterial cellulose synthase in 1987 (36, 37). While there is a great body of research on c-di-GMP, cyclic di-adenosine monophosphate (c-di-AMP) and guanosine-adenosine-monophosphate (c-GAMP) have only experienced limited research due to their more recent discovery in 2008 and 2012, respectively (38–42). They have been demonstrated to control the cell cycle, differentiation, pathogenicity, biofilm formation, and other functions (36). The majority of CDN signaling pathways regulate how well bacteria interact with abiotic surfaces, as well as with other bacteria or eukaryotic cells. They regulate the transition of many bacteria from their motile to sessile states, a key step of biofilm formation, and in pathogenic bacteria from a virulent state in acute infections to their less virulent but more resilient state. The formation and growth of biofilms in industrial and medicinal contexts could be managed by altering the CDNs signaling pathways in bacteria.

1.2 Riboswitches and aspects of riboswitch-based gene regulation

Riboswitches are RNA-based regulatory switches that modulate gene expression (21). 55 distinct classes of natural riboswitches combine the ability to detect and bind a low molecular weight connate ligand with the ability to change/stabilize its structure upon ligand binding. An extensive list of riboswitches and their structures until 2017 was procured by Lotz and Suess (43). The ligands range from ions, coenzymes, amino acids, nucleobases, nucleotide derivatives to CDNs, including c-di-GMP and c-GAMP (20, 40, 41, 44). Most riboswitches are located in the 5´-untranslated region (5´-UTR) of mRNAs between transcription start sites and the coding genes (45–47). Riboswitches are frequently discussed in different chapters to highlight specific aspects. This chapter intends to provide a general introduction to theoretical concepts of riboswitch regulation and specific aspects of ligand binding. In chapter 1.1.2 , general concepts regarding RNA structure formation were already discussed. Further reading can be found regarding CDN riboswitches, transcriptional riboswitches, cotranscriptional riboswitch folding in chapter 2.1 , regarding targeting riboswitches in chapters 3.1.1 and 3.1.3 , and in the results and discussion sections of this work. Examples of computational modeling of riboswitch are found in chapter 1.6.2 .

The discovery of cyclic-di-nucleotide riboswitches

In 2007 Weinberg *et al.* applied a computational pipeline on comparative genomics to bacteria and identified several riboswitch candidates, one of which was associated with the GEMM motif (48). The GEMM (Genes for the Environment for Membranes and for Motility) is associated with virulence in *Vibrio cholerae, Bacillus thuringiensis*, and *Clostridium difficile* (49–51). Riboswitches are also associated with utilizing metal ions as electron sinks under oxygen-deprived conditions in *Geobacter.* Their research found 21 covarying paired positions for a total number of 322 representatives of the GEMM motif. The consensus sequence contained two stems and was classified as probably a riboswitch. These stems are equivalent to the P2 and P3 stems described in this work. The consensus sequence also contains the bulge found in the P2 of the pilM riboswitch. A P1 stem equivalent structure was not found to be conserved throughout the representatives.

Construction and function

Riboswitches are structured RNAs. They contain two domains, a ligand-binding aptamer domain (sensor motif) and a regulating expression platform (regulator motif) (21). Riboswitches contain an aptamer that binds the inducer molecules with high affinity and specificity. Riboswitches can operate on the transcriptional level through transcript termination or on complete transcripts as translational switches where riboswitch structure can influence the accessibility of the Shine-Dalgarno sequence, mRNA splicing, and mRNA degradation (52). In response to changes in the intracellular concentration of a specific ligand, a commonly small inducer molecule, riboswitches regulate the downstream genes' expression level. The ligand binding stabilizes the aptamer structure and influences the cotranscriptional formation of downstream RNA structures. The ligand-free unbound aptamer is called *apo,* and the ligand-containing bound complex is referred to as *holo*.

In transcriptionally active riboswitches, these downstream structures feature a terminator stem. When a terminator stem is formed, the transcription is stopped, and the downstream genes are not transcribed. If an mRNA of a gene is not transcribed, the gene is not expressed and turned off. Riboswitches can act as ON or OFF switches depending on whether the ligand binding to aptamer prevents or favors the formation of the terminator stem. In ON switches, the aptamer and terminator compete, while in OFF switches, the aptamer and terminator coexist, competing with an antiterminator fold. The antiterminator forms in the absence of the ligand and persists during transcription, preventing terminator formation.

Structural rearrangements associated with ON and OFF signaling can be readily understood through the PATH representation(31, 53, 54). Every riboswitch contains four key strands: P, A, T, and H (Figure 4).

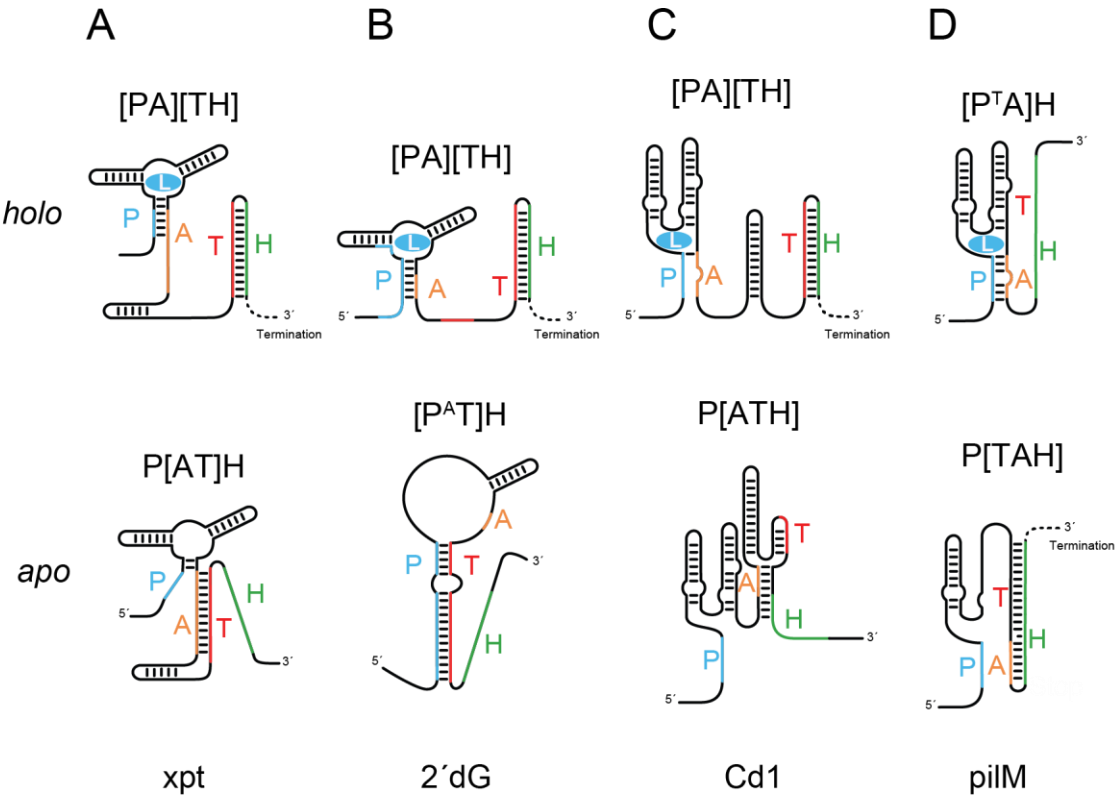


Figure 4 Schematic representation of the ligand-dependent secondary structure rearrangement of riboswitches. Comparing the ligand bound state (*holo*) and ligand free state (*apo*) of (A) xpt guanine-sensing riboswitch from *B. Subtilis*, (B) 2′dG riboswitch from *M. florum* (favored of two *apo* structures is shown), (C) Cd1 c-di-GMP riboswitch from *C. difficile*, and (D) pilM c-GAMP sensing riboswitch from *G. metallireducens* (1, 31, 53). Structural rearrangements are described by four distinct sequence segments in the mRNA chain: 5′-aptamer strand (P), aptamer-stabilizing strand (A), switching strand (T), and terminator strand (H). Ligand binding junctions are indicated in blue.

When high ligand concentration signals OFF, the aptamer [PA] is formed when the 5′-aptamer strand P base pairs with the aptamer-stabilizing strand A. [PA] is stabilized by ligand binding and forms the *holo* complex. In many three-stem junction riboswitches, [PA] is the P1 stem. If [PA] persists in cotranscriptional refolding, the switching strand T and the terminator strand H form the terminator [TH]. If low ligand signals ON, [PA] is unstable enough to refold, forming an antiterminator. The antiterminator prevents terminator formation (at least transiently). It utilizes strands from the PA. The A strand interacts with T to form the [AT] like in the xpt riboswitch (31) (Figure 4A). When T interacts with P, this forms an [P^{A}T] antiterminator as found in the ribD or 2′dG riboswitch (53–55) (Figure 4B). The antiterminator can integrate the upstream part of the H strand to form [ATH], like in the Cd1 riboswitch (56) (Figure 4C). The lower part of the H strand needs to remain unpaired to avoid exerting terminator-like behavior.

ON switches invert the logic of OFF switches. Ligand binding stabilizes an antiterminator fold, and the absence of ligand results in refolding of the A strand and terminator fold. An efficient way to accomplish these regulatory requirements is to locate the T strand upstream of the A strand. The PTAH sequence is, for example, found in the pilM and pbuE riboswitch (40, 41, 57) (Figure 4D). Stabil aptamer and antiterminator are represented as [P^{T}A], and the terminator is [TAH].

Since regulation occurs cotranscriptionally, the regulatory mechanism is limited to the transcription time and the sequential nature of transcription. The RNA continually adds bases that can influence the structural equilibria. While riboswitch aptamers bind their ligand with high affinities, the limited time frame of ligand-binding competent aptamer can dramatically increase the concentrations of ligand required to influence regulation. Equally, rapid folding structures with low kinetic barriers drive regulation if they are stable enough to persist throughout the transcription timeframe (58, 59). A more detailed introduction to the topic is found in chapters 1.2.1 , 1.2.2 , and 1.2.3 .

The non-equilibrium nature of this riboswitch regulation requires a detailed kinetic analysis and a deterministic simulation of the system to make conclusive statements about the regulatory mechanism and its dependence on ligand concentration, transcription speed, and other biological parameters (60).

Regulatory finetuning through competing folds, combined aptamers, and external factors

Moving on from simple regulation cases with bipolar *apo holo* equilibria, additional finetuning mechanisms have been described. An example of this fine tuning can be found in the add riboswitch from *Vibrio vulnificus* (24). The add riboswitch aptamer can fold into two *apo* states. The *apo*A state represents a typical aptamer that binds ligand and forms *holo* complex. *apoB* does not bind ligand. Reduced temperature decreased the dissociation constant (K_D) of *apoA*, because the add riboswitch regulates under thermodynamic control. This change would cause a drift away from a high dynamic range to a diminished range at low temperatures. A temperature-dependent equilibrium that favors aboB at low temperatures counteracts this trend and allows a high switching efficiency over an extensive temperature range.

Riboswitches can also act in tandem to achieve higher regulatory complexity (61, 62). An obvious example is two complete riboswitches ahead of a gene. The gene is regulated and dependent on two different cognate ligands. The use of OFF and ON switches is possible and enables Boolean logic. Even the multiple uses of the riboswitches that bind the same ligand can be used for fine-tuning. Riboswitches can be combined cooperatively (Figure 5C). Both riboswitches act independently to produce a combined outcome (Figure 5D). The tandem regulation is equivalent to multiplying their relative dynamic range. A gene output of one riboswitch of 10%, 50%, and 90% would result in a tandem output of 1%, 25%, and 81%. The gene output has a great dynamic range and is more digital. In some cases, two riboswitch aptamers can share one expression platform (Figure 5D). The aptamers can bind cooperatively and also produce a more digital genetic outcome (Figure 5A). The last example of tandem riboswitches uses the ligand-binding-associated conformation changes of an upstream aptamer not to influence an expression platform but instead interact with a downstream aptamer or another regulatory RNA like a ribozyme (Figure 5E).

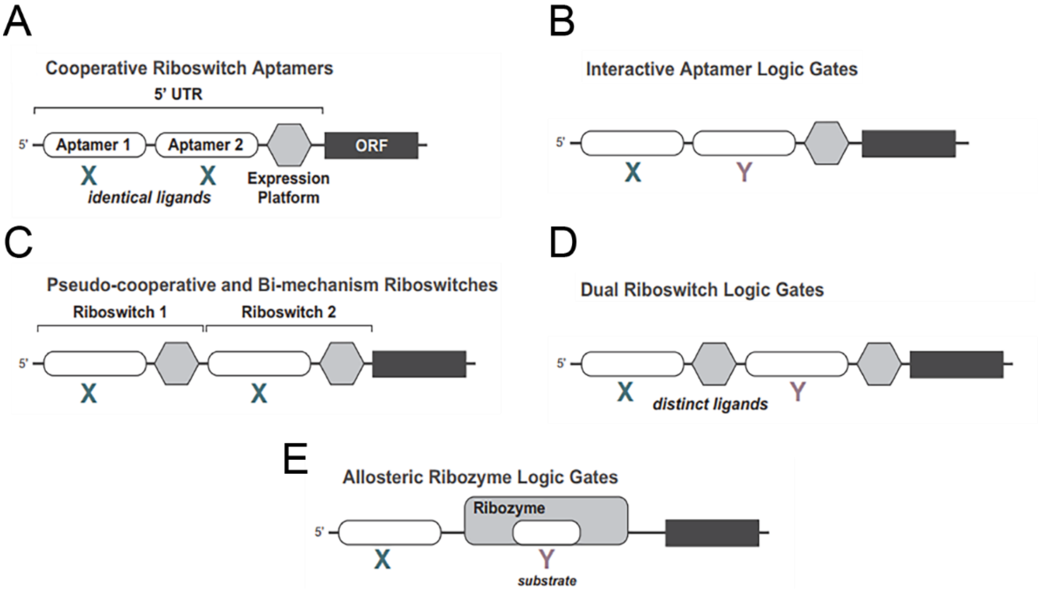

A
Cooperative Riboswitch Aptamers
5' UTR
5'
Aptamer 1
Aptamer 2
ORF
X
X
identical ligands
Expression Platform
B
Interactive Aptamer Logic Gates
5'
X
Y
C
Pseudo-cooperative and Bi-mechanism Riboswitches
Riboswitch 1
Riboswitch 2
5'
X
X
D
Dual Riboswitch Logic Gates
5'
X
distinct ligands
Y
E
Allosteric Ribozyme Logic Gates
Ribozyme
5'
X
Y
substrate

Figure 5 Previously known tandem architectures for riboswitches and their established functions. (A) Cooperative riboswitch aptamers carry highly similar aptamer domains that bind chemically identical ligands and associate with a single expression platform. Examples of this riboswitch architecture demonstrate cooperative ligand binding and a steeper dose-response curve. (B) Interactive aptamer logic gates are formed by two adjacent aptamers that respond to different target ligands, here depicted as X and Y, and associate with a single expression platform. Ligand binding by one aptamer affects the function of the adjacent aptamer. (C) Pseudo-cooperative and bi-mechanism riboswitches involve the tandem arrangement of independently functioning riboswitches that respond to chemically identical ligands. For a bi-mechanism system, each riboswitch operates with a different regulatory mechanism (e.g. one transcriptional and one translational). (D) Dual riboswitch logic gates involve the tandem arrangement of independently functioning riboswitches that respond to different target ligands. (E) Allosteric ribozyme logic gates involve allosteric regulation by an aptamer for the function of a ribozyme that requires a second distinct compound for its activity. Figure was upscaled, rearranged, and initially published by Sherlock et al. under Creative Commons Attribution License (62).

While riboswitch regulation is primarily believed to be independent of protein cofactors, protein-coupled regulatory systems have been found. Since the publications of Wickiser *et al.,* further research on the ribD riboswitch found additional regulatory mechanisms. The protein RibR binds to the ribD riboswitch sequence and counteracts the OFF regulation of the ribD riboswitch at high FMN concentrations (63–65). Pausing might also be critical for these novel ribR interactions. Similarly, the Vc2 riboswitch, a kinetically controlled translation riboswitch, involves a regulatory network enabling three distinct lifestyle states (49). Vc2 was found, in addition to regulating downstream genes, to prevent the degradation of an upstream non-coding RNA in a novel mechanism (66).

1.2.1 Kinetic and thermodynamic control of transcriptional riboswitches

The discussions in the literature surrounding kinetic control are a core component of riboswitch function. As a result, multiple criteria are used to define a transcriptional riboswitch as kinetically controlled. They all refer to a riboswitch aptamer not operating under equilibrium conditions. These non-equilibrium conditions are observed when the median response (T50) is larger than the K_D and the ligand-riboswitch complex has not had enough time to dissociate (criterium K1) or has not had enough time to bind/associate (criterium K2) (57, 67–69).

The criterium K2 is used in this work. It describes the inability of an aptamer to saturatingly bind its cognate ligand during transcription despite the ligand being present in concentrations orders of magnitude higher than the K_D of the aptamer. Aptamer saturation is prevented by transcription moving rapidly from transcribing the aptamer to sequences that refold the aptamer, preventing binding. These aptamer competing folds are terminator or antiterminator depending on the riboswitch being an ON or OFF switch. How fast the RNA polymerase needs to transcribe depends on the ligand-binding competent window length, the ligand concentration, and the association rate constant (k_{on}) of the ligand to the aptamer.

Similarly, thermodynamic control can be considered when the aptamer has enough time to reach complete equilibrium (criterium T1) or reaches 95% of the equilibrium distribution (criterium T2). At complete equilibrium, T50 equals K_D, the riboswitch is sensitive to changes around its K_D, and equilibrium is approached after transcription times larger than the inverse dissociation rate constant (k_{off}^{-1}). Suppose the upper limit concentration of a regulatory riboswitch ligand cotranscriptionally binds to a riboswitch aptamer and results in a 95% *holo* state population. In that case, it regulates very similarly to how a riboswitch would regulate at equilibrium in a large excess of ligand. The criterium T2 is related to K2. If a system under K2 kinetic control is provided with increasingly more transcription time or increasingly higher ligand concentration, it will transition to T2 thermodynamic control. T2 control can operate at a regulatory concentration above the K_D, but the closer it gets to reaching 100% equilibrium distribution, the more narrowly its dynamic range will match T1. The levels for criteria K2 and T2 were introduced by Wickiser *et al.* (57).

Wickiser *et al.* investigated the pbuE riboswitch (adenine sensing) and outlined the concept of kinetic riboswitch control. They suggest that if the riboswitch approaches saturation (>95% equilibrium level binding) with bound ligand during the transcription time of the binding-competent window, it should be classified as thermodynamically controlled. If the riboswitch does not reach a 75% equilibrium level, binding is kinetically controlled, and the 75-95% range should be considered mixed control. It is shown that the lifetime of the kinetic control regime is highly dependent on the concentration of the ligand. They are inversely proportional. A 10-fold increase in concentration results in a 10-fold reduction of the time frame that should be considered kinetic control.

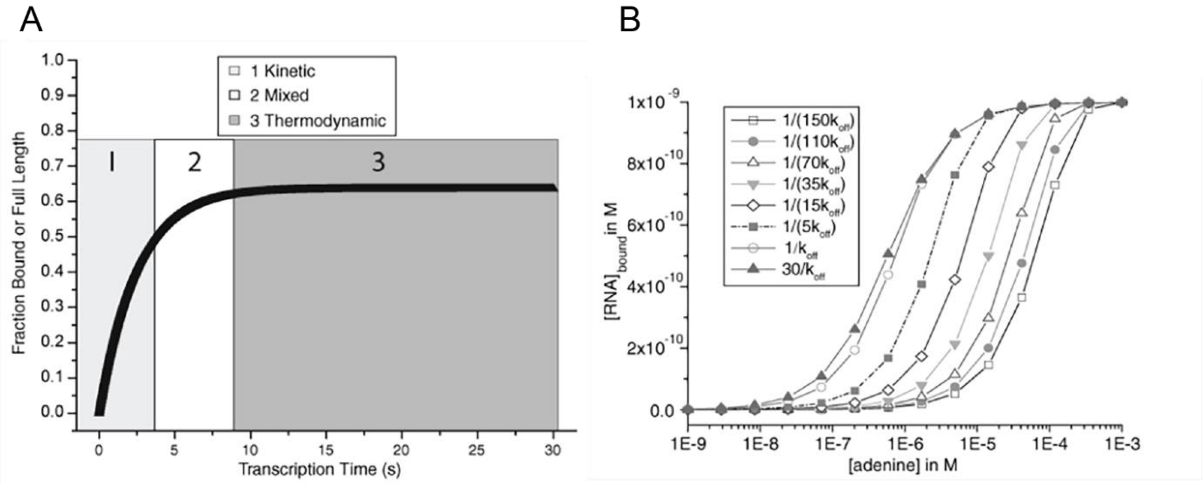


Figure 6 Kinetic simulation of binding and genetic decision using kinetic constants determined at 25 °C. (A) A sample set of data. The y-axis represents the fraction of bound aptamer and the fraction of full-length transcript. The purely kinetic, mixed, and purely thermodynamic zones of character are marked by the light gray [1], white [2], and dark gray [3] background. The partition between zones was set by defining the mixed region from approximately 75-95% of the fraction bound; time domains below and above this zone are designated kinetic and thermodynamic, respectively. (B) Simulated dependence of the extent of adenine binding to the riboswitch at 25 °C using the kinetic parameters reported in Figure 4 of Wickiser *et al.* (57). A 1 nM concentration of RNA is assumed. The different curves correspond to different values of ΔtRNAP, expressed as factors of $1/k_{off}$, where k_{off} is the dissociation rate constant. The maximum value of ΔtRNAP assumed in the simulation is $30/k_{off}$, or about 200 s. Under these conditions binding is at equilibrium, and median response (B50) is equal to the KD. Deviations from equilibrium, particularly at low adenine concentrations, are apparent when ΔtRNAP=$1/k_{off}$. As ΔtRNAP becomes progressively smaller, B50 moves to higher adenine concentrations. A noteworthy feature of the curves is their asymmetry and the abrupt transition to saturation binding when the system is under kinetic control. Figure reprinted adapted from Wickiser *et al.* (57).Copyright 2005 American Chemical Society.

Wickiser *et al.* state that a concentration of 10 µM adenine would result in less than a 1 s transcription time frame to be considered kinetically controlled. They also state that based on natural cellular concentrations of 30 µM adenine, which are considerably larger than the K_D of the riboswitch, they tentatively conclude that the riboswitch is kinetically controlled. At the 30 µM concentration, given the k_{on} rate of 263000 $M^{-1}s^{-1}$, the kinetical control level (75%) would be surpassed after 0.18 s, and the thermodynamic control level (>95%, criterium T2) would be reached after 0.38 s. Given that the ligand response window is 30 nt, it takes 1.5 s at a transcription speed of 20 nt/s (2 s based on their estimates) to transcribe this stretch of RNA. Based on their own classification, the pbuE riboswitch should be considered T2 thermodynamically controlled. The short transcription time at the *in vivo* concentrations is still long enough to bind and saturate aptamer. Yet Wickiser *et al.* state that without estimating the lifetime of two possible putative pause sites, no conclusions can be drawn about their impact on the thermodynamic and kinetic character of this riboswitch. Because prolonged pausing could allow the riboswitch to meet the T1 criterium, and no pausing could prevent equilibrium formation resulting in K1 kinetic control, Wickiser *et al.* are likely referring to these widely used criteria (K1 and T1) in their publication that suggested the use of other criteria (K2 and T2). They determined the k_{off} rate of the pbuE riboswitch as 0.151 s^{-1}. Therefore, equilibrium would be approached after 6.6 s, and even a short pause site could push the system to thermodynamic control.

A new method for sampling cotranscriptional RNA conformational ensembles was applied to the pubE riboswitch by Sun *et al.* (70). They modeled aptamer structure as stabilized by adenine binding (-5.3 kcal/mol) and found that the riboswitch aptamer transitions to over 95% ligand-bound aptamer within 1 s at 20 nt/s transcription speed and without pausing. This behavior also meets the Wickiser *et al.* criterium T2 of approaching saturation and being classified as thermodynamically controlled. Betay, in a review discussing Wickiser *et al.,* estimated transcription time to be around 0.5 s and attested kinetic control (K1), given the equilibration time of 6.6 s (68). Because T2 describes near-equilibrium conditions and K1 classifies the riboswitch as kinetic until equilibrium is reached, the two criteria inherently overlap.

The inverse k_{off} rates of Cd1 and pilM are 220 s and 270 s. Criteria K1 and T1 will classify Cd1 and pilM as kinetically controlled unless a remarkably long pause site is observed in future experiments. K1 and T1 are very limited in their descriptive power. They will correctly state that many transcriptional riboswitches are kinetically controlled and never reach equilibrium conditions before a regulatory decision is made. They provide no understanding of whether the system approaches a full population of the *holo* state, equilibrium-like conditions in large ligand excess. They, therefore, fail to describe the regulatory nuance at ligand concentration drastically higher than K_D, which can be understood through K2 and T2.

Criteria K2 and T2 described a system based on the degree of ligand binding. Given very slow k_{off} rates, fast k_{on} rate and high ligand concentrations compete with short transcription times. Depending on where the exact values of the three parameters lie, Cd1 or pilM could operate under kinetic, thermodynamic, or mixed control in the transition zone (75%-95%). Assuming a rigid regulatory range of, for example, 100 nM to 100 µM which the organism can realize to regulate a transcriptional riboswitch, a riboswitch could evolve a k_{on} rate, a specific length of ligand-binding window or pause sites to regulate a gene through kinetic or thermodynamic control. The regulatory range could be orders of magnitude above the K_D of the system. Under kinetic control, the genetic output would be linearly proportional to the ligand concentration, but at least 25% of the absolute dynamic range of the switch would not be utilized. Thermodynamic control would utilize at least 95% of the absolute dynamic range. Because thermodynamic control approaches saturation, the linear ligand response is lost towards the upper range of the regulatory range. These nuances are particularly interesting in the context of regulatory ON and OFF switches. Under K2, an ideal ON switch can regulate gene expression between 0 and 75% of all mRNA transcription initiations. An ideal OFF switch would regulate between 100% and 25% gene expression. The relative dynamic range of ON switches is huge and inherently limited for OFF switches under K2 kinetic control. OFF switches can increase their relative dynamic range through T2 thermodynamic control. ON switches only gain marginal increases in the absolute dynamic range. The ON and OFF switches studied in this work seem to abide by these considerations. (chapter 2.3.8 and 2.3.9)

1.2.2 Pausing and riboswitch function

Pausing has extensively been shown to be an essential requirement for riboswitch function (30, 31). Pausing has also been established to increase the sensitivity of transcriptional riboswitches by allowing more time for ligand to bind. The pausing delay thereby reduces the ligand concentration necessary to populate the aptamer (54, 57).

An early example of a pause site analysis was published by Wickiser *et al.* (55). They determined kinetic rates for three transcript lengths at four temperatures (Table 1). The transcripts represent riboswitch aptamer (ribD^{165}) and two extended versions of the folding accessible length at two pause sites (ribD^{200} and ribD^{230}). These two pause sits were observed in single-round transcription PAGE analysis and are both located before the antiterminator can form and compete with the aptamer P1. They determine the relative populations of full-length and terminated transcripts for 1 nM and 100 µM FMN for various conditions. The median response (T50) is obtained for FMN concentrations between 0.16 µM and 2.83 µM. At 1 µM FMN and 20 °C, this is equivalent to an association time ($t_{on}=(k_{on}*c_{Lig})^{-1}$) of 8 s for ribD^{200} and 65 s for ribD^{230}, matching the lifetimes determined for the pause sites at 10 s and 60 s, respectively. A transcription rate of 20 nt/s for a binding-competent window of 65 nt and the k_{on} rate of ribD^{165} at the same concentration yields a transcription time of 3.25 s and an association time of 5.5 s. Their analysis supported the hypothesis that FMN binding occurs primarily while RNA polymerase is paused, given that pausing durations matched the length of the corresponding association times at those sites, while the rest of the transcription occurred faster than the association time during elongation.

A recent analysis of the preQ1 riboswitch and RNA polymerase pausing found that RNAP stabilizes the native fold of the riboswitch, and binding of the ligand signals RNAP to release from the pause site (71).

1.2.3 Global k_{on} as a descriptor of ligand binding under kinetic control

Our computational analysis of the pilM and Cd1 riboswitch used global k_{on} rates to describe ligand binding, simplifying the underlying processes. The ligand binding to a riboswitch aptamer occurs in multiple steps with corresponding transition rates. The two consecutive steps of the system are ligand binding and binding pocket/aptamer refolding. While these steps are separate and can be individually observed, most methods determine a global K_D. This K_D describes the total affinity of the ligand to the aptamer, which stems from the surpassing of both kinetic steps and any other contributing effects. We think that the rates associated with this K_D, the global k_{on} and k_{off}, are a great descriptor of ligand binding and introduce limited errors to our model. Still, we want to present the position of a fellow scientist who suggests that ligand binding should be understood through the thermodynamic equilibrium of the initial binding step.

Burnouf *et al.* examined the mechanistic balances between TPP riboswitch refolding and ligand binding. They determined the isolated individual contributions of all involved rates. They found that the global K_D of 24 nM results from a k_{off} rate in the 10^{-4} s^{-1} range divided by a k_{on} rate in the 10^5 $M^{-1}s^{-1}$ range (Table 1) (72). The observed k_{on} is the same order of magnitude as the rates observed for the pilM aptamer in this work, while the pilM K_D and k_{off} rate were one order of magnitude higher. They also determined specific rates for the binding and the refolding. The binding (k_B) and unbinding (k_{UB}) rates were 26200 $M^{-1}s^{-1}$ and 0.09 s^{-1}, corresponding to a K_D 3 µM for initial binding. The rates for aptamer folding (k_F) and unfolding (k_{UF}) were 1.1 s^{-1} and 0.008 s^{-1}. They argue that total K_D is much lower than the common *in vivo* TPP concentrations. The k_{off} value obtained corresponds to a TPP release time of 1 h, much longer than necessary for complete riboswitch transcription, in line with kinetic regulation. Yet, their results also show a much higher K_D value for the primary TPP binding. As a result, the initial binding to the riboswitch would be thermodynamically controlled and sensitive to changes at *in vivo* TPP concentration.

They did not address that k_F is more than ten times faster than $k_{UB,}$ and the riboswitch would rapidly fold. Forming an isolated binding and unbinding pre-equilibrium with the given rates seems impossible, and we do not share their opinion that a thermodynamic pre-equilibrium can explain the concept of kinetic riboswitch control.

The use of global k_{on} rates, while simplifying the underlying multistep reactions, seems to allow an accurate description of the transitions of the system. The plausible transition states would readily convert into the final or initial state and not accumulate. An analysis of the multistage ligand binding kinetics to the rA71 riboswitch found an aptamer folding rate of 107s^{-1} for the transition of an initial binding state to a stable bound complex (73). In their fast mix-and-inject approach, they observed that the initial binding complex is most present after 10 ms at a relative population of around 20% and drops to baseline levels after 100 ms.

Bourouf's work was followed up by Guedich *et al.* (74). They also challenged the hypothesis of a thermodynamically controlled pre-equilibrium, stating that regulation of kinetic nature and the ability of the two riboswitches studies to detect a low TPP concentration appears to be correlated to the K_D for the initial step, not to K_D overall. They concluded from their results that kinetic regulation implies that the in-cell ligand concentration has to be of the order of the inverse product of the association rate constant and the transscription time. This relation can be restated as the association time has to be smaller than the transcription time.

Kinetically or thermodynamically controlled transcriptional riboswitch regulation is also discussed in chapter 1.6.2 , especially through the work by Beisel and Smolke, which analyzed this topic comprehensively (75).

1.2.4 k_{on} and k_{off} rates of different riboswitches

Given the great descriptive quality of kinetic parameters of riboswitch aptamers, previously published kinetic values (Table 1) hold information on riboswitch function, regulatory ranges, temperature dependence, and the influence of sequence truncations. The collected values from the literature are all small molecule binding riboswitches. The aptamers range in affinity from µM into the low pM range (seven orders of magnitude). The fastest k_{on} rates were observed for the FMN binding riboswitches. The highest value was observed for ribD with $3.99*10^5$ $M^{-1}s^{-1}$ at 310 K. At 298 K, the *Bordetella pertussis* FMN riboswitch has the fastest rate with $3.02*10^5$ $M^{-1}s^{-1}$. Very low k_{on} rates were observed for the 2′dG riboswitch, ranging from $1*10^3$ $M^{-1}s^{-1}$ to $8.9*10^3$ $M^{-1}s^{-1}$, depending on transcript length. Also low was the k_{on} rate at the second pause site of the ribD riboswitch $1.53*10^4$ $M^{-1}s^{-1}$ to $6.98*10^4$ $M^{-1}s^{-1}$, depending on temperature. The k_{on} of all compared riboswitches differs by less than three orders of magnitude. Larger differences in the K_D are caused by larger differences in k_{off} rate. This trend was also observed within the FMN riboswitches when comparing different *wt* aptamers and in a subsequent mutational analysis of the FMN riboswitch of *Fusobacterium nucleatum* (76). Small temperature-dependent increases in the K_D were observed for different lengths of the ribD riboswitch. Interestingly, the same temperature increase resulted in more significant increases in k_{on} and $k_{off,}$ which balanced, yielding smaller increases in K_D. Truncations of the Vc2 and 2′dG riboswitch show similar k_{on} and k_{off} rates for multiple truncation lengths until truncations result in large increases in K_D, where other rates could not be determined or were similarly impacted.

Table 1 Published kinetic values of various riboswitches. *values were calculated based on corresponding values provided in the related sources.

Authors	Method	Construct	Temp/K	K_D/µM	k_{on}/$M^{-1}s^{-1}$	k_{off}/s^{-1}
Burnouf *et al.* (72)	ITC	thiC (TPP riboswitch)	303	0.024	20000	0.00029
	SPR		298	0.0023*	126000	0.00029
Lang *et al.*(77)	stopped-flow fluorescence	U62AP thiM^{81} (TPP riboswitch)	298	0.495	86600	0.0429*
		U62AP thiM^{151} (TPP riboswitch)	298	0.420	81300	0.0341*
Wickiser *et al.* (55)	stopped-flow fluorimeter	ribD^{165} (FMN riboswitch)	288	0.0111		
			293	0.0112	183000	0.000205*
			298	0.0122	197000	0.00024*
			303	0.0145	423000	0.000613*
			310	0.0216	644000	0.001391*
		ribD^{200} (FMN riboswitch)	288	0.0273		
			293	0.0270	130000	0.000351*
			298	0.0307	182000	0.000559*
			303	0.0372	184000	0.000684*
			310	0.0559	399000	0.00223*
		ribD^{230} (FMN riboswitch)	288	0.0792		
			293	0.0893	1530	0.000137*
			298	0.1128	2060	0.000232*
			303	0.1366	2590	0.000354*
			310	0.1609	6980	0.001123*

Rode *et al.* (76)	stopped-flow fluorimeter	FMN riboswitch of *Fusobacterium nucleatum*	298	0.252	150000	0.0379
		of *Clostridium difficile* 630	298	0.321	68000	0.0220
		of *Bacilus anthracis* str.CDC 684	298	0.122	104000	0.0128
		Bacillus halodurans C-125	298	0.107	80000	0.00824
		Pasteurella multocida 36950	298	0.0535	128000	0.00687
		Desulfitobacterium hafniense	298	0.0445	77000	0.00342
		Bordetella pertussis CS Chromosome	298	0.0055	302000	0.00167
		Pseudomonas flourescens pf0-1	298	0.0032	225000	0.00072
Wickiser *et al.* (57)	2AP stopped-flow fluorimeter	pbuE^{70} (adenine riboswitch)	298	0.5809	263000	0.151
Greenleaf *et al.* (78)	force-extension curves	pbuE (adenine riboswitch)		3*	80000	0.2
Helmling *et al.* (54)	2AP stopped-flow fluorimeter	C74U dGsw^{78} (2′dG riboswitch)	298	430	1000	0.43
		C74U dGsw^{79}	298	307.7	1300	0.40
		C74U dGsw^{80}	298	70.0	5000	0.35
		C74U dGsw^{85}	298	76.6	4700	0.36
		C74U dGsw^{90}	298	35.6	8900	0.37
		C74U dGsw^{94}	298	56.6	6000	0.34
		C74U dGsw^{96}	298	66.0	5300	0.35
		C74U dGsw^{100}	298	66.7	5400	0.36
		C74U dGsw^{110}	298	66.7	5400	0.36
		C74U dGsw^{122}	298	41.8	5500	0.23
Gilbert *et al.* (79)	2AP stopped-flow fluorimeter (10 mM Mg^{2+})	C74U xpt-pbuX (guanine riboswitch)	303	0.017	150000	0.020
	(1.25 mM Mg^{2+})		303	0.77*	44000	0.034
Steinert *et al*. (31)	rapid sample mixing NMR of RNA reporter signals (600 µM hypoxanthine)		283		Biexponential: 0.21 s^{-1}, 0.02 s^{-1}	
Smith *et al.* (80)	Gel shift assay	*wt* Vc2^{110} (c-di-GMP riboswitch)		0.000011*	17000	0.00000018
Smith *et al.* (81)	Gel shift assay (10 mM Mg^{2+})	Vc2^{93}		3.4		
		Vc2^{94}		0.28		
		Vc2^{95}		0.063	33000	0.00022
		Vc2^{96}		0.0054	33000	0.00017
		Vc2^{97}		0.0046		
		Vc2^{98}		0.0059	8300	0.00005
		Vc2^{99}			22000	
		Vc2^{100}			20000	

1.3 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) spectroscopy observes the frequencies of nuclear spins to obtain information about the physical properties of molecules. While NMR can also be applied to solid-state systems with great success, it is uniquely capable of investigating chemical compounds in solution. NMR is the only method that allows obtaining atomic resolution structures of biological systems under *in vivo* equivalent solutions (82).

NMR spectroscopy is based on the phenomenon of nuclear magnetism and measures the interaction of nuclear spins with strong magnetic fields. The stronger the external magnetic field, the bigger the energy gap between the parallel and antiparallel spin state. Despite a powerful magnetic field 500,000 times stronger than Earth's magnetic field, the energy levels are still close enough to be similarly populated through Boltzmann distribution at room temperature. Only 1 in 10,000 spins contribute to the spins manipulated and detected in NMR. NMR spectroscopy rotates the z-magnetization with electromagnetic pulses out of alignment with the external magnetic field and into the transversal plane (XY). In the XY plane, the phase coherent magnetization precedes to precess and is detected. The magnetization is detected as an induced electric current in small wire coils positioned around the biological sample and inside the magnet. Different nuclear spins absorb and emit electromagnetic radiation at specific frequencies depending on their chemical environment. In several fields of bioanalysis, including organic chemistry, metabolome analysis, materials research, and structural biochemistry, NMR spectroscopy is employed.

NMR as a phenomenon was first identified in 1945. Nuclear spin energy levels were observed to split into multiple states in a uniform magnetic field. A transition between them could be created by emitting radio waves at a frequency matching the energy difference between the states. Since the development of pulsed Fourier transform NMR and multidimensional NMR spectroscopy in the late 1960s, the significance of NMR has significantly increased. Throughout the history of NMR, more physical processes like relaxation, nuclear Overhauser effect (NOE), coupling between spins, and the effects of partially oriented media were discovered and developed into more sophisticated methods. Examples of these methods allowed to determine the specific proximity of atoms (NOESY, COSY, and MQC), to determine the specific orientation (RDS) of domains of macromolecules, or to suppress solvent signals (JRE and watergate). These advances were accompanied by advances in hardware from superconductors that allowed strong magnetic fields over microelectronics that enabled highly sensitive detection and signal processing to sample preparation with isotope labeling and glass tubes that allowed further sample volume reductions.

1.3.1 NMR of RNA

NMR spectroscopy is an effective method for examining the structure and dynamics of RNA molecules in solution and their interactions with ligands such as proteins, other nucleic acids, low molecular weight molecules, ions, and solvent molecules (83). NMR spectroscopy has solved a large part of all three-dimensional nucleic acid structures. NMR spectroscopic studies yield a wealth of information. Due to inherent size limits, which constrain how much information can be obtained from a system, up to 100 nucleotides can yield information with intermediate resolution, and 50 or fewer nucleotides allow for high-resolution structure characterization. These NMR studies yield structural information, for example, about the base-pairing pattern. They cover standard and non-standard Watson-Crick base pairs and allow to verify and predict RNA secondary structure components and the identification of base pair dynamics. Dynamics data characterize conformational equilibria, such as those found between hairpin and duplex structures. NMR can yield information regarding ion binding to RNA and the specificity of those sites. NMR spectroscopy also allows resonance assignment of RNA and chemical shift analysis, enabling a delineation of secondary structure motifs. Hairpins and bulges are examples of such secondary structural motifs. The local structure and dynamics of RNA and its global structure can be generated through the analysis of residual dipolar couplings. And surfaces of RNA interactions with small molecule ligands, other RNA molecules, or proteins can be characterized.

RNAs are commonly formed from 5 elements hydrogen, carbon, nitrogen, oxygen, and phosphorous. While all these elements have NMR active isotopes (^{1}H, ^{13}C, ^{15}N,^{17}O, and ^{31}P), these active isotopes are the most common isotopes for hydrogen and phosphorus. ^{13}C and ^{15}N are naturally present in 1.1% and 0.37% relative abundance. Isotopically enriched RNA precursors are commercially available and can be used to increase the relative abundance of ^{13}C and ^{15}N. ^{17}O-modified precursors are not readily available and would likely be prohibitively expensive due to the lower relative abundance of 0.037% of ^{17}O. Beyond that, ^{17}O signals in larger biomolecules would suffer from extensive line broadening due to quadrupole transverse relaxation. Despite hydrogen and phosphorous being highly abundant NMR active elements, hydrogen contributes substantially more information during NMR analyses. Hydrogen atoms are more abundant and distributed throughout the RNA molecule. The hydrogen spin is the more sensitive probe, has a favorable ratio between peak and spectral width, and carries information on its direct environment in signal position due to chemical shift. The proximity to other hydrogen spins and hetero atom spin allows multidimensional NMR analyses. The innate properties of some hydrogen atoms to be exchanged with solvent hydrogen atoms and others to remain covalently bound allows the distinction of these atoms, making them a direct reporter of Watson-Crick base pairing, which limits the exchange of these hydrogen atoms.

1.3.2 Challenges of NMR analysis

The analysis of RNA using NMR is usually carried out in three steps. First, a large-scale NMR sample needs to be produced, followed by a comprehensive assignment of NMR resonance signals, and ending in the interpretation of additional NMR parameters that contain information on the three-dimensional structure or characteristics of an interaction of interest. The preparation of milligram amounts of RNA in an isotope-labeled (^{13}C,^{15}N) form can be carried out in multiple ways, most commonly through *in vitro* transcription (Chapter 5.2.3). The subsequent NMR sample preparation requires either D_2O as a solvent or spectroscopy signal suppression to mitigate the very strong H_2O

signal. The assignment of each NMR-active atom (^{1}H, ^{13}C, ^{15}N, ^{31}P) in a molecule to its respective resonance for RNA is more challenging than the protein resonance assignment. The chemical-shift dispersion in the spectra of RNA, a biopolymer composed of only four distinct nucleotides, is considerably lower than that of proteins, which contain 20 distinct amino acids. Additionally, there is only one dominant secondary structural component of RNA, the A-form helix. Numerous nucleotides from helical secondary structures encounter a similar chemical environment, leading to similar chemical changes. The only noncanonical structural features where chemical-shift dispersion is frequently seen are hairpins, bulges, or internal loops.

The unambiguous assignment of NMR resonances relies on reducing ambiguity in the system. This reduction can be accomplished through modification, marking, or removing one or several resonances of the RNA. The methods to accomplish these modifications can be grouped into preparative and spectroscopic. Preparative modifications are all physical changes to a sample or the preparation of a new sample that differs in one or more physical parameters. Examples of changes to a sample include changes in buffer, exchanging isotopes from ^{1}H to ^{2}D, adding small molecules or salt, Mg^{2+} ions (bound by specific binding sites in a large fraction of RNA structures), adding binding partners like proteins or other RNAs, agents that increase relaxation, or compounds that contain NMR active spins like isotopically labeled CDNs or ^{19}F-containing small molecule fragments. Additional sample preparations that can help to obtain a complete assignment involve isotopically labeled RNA bases, mutation of the RNA sequence, or truncations to the RNA sequence. Preparing and solving adequately sized truncated secondary structure elements and reconstructing the complete system is sometimes called the ‘divide and conquer’ approach. The method to prepare several RNAs of different lengths to analyze their transcriptional folding introduced by Helmling *et al.* has the innate benefit of providing information that links signals to specific lengths of the RNA (84). Incrementally longer RNAs contain information that can aid in assigning the corresponding signals. This method was also used throughout this work.

The use of multidimensional heteronuclear NMR experiments on completely or selectively labeled RNA molecules increases the resolution of NMR experiments by combining a proton chemical-shift dimension with one or two heteronuclear chemical shift dimensions. The dispersion of signals along multiple dimensions spectroscopically reduces ambiguity without the need to prepare additional RNA samples. Due to the increased effort, preparative means are only chosen when spectroscopy means have been tried and cannot resolve the ambiguity in question. The interpretation of NMR parameters such as NOE contacts, J couplings, residual dipolar couplings, and cross-correlated relaxation rates contribute to a higher understanding of the structure and aid in obtaining the unambiguous assignment. Therefore, obtaining a complete assignment and further high-level analysis of the system is not strictly sequential and instead commonly coincides.

1.3.3 Signal acquisition and assignment

NMR analysis of RNA molecules differentiates exchangeable and non-exchangeable hydrogen atoms. While exchangeable hydrogen atoms can be found on the base pairing side of the nucleobase and therefore act as a great reporter of solvent accessibility and conformational dynamics, the non-exchangeable hydrogen atoms are more abundant, can be measured in D_2O, and contribute a wealth of structural information. When imino protons are engaged in hydrogen bonding or protected from the exchange with the bulk solvent water, their signals provide essential information about base pairing in the RNA molecule (85). These signals are present in guanosine and uridine residues and are observed between 10 and 15 ppm. Contrary to the NMR of proteins, presaturation of the H_2O signal cannot be employed for solvent suppression to reveal exchangeable base proton signals in RNA due to the rapid exchange of imino and amino protons with water protons. The jump-and-return pulse pattern is helpful for solvent suppression of RNA samples despite the unwanted baseline distortions (86). These jump-and-return pulse sequence shortcomings can be solved using WATERGATE or other solvent suppression techniques (87).

Imino protons from canonical base pairs, such as A:U and G:C, typically exhibit chemical shifts of 13-15 ppm and 12-14 ppm, while imino protons from noncanonical base pairs, such as the G:U base pair are shifted upfield to 10-12 ppm. The RNA sequence significantly impacts how the imino proton signals shift due to its influence on the chemical environment. Imino proton signals can provide an understanding of the secondary structure of RNA without the requirement to solve the complete structure. Additionally, monitoring temperature-induced changes in the imino proton spectra may be used to examine the conformational stability of RNA molecules. Cytidine amino proton signals are observable when engaged in hydrogen bonds, whereas guanosine and adenosine amino protons are rarely observed. Scalar couplings across an NH-N-type hydrogen can be investigated by performing an HNN-COSY experiment on ^{15}N-labeled RNA (88). 2D NOESY spectra are used to assign the imino proton signals progressively along the RNA stem and in line with the secondary structure of RNA. In A:U base pairs, there is a strong NOE cross peak between the uridine H3 and the adenosine H2. NOE cross peaks can also be observed between the amino protons of cytidine and the guanosine H1 in G:C base pairs. A very intense NOE cross peak between the guanosine H1 and the uridine H3 can be used to identify G:U base pairings, whose imino proton signals resonate at 12-10 ppm. To distinguish the imino proton signals from guanosine and uridine, heteronuclear multi-quantum coherence (HMQC) spectra of RNA can be acquired. Imino proton signals of guanosine and uridine are distinguished because their ^{15}N chemical shifts are usually 10 ppm apart, at around 150 ppm and 160 ppm, respectively. The minimum requirement for the spectra is a ^{15}N labeling of either the G or the U bases, but usually, a more comprehensive labeled scheme is chosen to allow for additional NMR experiments on the same sample.

D_2O is a suitable solvent for NMR measurements of non-exchangeable protons because signals adjacent to the water signal can be detected without needing a water suppression pulse technique. Repeated lyophilizing of the solvent water and rehydration with D_2O is performed to exchange the protons in the RNA sample completely. In fully ^{13}C-labeled RNA samples, C-edited experiments may effectively suppress the water signal and allow NMR analysis without needing D_2O exchange. The overlapping signals of H2′, H3′, H4′, H5′, and H5″ are found at 3.5-5 ppm. There is usually no overlap between H1′ and the aromatic protons of the nucleobases. The analysis of NOE connectivities is used to determine the sequence of bases and ribose H1′ signals in helical A-form RNA. Strong NOEs may be observed between H2 *i* and H1′ *i*+1) in helical A-form RNA because their separation is less than 4 Å, allowing their sequential assignments. The H1′ of residue *i* exhibits NOE cross peaks to the H8/H6 of residue *i* (intra-residue) and residue *i*+1 (inter-residue). Spin-diffusion-mediated NOEs are generated by long mixing times of >400 ms. They aid the sequential assignment with stronger cross peaks since the inter-residue H1′-H8/H6 distance in helical A-form RNA is more than 4 Å. Sequential assignment is further aided by the inter-residue NOE cross peaks between base protons. NOEs between base protons and the sugar protons H4′, H5′, and H5″ can be observed but are typically challenging to identify due to low dispersion.

Connectivities between non-exchangeable base protons and exchangeable imino protons are crucial for the NOE-based assignment mechanism. Using NOE connectivities to its amino protons, the H5 proton of cytidine may be determined. The NOE cross peaks of pyrimidine H5 and H6 signals in NOESY experiments can be verified by TOCSY experiments. The difference in the chemical shifts of the C5 resonances in a ^{13}C-^{1}H SQC can be used to distinguish between the signals of cytidine and uridine. NOE connections to the uridine imino protons can be used to assign adenosine-derived H2 signals or they can be identified in ^{13}C-^{1}H SQC (150 ppm in 13C).

1.4 NMR-based methods for drug discovery

Many RNA targets fall within the size limit that can be studied through structural NMR (7, 89). The two benefits of NMR as a screening tool are pinpointing the precise binding site of small compounds, even ligands with affinities in the millimolar range, and overserving the conformational changes generated by ligand complexation (90). Different NMR methods have been developed to aid in the search for structure-activity connections and to develop new macromolecule binders from fragment libraries (91–95). One-dimensional (1D) and two-dimensional (2D) NMR spectroscopy are often used in drug development.

A molecule's composition, atom types, chemical environment, and bonding patterns may all be learned from 1D NMR data. Nuclear Overhauser effect spectroscopy (NOESY) and correlation spectroscopy (COSY), two 2D NMR methods, can provide more details regarding the connectivity and spatial interactions between atoms in a molecule but usually require longer acquisition times. The long acquisition time limits these methods to a throughput of one sample measurement per day, commonly containing a mixture of ten compounds (91). Higher throughput can be accomplished in NOE-pumped 1D spectra (240 compounds/day) or in $^{1}H/^{15}N$ HSQC spectra which can be acquired in minutes despite being two-dimensional (1000 compounds/day). A disadvantage of HSQC spectra analysis is that it is not ligand observed. Instead, changes to macromolecule signals are observed. While this provides direct information on the interaction site, it introduces a size limit of 40 kDa due to signal boarding at larger molecule masses. Furthermore, the screening of mixtures will necessitate control experiments to determine which small molecule is interacting with the macromolecule.

Other ligand-based methods for discovering and quantifying RNA-ligand interactions compare the resonance of free and bound ligands (96–101). These ligand-based approaches are especially helpful in the medium-low affinity range, with K_D higher than 100 µM, and rely on transferring NMR parameters from the RNA-bound to the ligand-free state (102). These NMR techniques enable the screening and detection of interaction systems in the fast-exchange regime, with the dissociation rate constants (k_{off}) ranging from 1000 s^{-1} to 100 000 s^{-1}. Ligand-based NMR methods have three key advantages. Macromolecule labeling is not necessary because only the ligand signals are observed, the NMR size limit is circumvented, very high-molecular-weight RNAs even aid analyses with their fast signal relaxation allowing easily observable ligand signals, and thirdly only micromolar amounts of the macromolecule are required to observe effects on the ligands, which are provided in 5-100 fold excess (103). There are multiple methods for ligand observed screening. They utilize NOE effects, saturation transfer difference (STD), or increased transversal relaxation during the interaction of a binder with the macromolecule.

NOE screening methods are based on transferred-NOESY (tr-NOESY) (104). When bound, the ligand behaves like a component of the macromolecule and exhibits the corresponding NOE behavior. Since the ligand retains these NMR characteristics after dissociation, the free ligand spectrum provides information about the ligand in the bound state. The first examples of the effect were observed by

Albrand *et al*. (105). More advanced methods can be used to investigate two or more ligands that are bound in close proximity through inter-ligand Overhauser effect (ILOE) (106, 107). These methods are conceptually based on experiments to investigate intermolecular NOEs of a small molecule with interacting residues of macromolecules (108, 109). One of the earliest examples of this elucidated the interaction between ethidium bromide and a poly-U-RNA (110). Other advanced methods that were more recently developed observe subsequent binding to the same macromolecule through inter-ligand NOE for pharmacophore mapping (INPHARMA) (111).

Saturation transfer difference (STD) NMR spectroscopy is commonly used to identify novel binders. It is a potent tool to confirm the interaction between small molecules and their receptor, allows molecular-level insight into the binding process, and gives a detailed understanding of which part of the ligand is directly involved in the binding process (103, 112). This method can be used on fast-exchanging systems, commonly small molecules with weak binding affinities. The corresponding K_D values are typically in the millimolar to micromolar range. For STD NMR, a macromolecule resonance is selectively saturated by applying a pulse train in the range of seconds. Spin diffusion and intra-molecular NOEs rapidly spread the saturation throughout the macromolecule. The spin diffusion is followed by a magnetization transfer through intermolecular NOE from the macromolecule to the ligand protons. Intermolecular saturation transfer and chemical exchange occur while the interacting ligand is bound to the macromolecule. Different nuclear spins of the ligand will experience different amounts of saturation transfer based on the interaction time and their proximity to the macromolecule during the interaction.

In practice, a sample having a nanomolar to a micromolar concentration of macromolecule and a 100-1000-fold molar excess of ligands is used to collect two distinct 1D-NMR spectra. One corresponds to an off-resonance experiment, recorded as reference, and involves irradiating a signal-free area (commonly 40 ppm). The other experiment, on-resonance, saturates a frequency with a signal exclusive to the macromolecule. Typical ranges are 6.5 to 8.5 ppm and 0 to 2 ppm. The off- and on-resonance spectra are subtracted to yield the STD spectrum. This spectrum only contains the signals of the ligands that experienced saturation transfer. It takes relatively little time, on the order of 100 ms, to transfer saturation from the macromolecule to the ligand (102). As a result, if the ligand's k_{off} rate is rapid and a high excess of a ligand is present. Stronger STD signals will be observed for weaker binders. These effects taper off at very low K_D values around 10 mM, where STD signals become extremely weak due to insufficient saturation transfer.

A popular 1D saturation transfer technique for fragment-based screening is WaterLOGSY (Water-Ligand Observed through Gradient Spectroscopy) (101–103, 107). WaterLOGSY is based on the NOESY experiment, and like STD methods, it utilizes an intermolecular NOE, and spin diffusion transfers magnetism. Its unique feature is the inclusion of water molecules in the transfer mechanism. WaterLOGSY, first introduced by Dalvit *et al.,* selectively transfers a part of the large bulk water magnetization to the macromolecule and subsequently to the ligand via the binding interaction (97, 101). A water-macromolecule-ligand or macromolecule-ligand complexes, whose rotational correlation times result in negative cross-relaxation rates, display a negative NOE. Non-binding small molecules will tumble significantly faster, resulting in a positive NOE. As a result, the resonances of non-binding molecules in this experiment have the opposite sign and are often weaker than those of the interacting molecules. WaterLOGSY was found to be 3.2 to 16 times more sensitive in three test systems than STD measurements (113). Increased sensitivity is directly correlated with a reduction in measurement time and the overall cost of the corresponding screening.

The 1D-NMR screening method used in this work was NMR relaxometry. For this fragment-based screening, Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences are employed. A spin-lock field causes relaxation to be active, suppressing homonuclear scalar coupling development. Fast relaxing signals are eliminated from the spectrum when this delay is set to several hundred milliseconds. A decrease in the ligand's T2 relaxation time, indicative of the complex formation, results in line broadening and a decrease in signal intensity(102, 103). The ligand signal decrease is compared to a reference measurement that does not contain the macromolecule but has the same CMPG pulse durations applied. The data quality can be further improved by correcting these values with additional measurements that contain a minimal relaxation delay. Further details are discussed in chapter 3.2.1 .

CPMG-based FBS has several advantages. The method is macromolecule agnostic. No excitation of a specific macromolecule frequency is required. The independence of macromolecule-specific parameters allows the measurement of several different macromolecules without changes to the methodology and requires only a single set of macromolecule-free reference measurements reducing total measurement time compared to NOE transfer screening. The measurements do not use a large excess of ligands, which beyond cost savings, allows for multiple compounds in ligand mixture samples. The results are easier to interpret compared to WaterLOGSY. While CPMG measurements only experience signal loss due to enhanced relaxation, WaterLOGSY signals are comprised of the overlap of free and bound signals with opposing signs and the resulting signal is influenced by the relative ratio of free to bound ligand, water residence time and rotation correlation time (97). While rotation correlation time also influences CPMG measurements for each macromolecule hits with a more substantial degree in signal intensity will correspond to a higher affinity binder (114). Furthermore, Aretz *et al.* showed that similar inferences about the interaction strength could not be made based on STD experiments (115). The use of isotopically labeled compounds can further increase CPMG screening efficiency. Research frequently relies on ^{19}F-labeled compounds due to their favorable combination of a wide spectral range and sharp signals (116). In this work, a ^{19}F-containing subset of the iNEXT library was distributed into 5 mixtures and subsequently screened (Chapter 3 : ^{19}F NMR-Based Fragment Screening of RNA).

There are additional examples of less commonly used techniques that use changes in diffusion speed to confirm binding (DOSY) or paramagnetic labels that cause relaxation enhancement (117, 118). These techniques are usually time intensive or require specific constructs, which makes scale-up difficult and not generally applicable.

1.4.1 Comparison of FBS and high throughput screening

Fragment-based screening (FBS) has increasingly been applied since 1996 to the point that it is now a widely used approach in both commercial ventures and academia. At the time, it challenged previously established methods for lead to drug development, which relied on high throughput screening (HTS) of large chemical libraries or natural products followed by the synthesis and optimization of structurally related analogs (rational drug design). HTS used huge libraries (up to millions) of higher molecular weight compounds (300-500 Da). The aim was to obtain a compound with optimized chemical, physical, and biological properties before further developing it into a clinically trialed drug. As a result, HTS tends to find leads with several interaction points to the macromolecule targeted. Some of these interactions might not represent high-quality contacts and can even reduce the binding affinity. All these effects are compounded into the total binding affinity that, if high enough, is categorized as hit HTS screening (Figure 7).

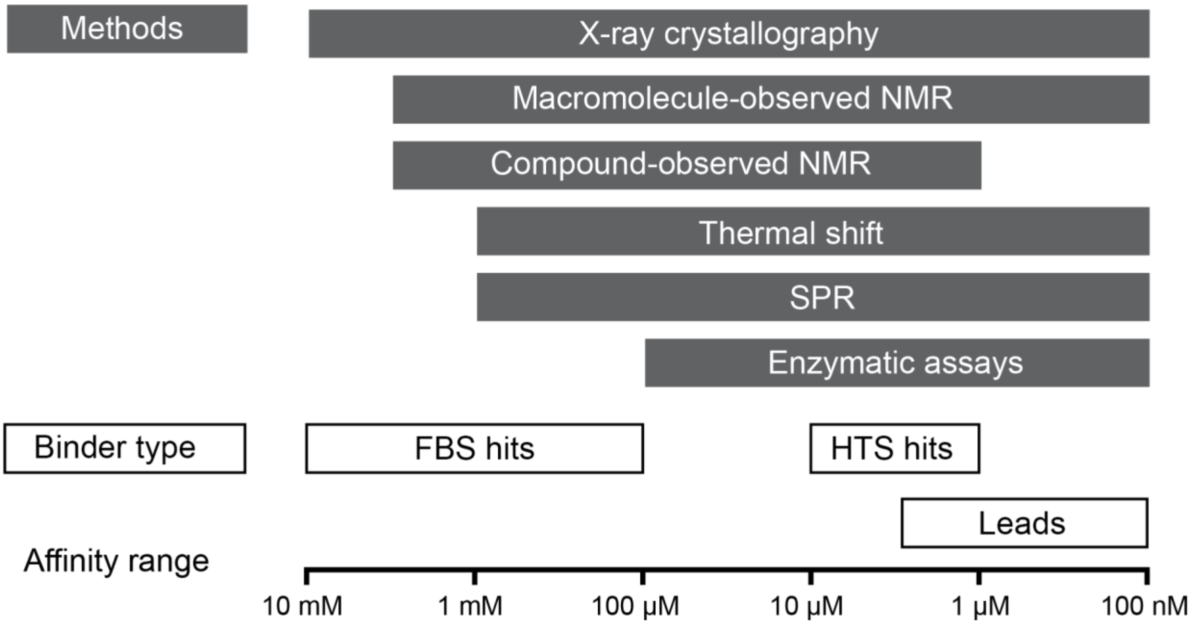


Figure 7 Sensitivity of screening techniques: Detection limits of common screening methods and common hits affinities for 150–250 Da FBS fragments, HTS, and lead compounds. Based on Price *et al.* (119).

Shuker *et al.* designed FBS referred to as structure-activity relationships (SAR) at the time. Their goal was to find compounds that produce one or two good quality interactions with a target of interest from a modest number (thousands) of low molecular weight compounds (250 Da) screened (120). They used NMR to show the feasibility of FBS to obtain binding fragments from a small library and how two weakly binding hits can easily be optimized and linked to produce a 19 nM binder. The fragment-based approach had two guiding principles that have distinguished it from HTS and have been essential to its success (121).

Firstly, screening molecular fragment libraries rather than bigger molecules is a more effective way to explore chemical space. While the number of possible drug-like compounds with up to 30 heavy atoms is estimated to be more than 10^{60}, the number of potential fragments with up to 12 heavy atoms (excluding three- and four-membered ring structures) has been calculated at 10^{7} (122). The relative fraction of the "druglike" chemical space covered in a 10^{6} compound HTS is substantially smaller than the relative fraction of "fragment-like" chemical space screened in 10^{3} fragment screening.

Secondly, the likelihood of introducing steric conflicts that impede binding increases with molecular size. Therefore, FBS yields higher hit rates with on average lower affinities (millimolar range). The added benefit is a higher binding efficiency per atom and a lower likelihood of overcompensated negative interaction (123, 124). Even a Lipinski's ´Rule of Five´ equivalent term, the ´Rule of Three´, has been termed (125). Congreve *et al.* suggested the library design should focus on fragments in which molecular weight is <300 Da, the number of hydrogen bond donors and acceptors is 3 or less, ClogP and NROT are smaller than 3, and PSA is smaller than 60.

Nuclear magnetic resonance (NMR) has been widely employed in fragment-based research because it can detect low-affinity fragments' binding and provide information on the structural basis of fragment-target interactions (102, 103, 107, 121, 126). The obtained structural information can aid the next steps of fragment modification and choose the appropriate linker length when connecting fragments.

1.4.2 Experimental challenges of FBS

Other methods applied in fragment-based screening are ITC, thermoflour, thermal shift, spectral shift, mass spectrometry, surface plasmon resonance (SPR), equilibrium dialysis, and enzymatic assays (Figure 7). While these methods provide more limited insight into the binding mode of an interacting fragment, they can be applied with substantially lower upfront or maintenance costs. Methods can be grouped based on different parameters, such as their binding classification. Depending on the method, binding is observed as a change in ligand properties. The properties can be observed through signal frequency or intensity, as a change of target properties like a melting point or structure, or additional physical evidence of binding like heat release or covariance. The observation of binding can be direct, like a change of electron density inside a screened structure, or indirect through the release of a known binder or a downstream induced enzymatic reporter reaction.

Depending on the method, it is possible to investigate mixtures of several fragments. Multi-fragment samples dramatically increase the throughput of these methods and make some financially challenged methods commercially competitive. Yet, the investigation of mixtures usually comes with the necessity to deconvolute the mixture to find the binder from the mixture. This deconvolution can be accomplished through orthogonal screening techniques. Deconvolution is not required when the method inherently allows distinguishing the fragments of the mixture, for example, NMR. Reduced need for deconvolution can be accomplished when fragment mixtures are designed to reduce signal overlap. This mixture design shifts resources from post-screen to pre-screen assay design and is usually beneficial. The aspect of false positives and false negatives should be considered. All these aspects of FBS inform the choice of method, and the named methods and considerations should not be seen as a comprehensive list but rather help inform the decision made in the experimental section of the work (Chapter 3 : ^{19}F NMR-Based Fragment Screening of RNA). Until 2016, FBS was used to develop at least 30 drugs that entered into clinical trials and have resulted in the FDA-approved drugs Vemurafenib (Plexxikon) and Venetoclax (AbbVie, Genentech) (127).

1.4.3 Targeting RNA with fragment-based screening

The approach of fragment-based drug discovery (FBDD), also known as fragment-based screening (FBS), identifies low molecular weight chemical starting points for drug discovery (119). The method involves screening libraries of low molecular weight fragments for their binding capability to large biomolecules. Although commonly found binding interactions are low in affinity, they represent structures that successfully overcome a high entropy barrier and generate effective, high-quality binding interactions with the macromolecule. These fragment leads can be optimized effectively and merged or linked into a high-affinity binder (120).

RNA is a highly relevant target for drug discovery. It is involved in a multitude of cellular processes and is a core target in the defense against viral infection (12). These cellular processes include protein synthesis, gene regulation, information storage, and catalysis (7). Successful examples of drug discovery on RNA targets were antibiotics targeting ribosomes (128–130). Ligand binding sites often include just one form of secondary structure. Secondary structure elements have, therefore, been a strong research focus, and secondary structure becomes the crucial factor in establishing the drugability of a targeted RNA (7). Additionally, due to the sequential nature of structure formation, the goal of disrupting tertiary structure can be accomplished by preventing two secondary structures from connecting. This approach finds further benefits in FBS. A comparatively simple target structure results in fewer but higher quality interactions with a lower likelihood of negative interaction. Hits can be subsequently combined to target multiple secondary structures of the RNA. Desirable binding positions are found in large asymmetric loop regions. The absence of stabilizing effects at the base of their stems results in increased accessibility and reactivity at terminal base pairs (131). Such motifs are common contributors to protein recognition, tertiary folding, and catalysis.

Several examples of FBS conducted on RNA in the literature utilize the methods outlined in this chapter (92–95). The observation of binding can occur through target signals, or the fragment signals are observed directly. Commonly, methods use STD, either directly on the target or indirectly by saturation of the water signal (WaterLOGSY), or relaxation enhancement (CPMG) of ^{1}H or ^{19}F signals.

A large screen of 4279 fragments conducted by Lee *et al.* observed changes in the imino spectrum of 32 nucleotide influenza A virus promoter RNA (132). Their initial screen used mixtures of 20 fragments and resulted in 7 hits. The compound which caused the strongest perturbation to the RNA signal was further investigated. A complex structure of fragment and target was solved, and to compound inhibited the replication of three types of influenza viruses in cellular assays with EC_{50} values ranging from 72 µM to 276 µM. A smaller screening of 250 fragments from a Maybridge “Rule of 3” collection in mixtures of 5-8 fragments was performed by Davidson *et al.* using STD measurements on HIV TAR RNA stabilized by a bound small molecule (133). After over a hundred fragments were identified as binding by STD (over 10%), 6 out of 20 fragments with the most intense STD signals were subsequently confirmed to bind the stabilized RNA while not interacting with unstabilized RNA. WaterLOGSY T2-relaxation experiments were used by Cressina *et al.* to validate the binders established in an initial equilibrium dialysis screening (134). Subsequently, Warner *et al.* solved crystal structures of four fragments bound to the TPP riboswitch using small-angle x-ray scattering (SAXS) experiments (135). They observed rearrangements into a structure distinct from that of the cognate complex highlighting that off-pathway conformations of RNAs can be targeted for drug development. T2-relaxation and WaterLOGSY were also applied in a screen of 15 SARS-CoV-2 RNAs with a 768 fragment library based on the methodology discussed in this work(136). Shortridge *et al.* utilized their previously published ^{1}H broadening methodology for screening (114, 137). They showed how the increase in T2-relaxation identified RNA binders to a pre-miR-21 stem-loop, a relevant target for oncology and metabolic diseases. Eighteen initial hits from a 420-compound library were further analyzed with an NMR pipeline removing mixture-only hits and subsequently confirming binding to the Dicer cleavage site of the RNA through intermolecular NOE. A different follow-up pipeline was published by Tam *et al..* An initial screening of 1000 fragments, in mixtures of 9 to 11, against an RNA hairpin from the *S. aureus* 50S ribosome subunit was followed by a virtual screen and subsequently followed by a functional assay (138). Four fragments were shown to have the best binding parameters based on both CPMG spectra and differential line broadening (DLB). A screen using ^{19}F labeled fragments was performed by Garavís *et al.* (139). They found 20 novel binders to a telomeric repeat-containing RNA (TERRA) from the 355 fragment library. A comparison of fragment binding capability to other nucleic acid macromolecules showed preferences for telomeric DNA G-quadruplexes over tRNA and DNA duplexes.

NMR screenings of RNA can yield lead structures that can be modified to improve the binding affinity to the target. One example of the capability to find novel structurally different binders and the feasibility of improving lead affinity was shown by Yu *et al*., after their initial screening on E. coli 16 S A-site RNA, they modified their strongest binder and found 3 modifications with even higher affinity, the highest being 9 µM (98). The combination with other methods can be utilized in the initial screen or to confirm and further analyze hits from other experimental methods or *in silico* docking (140–142). While NMR can provide validating evidence for hits for an initial screening, additional orthogonal methods are generally perceived as evidence of binding with higher confidence. Several examples of this were reviewed by Diethelm-Varela and Haniff *et al*. (93, 94).

1.5 Affinity and kinetic measurements using ITC

A core aspect of understanding a biological interaction is determining its strength. The strength of an interaction is usually determined by its dissociation constant (K_D). It is a system-specific equilibrium constant that describes the relative proportion of the concentrations of the free binding partners in relation to the concentration of the bound complex formed by the binding partners. A low K_D corresponds to a strong interaction since the bound complex is already the predominant population at low concentrations of the binding partners. Biological systems require K_D in the micromolar to nanomolar range. This results from the usual concentration of the biopolymers and the typical concentration of small molecules in the organisms. Transcriptional riboswitches are an excellent example to illustrate this. Because their switching regulation must occur during transcription, there is likely only one transcript with a fully formed aptamer per cell at a time. One transcript per cell puts the *in vivo* concentration at the inverse of the cell volume. A bacterial cell volume of 0.5 μm^3 corresponds to a concentration of 3.3 nM. Small molecules are, with rare exceptions, present in the millimolar range or substantially lower. Therefore, any interaction forming a stable riboswitch ligand complex has a K_D that must be micromolar or lower. This effect gets further increased by possible interactions of the small molecule with other macromolecules. Another limiting factor is the time window when aptamer is present due to transcription speed and refolding. Fast transcription speeds require higher ligand concentration to saturate the aptamer while pausing transcription can keep sensitivity ranges near the aptamer's K_D (57). A high-affinity binding is detrimental to biological function, and accurate measurements of this interaction are detrimental to understanding it.

While the systems are complex, the strength of the underlying interactions can be studied in isolation. With few exceptions, methods that allow determining K_D constants involve measuring the concentration or physical properties of one or more binding components or the formed complex while successively adding one of the binding partners to the system, a classical titration. Depending on the method, this requires labels or can be done label-free, for example, using isothermal titration calorimetry (ITC) or NMR. Physical properties that can be probed to determine K_D constants may include thermophoresis (MST), Spin behavior (NMR and ESR), absorption or emotion spectrum of the binder or a dye attached to the binder, (UV-vis, IR, Raman, fluorescence spectroscopy, SPR), changes to the structure or the rigidity of the binder (X-Ray, electrophoresis, chromatography, modification mapping sequencing), mass (mass spectrometry), heat transfer (DSC, ITC) and enzymatic activity of the binder or a reporter system (biosensors) (143). Since the observed binder needs to undergo a change that is detected to quantify the strength of the interaction, the measurement should be carried out at a concentration high enough to form a large population of the bound complex, and the inflection point of the titration curve is at the center of the data points (144). ITC was used in this work to determine the affinity of several riboswitch RNAs towards the cognate riboswitch ligand c-GAMP.

1.5.1 History of ITC

The first scientific uses of calorimetry date back to the 18th century, but it took until the 1970s for highly specific differential scanning calorimeters (DSC) to become commercially available (144). DSC investigates the transition temperatures and heat of thermally induced changes in the conformation of biopolymers compared to a reverence cell. The first commercial DSC machines were developed by P. Privalov and produced in Russia. The high sensitivity allowed the operation at small volumes and comparatively low concentrations. It made investigating biological samples more affordable and allowed the analysis of proteins that aggregate at higher concentrations. The method requires slow heating or cooling of the sample in the temperature range of the investigated transition. It requires a substantial amount of time for each measurement and limits the systems that can be studied due to the thermal degradation of biomolecules. J. F. Brandts further innovated the field during the 1980ies, upgrading DSC machines by the company MicroCal to operate as ITC machines. These featured a motorized injection tip with stirrer blades allowing injections into and mixing the measurement cell during operation (Figure 8).

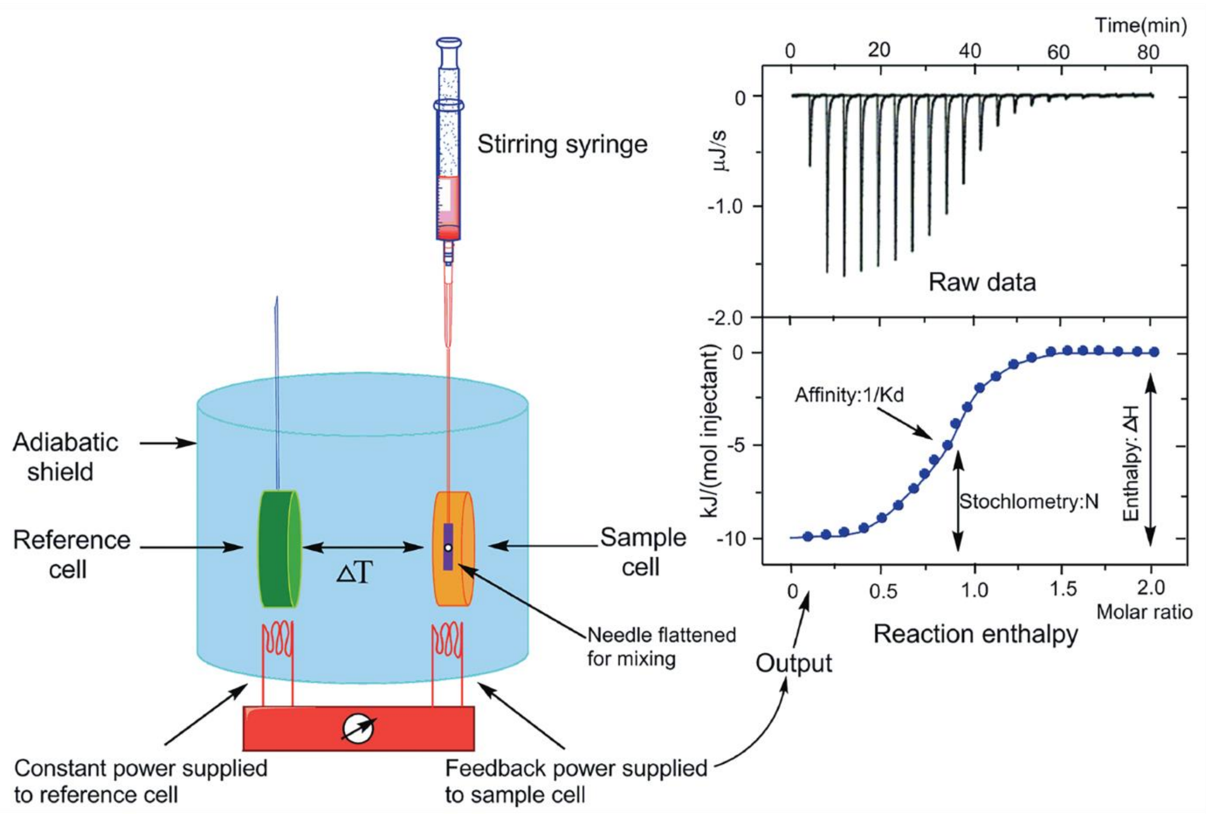

Stirring syringe
Adiabatic shield
Reference cell
ΔT
Sample cell
Needle flattened for mixing
Constant power supplied to reference cell
Feedback power supplied to sample cell
Output
Time(min)
0
20
40
60
80
μJ/s
0
-1.0
-2.0
Raw data
kJ/(mol injectant)
0
-5
-10
Affinity:1/Kd
Stochlometry:N
Enthalpy: ΔH
0
0.5
1.0
1.5
2.0
Molar ratio
Reaction enthalpy

Figure 8 Basic principle of isothermal titration calorimetry. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment (upper right) with its evaluation (lower right). The titration thermogram is represented as heat per unit of time released after each injection of the ligand into the protein (black), as well as the dilution of ligand into buffer (blue). The dependence of released heat in each injection versus the ratio between total ligand concentration and total protein concentration is represented. Circles represent experimental data and the line corresponds to the best fitting to a model considering n identical and independent sites. Figure was upscaled and initially published by Song *et al.* under Creative Commons Attribution License (145).

A measurement cycle consists of a small initial priming injection followed by multiple injections of one of the binding partners in 1-25 µL aliquots. Changes in the energy needed to heat the measurement cell in relation to the reference cell are measured and graphed over time. Isothermal titration calorimetry is a quasi-isothermal method. The sample is actively heated to compensate for possible endothermic changes. This heating results in an insignificant temperature increase over the entire measurement cycle. A unique feature of ITC is that K_D and binding enthalpy are determined in the same measurement. The measurements yield free binding enthalpy and binding entropy values. These measurements can establish K_D values for any chosen temperature if the system tolerates the temperature in question.

1.5.2 NMR and ITC

ITC and NMR are uniquely complementary methods for the characterization of biological samples. Both methods are label-free and nondestructive. The same sample can be measured in both setups without extensive sample recovery. NMR and ITC detect completely independent physical properties of the sample and can therefore cross-validate data on the same sample. In this work, binding windows of RNA were established through binding associated changes in the NMR spectra of RNA samples and the determination of binding affinities (K_D) of the corresponding constructs. Both methods share the necessity for comparatively high sample amounts to overcome sensitivity limitations. The chance to obtain biophysical data from two methods might entice researchers to overcome the fear of initiating these financially costly methods. When engaging in these multi-method approaches investigating RNA, significant care should be applied to prevent issues stemming from contamination, especially RNases. The biggest threat to sample purity stems from the ITC sample cell, permanently attached to enable high sensitivity. It requires special preparation to mitigate the RNase risk. NMR investigations should precede or at least promptly follow up ITC analysis when possible. The last significant caveat to ITC and NMR analysis is the sensitivity to bubble formation. When using small-volume NMR Shigemi tubes or ITC, the formation of air bubbles during measurement runs can make the acquired data unusable. Careful preparation and sample cell filling are required. Still, the storage of sample and buffer at low temperatures and the detection at higher temperatures might make occasional bubble formation due to changes in gas solubility an unavoidable problem. Continuous checks for these errors in the data are required, and possible failures should be expected. Accordingly, the preparation scale of the biological sample should be increased to allow additional necessary measurements.

1.5.3 Kinetics studies using ITC

The capability to obtain kinetic information from chemical reactions has long been known and was proposed by Sturtevant (72, 146). Many decades and remarkable scientific advances allowed a continuous move towards smaller sample sizes and single measurements rather than multiple at varying conditions (147). Burnouf *et al.* showed the general applicability of ITC measurements for kinetic analysis (72). They utilized ITC as a multiple injection method (MIM) in which compound A, in the measuring cell, is combined with compound B, which is added in small aliquots at regular intervals. The heat evolved at each titration step is obtained by integrating the heating power peaks. The analysis of these integrals allows the determination of the variation of enthalpy, the affinity, and the stoichiometry of the interaction. Their method KinITC (kinetic ITC) was used to analyze titration curves and extract the kinetic data from their shapes. Burnouf *et al.* used KinITC in their first paper to investigate the TPP riboswitch (72). Their method, paired with surface plasmon resonance (SPR) and hydroxyl radical footprinting, enabled them to show and validate the power of KinITC as an analytical tool. It is currently marketed by the company Affinimeter.

1.6 Simulation of riboswitch RNA folding

The computational description of RNA folding can refer to two different concepts. It, firstly, can refer to the general behavior of RNA to form stable structures and transitions between different stable structures. Secondly, RNA folding can refer to the process of initial structure formation, which occurs during the transcription of RNA. This second process is called cotranscriptional folding. General folding and cotranscriptional folding contain substantially different challenges.

In the simulation of general folding, specific RNA sequences are simulated and folded, and the sequence of these RNAs is always used in its entirety. These simulations calculate the energies of stable structures or transition states to resolve ambiguity about the exact structure (148–151). The simulations determine base pairing partners or the relative fractions of possible RNA folds and result in an understanding of the long-term behavior of the RNA structure in solution. Translational riboswitches are fully transcribed before they are required to function. The simulation of translational riboswitches as the classical folding simulation will accurately describe the underlying folding.

1.6.1 The cotranscriptional folding problem

Further increasing the difficulty of simulation, cotranscriptional folding adds the challenge that the RNA of interest is successively extended (152, 153). While initial structures are forming, adding new nucleobases changes the underlying energy landscape and structural equilibria. The formation of new structures becomes conditional on unfolding initially formed structures. The unfolding depends on the stability of the corresponding structures and the timeframe provided by the transcription speed. The sequential nature of cotranscriptional folding can trap conformational assembles in local minima of a global energy landscape. In protein science, this is called the folding problem (154–156). These conformational ensembles can persist for a long duration, exert biological functions, and might never be provided with enough time or the conditions to fold into the absolute minimum energy state of their full-length RNA sequence.

Transcriptional riboswitches, in contrast, rely on the formation of structures during transcription. They contain decision points that can terminate transcription and regulate gene expression. These decision points are reached within seconds of the transcription start. Hence the effects of cotranscriptional folding dominate transcriptional riboswitch function. Therefore, transcriptional riboswitch folding can only be simulated as cotranscriptional folding. Simulations of cotranscriptional folding exist on a broad spectrum ranging from purely *in silico* to highly informed by additional data from experiments. In silico models aim to provide reliable predictions with the long-term goal of universally determining transcription folding without requiring experiments (157). Other models might add further data like determining stabilities, the affinity of binding partners, or predominant fold structures. While this reduces possible errors from the assumptions and approximations made *in silico*, possible errors from the experiments are introduced.

The first example of a cotranscriptional folding simulation software was introduced by Proctor et al. in 2013 (158). Their software COFOLD was a conceptually new technique for predicting RNA secondary structure that explicitly accounts for the impacts of cotranscriptional folding. In terms of prediction accuracy, their technique vastly outperformed the state-of-the-art at the time, especially for lengthy sequences longer than 1000 nt. COFOLD combined thermodynamic energy scores with a state-of-the-art method for predicting the secondary structure of RNA, RNAFOLD(159), with a scaling function that accounts for the effects of kinetic folding. By considering the effects of kinetic folding, COFOLD enhanced RNA secondary structure prediction rather than directly simulating the folding route. COFOLD's predictions of RNA secondary structures were tested on a broad range of known RNA secondary structures, and their results demonstrate a considerable improvement in prediction accuracy, especially for long RNA sequences like ribosomal RNAs. COFOLD uses a combination of thermodynamic and kinetic considerations to direct the structure prediction process but could not account for trans-interactions with other molecules. Interactions with macromolecules such as other RNAs or proteins, are another crucial part of cotranscriptional RNA structure formation and were an at that time insurmountable challenge. The authors stated that explicit knowledge of the binding sites and timing of these interactions with respect to the transcription of the RNA would be required in order to incorporate them into a predictive model. Information which at the time was limited to a small number of RNAs.

The cotranscriptional RNA folding model (kinfold, ViennaRNA) presented by Hofacker *et al.* simplifies RNA folding to the degree that it acts as a suitable trade-off between accuracy and computational effort (157). Since secondary structure formation absorbs the majority of the free energy required for RNA folding, RNA structures are represented by secondary structures. This description does not account for tertiary interactions, nor can RNA-ligand binding be described. In their study of the 2′-deoxyguanosine (2′dG)-sensing riboswitch from *Mesoplasma florum* they used binding motifs and affinities derived from in vitro experiments. At present, the use of molecular dynamics at atomic resolution is still out of reach due to the system size and time scales involved. They outline a computationally effective method for predicting cotranscriptional RNA folding with a trans-interaction (ligand binding) present. Their model requires a completely formed aptamer structure to bind a ligand, representing ligand binding as an all-or-none process. A structure is deemed binding competent if it includes the whole aptamer. This assumption disregards the affinities of partly produced aptamer intermediates that bind ligands. Kinfold was employed to simulate 10000 trajectories of the developing 2′dG riboswitch (dGSW) with and without ligand beginning from an initial chain length of 40 nt. For ligand stabilization, 8 kcal/mol of additional energy was attributed to the corresponding structures. Comparison with experimental data led to the conclusion that 1 s is equivalent to 100,000 internal time units. As a result, a chain elongation event occurring every 4,000 simulation steps corresponds to a transcription rate of 25 nt/s. Giving all aptamer motif structures that include the pattern bonus energy of 8 kcal/mol is the most basic model to describe binding. This approach is equivalent to providing indefinitely high ligand concentration and infinitely quick binding. The presence or absence of the aptamer motif and the terminator hairpin were then used to categorize each of the 10,000 structures along trajectories, determining four different structural classes. The likelihood of finding the folding chain in an aptamer or a terminator for each time point was then calculated using this data. The model proved capable of reproducing experimental results and explained the switching behavior of the 2′dG riboswitch. Kinfold could be utilized to universally predict the cotranscriptional folding dynamics with and without ligands if binding affinities and a riboswitch's binding competent aptamer structure are known. Beyond Hofacker *et al.,* there are few practical *in silico* methods available for accurately simulating the folding behavior of nascent RNA, and several of the available options (RNA kinetics (160), Kinefold(161) and Gong *et al.* (162)) do not provide source code and therefore limit reproducibility and reliability.

A recent development is the use of artificial intelligence for RNA structure prediction. Ufold uses an image-like representation of RNA sequences based on fully convolutional networks (FCNs) (163). It outperformed previous methods on within-family datasets and matched performance when trained and tested on distinct RNA families. Currently, performance seems limited by the quality and the number of experimentally resolved RNA secondary structures.

1.6.2 Computational modeling of specific riboswitches

Some computational models introduced additional constraints on the studied systems to increase the understanding of specific riboswitches but lost their general applicability. These simulations usually employ experimentally determined transition rates, limit the number of states a riboswitch fold can acquire, prohibit transitions based on empirical knowledge about the studied riboswitch, and aim to answer specific questions about the system. An early example of this was introduced by Wickiser *et al.* (57). They investigated the combined effects of three biologically active ligands of the pbuE riboswitch. 2,6-diamino purine (DAP) binding was thermodynamically favored, while adenine binding was favored due to a higher association rate constants. They further showed that the riboswitch could be either kinetically or thermodynamically controlled depending upon the time scale of transcription and other variables like temperature. They simulated the concentration of bound riboswitch RNA in dependence on ligand concentration for different transcription durations. At long transcription times between $1/k_{on}$ and $30/k_{on}$ (about 6.5 s to 200 s), the riboswitch is in thermodynamic equilibrium, and the system has a low B_{50} inline with a high-affinity aptamer. Reducing the transcription time down to 1/150 k_{on} (about 0.04 s) successively increases B_{50}, higher ligand concentrations are required to achieve this same amount of bound RNA, and kinetic control takes over. They also observed a steeper response curve slope under kinetic control requiring 19 times less ligand to increase occupancy from 50% to 99%. A simulation of the pbuE riboswitch by Sun *et al.* found that when the aptamer is stabilized by -5.3 kcal/mol through adenine binding, it reaches 95% population without transcriptional pausing(70).

Beisel and Smolke developed design principles for riboswitch function in 2009 (75). They utilized Markov-style simulations of transcriptional, translational, and regulation through degradation to derive general principles of riboswitch function. They use EC_{50}, the effective ligand concentration at which a half-maximum response is achieved, and dynamic range (η), defined as the difference (not the ratio) between high and low protein levels (the regulatory output), to discuss the influence of specific transition rates of the regulatory systems. They found that non-transcriptional riboswitches can fall into a thermodynamically-driven, kinetically-driven regulatory regime, or a nonfunctional regime. When reversible rates (binding or folding aptamer competing fold) dominate, we find thermodynamic regulation, while a balance of reversible and irreversible rates is required for kinetic control. Domination by irreversible rates (termination rate) results in a loss of function, the dynamic range approaches zero, and ON and OFF have no difference in genetic output. Riboswitch performance from the non-functional regime is rescued when we account for ligand binding during transcription, transcriptional riboswitch activity (Figure 9A). Initially, the aptamer of transcriptional riboswitches can reversibly bind until the terminator is transcribed. Beyond this point, rates are no longer reversible. The system can bind ligand and progress towards termination or refold into aptamer competing form full-length transcript. Under these restrictions, EC_{50} and dynamic range showed dependences on specific rates. EC_{50} was maximized when RNA was readily transcribed to terminator length compared to the time required for dissociation ($1/k_{off}$). The EC_{50} behavior showed the same reduced ligand sensitivity under kinetic control as described by Wickiser *et al.* (57). Beisel and Smolke also determined the upper and lower boundaries of EC_{50}. The minimal EC_{50} value approaches the dissociation constant for transcription times longer than $1/k_{off}$, thermodynamical equilibrium. The ratio of irreversible rates determined the upper limit after the initial binding window. High terminator stem formation and antiterminator folding rates increased EC_{50}, while fast k_{on} rates decreased it. The upper limit is approached for fast transcription speeds under kinetic control when minimal time is available for equilibration. Dynamic range showed no shared dependence with the EC_{50} value. The dynamic range was maximized when the conformational transition towards aptamer competing fold occurs much faster than the formation of the terminator stem. The dynamic range was not dependent on the transcription speed. Full-length riboswitches operating under transcriptional termination strongly prefer the aptamer-disrupted conformation and exhibit negligible ligand binding affinity.

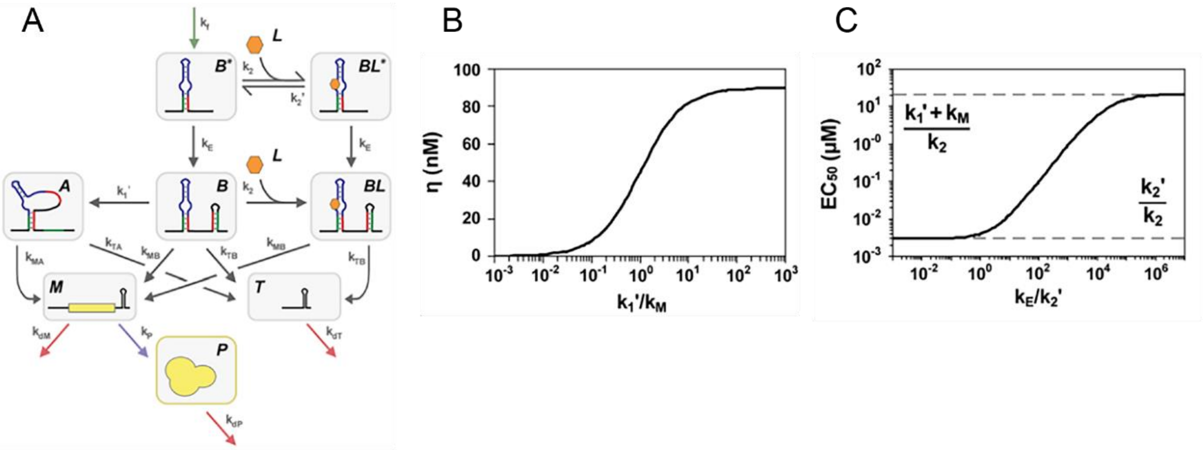


Figure 9 Rescuing riboswitch performance in the non-functional regime: (A) Riboswitch functioning through transcriptional termination general response relating ligand concentration [*L*] and regulated protein levels [*P*]. The aptamer is first transcribed in conformation [*B**] and can reversibly bind to form [*BL**] and release ligand before the terminator stem is transcribed (k_E). Terminator stem formation ($k_M=k_{MA}+k_{TA}=k_{MB}+k_{TB}$) occurs much faster than ligand release ($k_2´$) and the progression from

conformation *A* to *B* (k_1). Conformation *B* contains a formed aptamer that can bind ligand to form [*BL*]. Green arrow designates mRNA synthesis with biased transcriptional folding, red arrows designate species degradation, and blue arrow designates translation proportional to mRNA levels. Under transcriptional termination, riboswitches effectively choose between termination to form a truncated product [*T*] and extension to form the full-length mRNA [*M*]. (B) The competition between terminator stem formation (k_M) and the progression from conformation *B* to *A* (k_1') determines the dynamic range (η). (C) EC_{50} can be tuned independently from the dynamic range. The accessible range of EC_{50} values is bounded by the aptamer association constant ($K_2=k_2/k_2'$), the rate constant for the progression from conformation *B* to *A* (k_1'), and the rate constant for terminator stem formation (k_M). EC_{50} is tuned over this range by the rate constant representing the delay between aptamer formation and transcription of the terminator stem (k_E). Parameter values: $k_2=10^6$ $M^{-1}s^{-1}$; $k_2'=3*10^{-3}$ s^{-1}; $K_A=k_P*k_{MA}/k_M=10^{-3}$ s^{-1}; $K_B=k_P*k_{MB}/k_M=10^{-2}$ s^{-1}; $k_f=10^{-11}$ $M^{-1}s^{-1}$; $k_{dP}=10^{-3}$ s^{-1}; $k_{dM}=10^{-3}$ s^{-1}; $k_1'+k_M=20$ s^{-1}. Figure was upscaled, rearranged, and originally published by Beisel and Smolke under Creative Commons Attribution License (75).

Lin *et al.* quantitatively predicted the folding landscape of the FMN riboswitch and compared the add and the pbuE riboswitch folding landscape (164, 165). A critical analysis tool was contour plots of the riboswitches showing the protein production dependent on the binding rate k_b (k_{on}) and the folding rate ratios. The resulting contour plot can be subdivided into different regimes that correspond to different chains of transitions that result in the regulatory riboswitch outputs.

More examples of Markov simulations on riboswitches were carried out in the Schwalbe group by Steinert *et al.* and Helmling *et al.* and highlighted the time-dependent transitions of riboswitches (31, 53, 54). They experimentally obtained several of the required rates for their models through time-resolved NMR, stop-flow measurement, and time-resolved transcription.

Steinert *et al.* observed pause sites in the *xpt-pbuX* riboswitch sequence with two RNA polymerases. Their Markov model simulated riboswitch regulation under ligand-free and ligand-containing conditions with and without pausing. The pausing was introduced through a 70-fold reduced transcription rate in line with experimental observations. They show that only the combination of elevated ligand concentration and pausing drives a population transition favoring the ligand-bound state and resulting in OFF regulation through transcript termination.

Helmling *et al.* initially used a nine-state Markov model to investigate the ligand-free folding transitions. The model described transitions between the 2′-deoxyguanosine (2′dG)-sensing riboswitch aptamer fold and two competing antiterminator folds at three different transcript lengths. The lengths represented the initiation of antiterminator folding, the point where the antiterminator folding pathway can no longer be entered, and the point of regulatory decision. Their results showed a negative correlation between transcription speed and gene expression. They also provide a plot of the relative population of all nine states over time at a transcription speed of 10 nt/s and 40 nt/s. The increased gene expression at the slower transcription speed can be attributed to a higher population of the slower-forming but thermodynamically favored antiterminator conformation. In addition to their Markov model, they compared their *in vitro* characterized mechanism with *in silico* analysis conducted in the Hofacker group (53, 157). The barmap method could be used due to a high degree of secondary structure rearrangements to the different regulatory states of the riboswitch. While these secondary structures have well-established energy models, the aptamer tertiary structure had to be accounted for through experimentally derived ligand binding energies. A free energy bonus was added to every binding competent secondary structure. The model and the *in vitro* result describe remarkably similar behavior. The aptamer domain represents the dominant structure until $dGsw^{110}$, thermodynamic equilibrium shifts from PA to PT at ~$dGsw^{113}$, and the antiterminator state $P^{A}T$ is stable over a ~23−24 nt segment. The simulations suggested the formation of additional states, which may be indetectable by NMR, including an alternative aptamer conformation containing a helix P2*. Subsequently, Helmling *et al.* complete their Markov simulation in their second paper by describing the ligand-binding-related behavior (54). The dGSW contains an initial binding window of 13 nt length with full binding capacity followed by a 17 nt window with impaired (70%) binding. Their simulation added a ligand-bound state and simulated a fraction of ON signaling when 100% ligand-bound and 70% ligand-bound states were cotranscriptionally reached. They performed simulations of the conformational transition from aptamer to antiterminator between transcript lengths 113 and 137 at different transcription rates (0–100 nt/s) depending on whether 100% or 70% ligand binding was reached at $dGsw^{113}$. The $dGsw^{137}$ length is assumed to represent the latest possible regulatory decision point. The simulations show that the riboswitch performs best at transcription rates of 10–50 nt/s, in line with the typical speed of bacterial polymerases. Ligand-bound aptamer could no longer delay ON state folding sufficiently at slower speeds. At transcription rates exceeding 70 nt/s, ON state folding cannot compete with the transcription speed, and the ligand sensitivity of the system is lost. Only at intermediate speeds a maximum regulatory efficiency is achieved and limited to 40–50%.

Parra-Rojas *et al.* proposed a generalized modeling framework for the kinetic mechanisms of transcriptional riboswitches (166). Their formalism accommodates time-dependent transcription rates and changes in metabolite concentration and permits the incorporation of variations in transcription speed depending on transcript length. They derive explicit analytical expressions of the 2′dG riboswitch, refining the model of Helmling *et al.* discussed above. The two main changes represent the introduction of sink states and the merger of the two biological ON states into one total ON state fraction. As a result, the system size effectively reduces from twelve to nine dimensions. They discuss that for sufficiently large numbers of nucleotides separating all the different transcriptional intermediates, the transitions may be well approximated by a fixed waiting time. However, a more realistic simulation could be conducted, comprising windows of 1 nt, each traversed in exponentially distributed times.

Aware of these possible improvements, the simulation in this work described the cumulative effects of sink states instead as the sum of all lengths in the corresponding state after a time frame that allows over 95% of transcripts to reach the full transcript length. In addition, it is known that short RNA stems can rapidly refold at ambient temperatures. A 1 nt window approach was chosen for every transcript, with each state containing its specific structure-derived refolding rates. The results of this work are shown and discussed in chapters 2.1.3 , 2.2.2 , and 2.2.3 of this work. The only other examples of a 1 nt resolution simulation that utilized experimentally determined rates we are aware of was conducted by Gong *et al.* (162, 167–169). They determined refolding rates with a helix-based method where each elementary kinetic step is represented as the transformation between two secondary structures that differ by a helix (170). The model was validated against several experimental measurements (171).

Chapter 2 Cotranscriptional Folding Landscape of CDN-Riboswitches

2.1 Introduction

2.1.1 CDN-binding riboswitches

CDN-binding riboswitches are regulators of networks of genes for the environment, membranes, and motility (GEMM) (48). External stimuli cause changes in CDN concentration through cyclases forming CDNs and phosphodiesterases degrading CDNs, respectively. The resulting change in signaling molecule concentration regulates mRNA transcription at multiple riboswitch-controlled loci and leads to lifestyle changes, including biofilm formation and pili expression (36). C-di-GMP controls changes in the mobility lifestyles of gram-positive and gram-negative bacteria. For gram-positive bacteria, including *C. difficile*, it was shown that elevated levels of c-di-GMP lead to biofilm formation (36, 172). The c-di-GMP-induced lifestyle change is a key factor of pathogenicity in *Vibrio cholera* and *Clostridium difficile* (39, 173). Biofilm formation has been shown to increase antibiotic resistance *in vitro* in *C. difficile* (174–177).

Examples of c-GAMP-sensing riboswitches are found in multiple strains of gram-negative bacteria (39–41). Due to the high similarity in secondary structure, they were annotated as c-di-GMP-sensing riboswitches before c-GAMP was known to also act as a secondary messenger (41). Many c-GAMP riboswitches can also bind c-di-GMP, some with similar affinities and others with remarkably higher affinities for c-GAMP (41). The lack of selectivity could originate from the lack of necessity since the organism might only use one of the two messengers for signaling. Keller *et al*. demonstrated that c-GAMP riboswitches could bind c-di-GMP using a stably protonated adenine in the ligand binding pocket (178). Ligand affinity and selectivity between c-di-GMP and c-GAMP can be altered with a single point mutation in the lower part of the P2 stem (42). The pilM riboswitch of *Geobacter metallireducens* might have experienced particular evolutionary pressure because the bacterium uses c-GAMP and c-di-GMP signaling. *G. metallireducens* gene transcription is regulated by c-GAMP-sensing riboswitches at thirteen loci, and the c-GAMP aptamer appears multiple times in some of these loci (40). One of these loci is a polycistronic gene coding for pili. The encoded genes enable G. metallireducens to utilize metal oxides as an electron sink under anaerobic conditions. The electrically conductive pili form a root-like network connecting the organism to the metal oxide particles (179). This allows *G. metallireducens* to efficiently maintain its metabolism in anaerobic environments using ferrous oxide as an electron sink, predominating its habitat (180). The pilM riboswitch (Figure 10) acts as a transcriptional ON-switch. Increased stability of the aptamer against refolding after binding c-GAMP prevents the formation of the terminator and enables the transcription of the downstream genes.

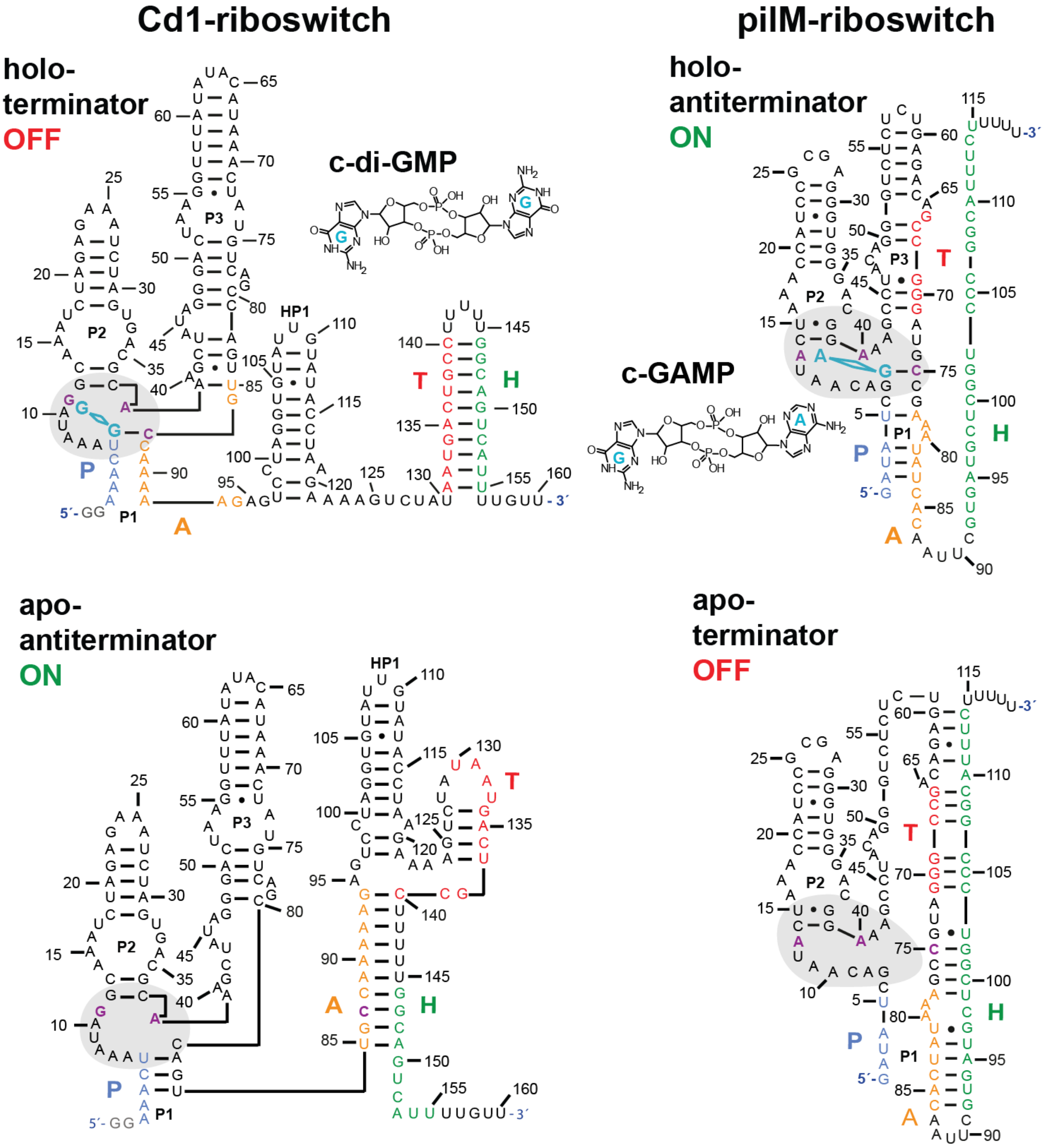

Cd1-riboswitch
pilM-riboswitch
holo-terminator
OFF
holo-antiterminator
ON
apo-antiterminator
ON
apo-terminator
OFF
c-di-GMP
c-GAMP
P1
P2
P3
HP1
P
A
T
H
5´-GG
-3´

Figure 10 Nucleotide sequence and secondary structures of full-length pilM and Cd1 riboswitches in their ligand-bound conformation with their cognate ligands c-di-GMP and c-GAMP, respectively. The bound ligands are indicated in blue, and stems are annotated with P1, P2, and P3. The nucleotides of the binding pocket are indicated in violet. The strands that contribute to the riboswitch function are color-coded in the PATH nomenclature in P(blue), A(orange), T(red), and H(green). The 5´-aptamer strand P pairs with an aptamer-stabilizing strand A to form the binding competent aptamer. The switching strand T and the terminator strand H are located downstream on the mRNA. For Cd1, the terminator conformation is formed through interactions of strands T+H and antiterminator through T+A+H. For pilM, the terminator conformation is formed through interactions of strands T+A+H and antiterminator through P+A. Additionally, the binding pockets are marked with a gray background. Figure as published in Landgraf *et al.* (1).

2.1.2 Transcriptional riboswitches

In the case of transcriptional riboswitches, regulation occurs during transcription by forming a terminator stem. If the terminator is folded, transcription stops. Otherwise, the expression platform folds into a non-terminating structure called an antiterminator (45). Depending on if the increased stability of a ligand-bound aptamer contributes to the stability of the terminator or the antiterminator conformation, riboswitches can act as ON- or OFF-switches. In an ON switch, an increase in the concentration of the cognate ligand will increase the proportion of RNA that folds into the antiterminator formation. This results in a higher proportion of the regulated gene being transcribed into a complete mRNA, allowing subsequent expression. This creates a sigmoidal dose-response curve of the riboswitch around the affinity value of its aptamer.

The requirements for adopting either the aptamer conformation or the terminator conformation in the pilM riboswitch are mutually exclusive. In the Cd1 riboswitch, these conformations coexist and compete with the formation of the antiterminator. Therefore, comparing the two riboswitches offers an opportunity to investigate the conformational transitions necessary to produce inverse regulatory mechanisms.

2.1.3 Cotranscriptional folding of riboswitches

The transcriptional gene regulation mechanism relies on mRNA structures that are populated cotranscriptionally. Only these structures contribute to the mechanism. The transcribed RNAs populate an ensemble of minimal energy states at a speed proportional to the height of the kinetic barriers to reach the corresponding states. An intrinsic feature of transcriptional riboswitches is that the regulatory mechanism only includes barriers that can be overcome during transcription. RNA transcription intermediates occupy low-energy structures on multidimensional folding landscapes that continually change with the extension of the RNA sequence. Initial local minima structures contribute to refolding and substantially impact the final RNA structure (33, 181). An inducer molecule can bind and stabilize an intermediate aptamer domain in the case of riboswitches. This aptamer domain stabilization causes equilibria to change in favor of terminator or antiterminator folds, both aptamer competing folds (45–47).

A study of cotranscriptional folding requires a methodology that can emulate the dynamic folding process that allows riboswitch gene regulation. Methods of study range from real-time observation of transcription, over roadblock systems halting transcription at desired stops, to detailed analyses of isolated key states. Realtime cotranscriptional folding was observed using single molecule force spectroscopy on an adenine-sensing riboswitch (182). Pausing was demonstrated as a crucial step for riboswitch gene regulation using kinetic models and PAGE analyses of single-round riboswitch transcriptions developed by Wickiser *et al*. (31, 55). Various chemical probes can be used to learn more about the structure of folding states. A technique was created by Strobel *et al.* that examined each transcriptional intermediate of an RNA in a single mixture through a roadblock assay followed by sequencing (183). Understanding the RNA residues' sequence length-dependent reactivity reveals essential details about the observed RNA structures. In addition to its remarkable capacity to study individual molecules, smFRET (Single-molecule fluorescence (or Förster) resonance energy transfer) also adds a thorough knowledge of cotranscriptional processes. One example is how interactions with RNA polymerase help to maintain the preQ1 riboswitch fold (71). NMR was applied to analyze cotranscriptional folding by studying successively longer tRNA fragments as early as 1980 (184).

Using a previously developed approach that offers specific base pairing data for each transcript length, it was possible to characterize CDN-sensing riboswitches and their regulatory outputs while learning essential details about ligand-dependent RNA refolding. Additionally shown by earlier research is the requirement of pausing for riboswitch functionality (31). Interestingly, the pilM riboswitch lacks any known putative pause sites beyond the termination-related site.

In this work, NMR is used to explore the structural bases of the regulatory output, while ITC is used to quantify binding affinities and kinetics. The experimental data formed the basis for Markovian simulations that assessed responses to changes in the transcriptional variables, including ligand concentration and transcription speed for transcriptional ON- and OFF-switches. The simulations also show how pausing affects riboswitch performance.

2.2 Results

2.2.1 NMR chemical shift assignment of the pilM riboswitch

Combining ^{1}H/^{1}H-NOESY spectra with and without the native ligand and information from the crystal structure of a very close homologous RNA (42) allowed an imino assignment (Figure 11). Multiple naturally occurring GU base pairs found in the secondary structure aided the assignment with their characteristic strong cross-peaks found between 12 to 9 ppm. The GU imino signals allowed a complete assignment of the P2 and the P3 stems of the pilM riboswitch. The assignment of the upper part of the P3 stem was challenging due to its palindromic sequence of GUGUG and the fact that G52 and G62 have the same chemical shift. A cross peak between a C2H2 side of A28 to the imino NH of G62 solved this assignment. This cross peak is in line with a distance of ~ 5 Å from A28 H2 to G62 H1, as found in the homolog crystal structure (42). This allowed a definite determination of G62 contributing to the cross peak despite G52 having the same spectral shift. A second cross peak is found in the spectrum of the *holo* aptamer RNA connecting U55, and the now assigned A28 enabled an unambiguous assignment of the palindromic region. An adenosine C2H2 side contributing to the cross peak was confirmed in NOESY experiments with selectively ^{13}C labeled adenosine residues. The C2H2 side peak of A28 with chemical shifts ($\delta(^{1}H)$ = 9.5 ppm, $\delta(^{13}C)$ = 156 ppm) lies remarkably far outside the typical chemical shift region. The high similarity between the full aptamer spectra with the P1-truncated pilM$^{14\text{-}75}$ aptamer confirmed that all large spin systems reside in the P2 and P3 stem (Figure 11C) of the aptamer.

Ligand addition produces several strong chemical shift changes in the pilM NOESY spectrum (Figure 11B). A large up-field signal shift can be observed for the lower GU base pair of U15 and G38 located close to the binding site in the lower part of the P2 stem. G38 shifts 0.2 ppm from 11.0 ppm to 10.8 ppm, and U15 shifts 0.6 ppm from 10.8 ppm to 10.2 ppm. G38 would be expected to have two cross-peaks, based on its proposed secondary structure, the U15 cross peak and a cross peak to the neighboring base G39. However, an additional crosspeak can be observed in the *apo* spectrum. In the crystal structure of the *holo* by Ren *et al.*, a G66 C37 base pair is located on top of G38 (42)*.* The distance between predicted imino protons of G38 (notation in structure is G39) and G66 is 5.539 Å (chain A) and 5.709 Å (chain R) for the two aptamers found in the unit cell of the crystal. This distance is beyond 5 Å and is not commonly observable in NOESY spectra. The additional cross peak of the *apo* pilM spectrum is likely caused by a G38-G66 cross peak (Figure 11A) which is no longer present after ligand addition. The distance between the iminos of G38 and G66 increases, and the cross peak becomes undetectable in NOESY experiments (Figure 11B). This aligns with the strong chemical shift perturbation near the binding pocket and indicates significant structural changes. A readily observable signal and a good indicator of ligand binding is the most downfield signal U53. It shifts 0.04 ppm downfield from 14.78 ppm to 14.82 ppm when the ligand is bound.

To aid the assignment of P1 stem imino signals, ^{1}H-^{15}N correlation spectra were acquired (Figure 11D). While no additional signals indicating a formed P1 stem were observed for the *apo* aptamer, four additional U signals in the spectral region from ^{1}H 13.1 ppm to 12.7 ppm and ^{15}N 161 ppm to 164 ppm were observed in spectra of the *holo* aptamer. No NOESY cross-peaks could be resolved in this region due to the limited chemical shift resolution and thus unambiguously assigned. The high degree of similarity to the crystal structure, the absence of the signals before the ligand is added, and the expected increase in stability at the top of the P1 stem as a result of ligand binding make it likely that these U signals correspond to the U residues found in the lower part of the P1.

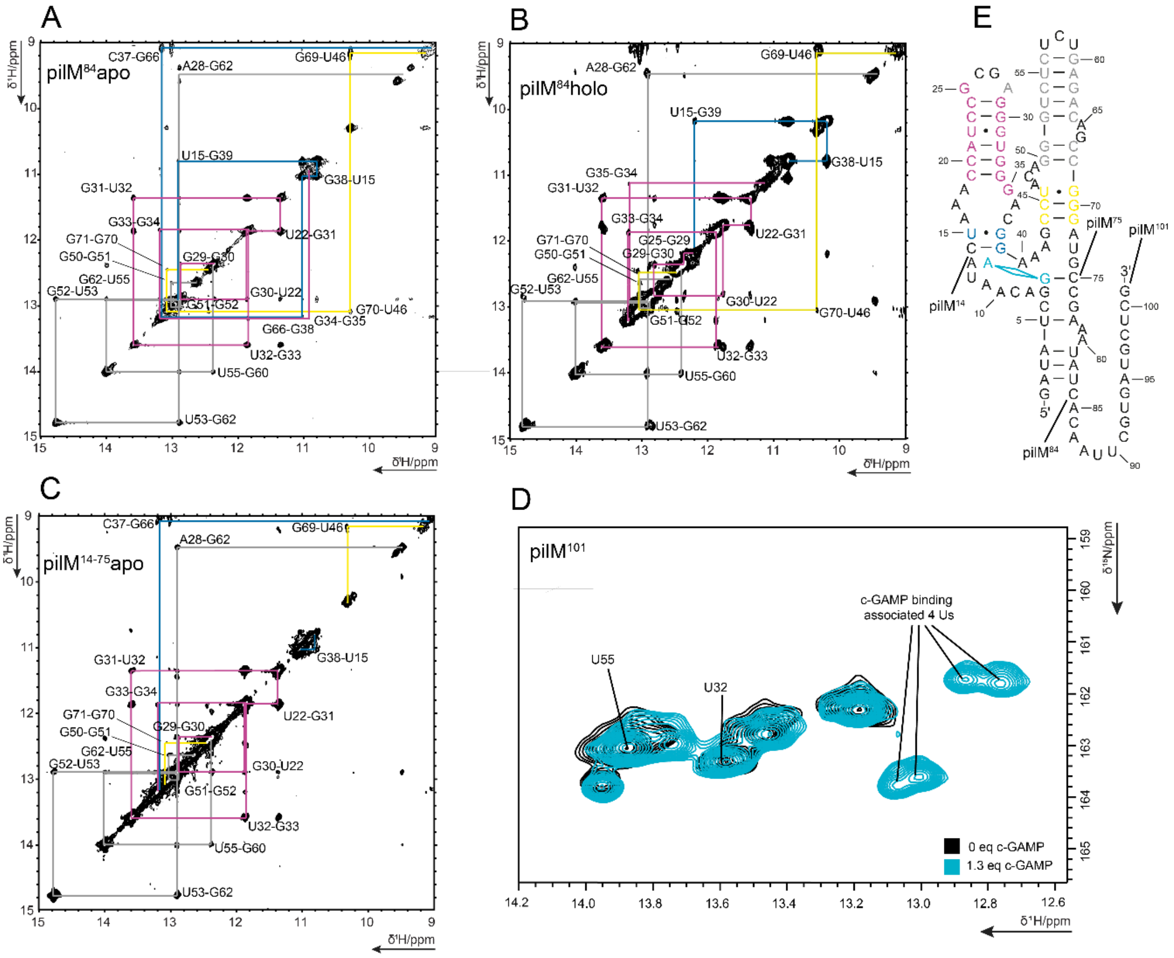


Figure 11 Imino region of $^1H/^1H$ NOESY spectrum for pilM^{84} in the *apo* state (A), *holo* state (B), and double truncated pilM$^{14\text{-}75}$ (C). (D) shows an overlay of the pilM^{101} *apo* (black) and a *holo* (blue) spectrum with four additional Us resulting from c-GAMP addition. (E) shows the pilM secondary structure indicating the key last residues and the start residue of pilM^{84}, pilM ^{101}, and pilM$^{14\text{-}75}$. The assignment walks of (A), (B), and (C) are indicated in the same color scheme as was used for residues in the secondary structures in (E). Figure as published in Landgraf *et al.* (1).

2.2.2 Affinity analyses of different transcript lengths

A transcriptional riboswitch must be able to switch within the time frame set by the transcription. The shortest and longest ligand binding transcripts were determined and confirmed through NMR and ITC. From the known aptamer lengths as in literature, the construct lengths were increased or decreased until ligand binding was no longer occurring. It was possible to distinguish between additional residues for stems that are anticipated to provide new imino signals and loop residues that will not contribute any more signals based on the projected secondary structure of pilM.

When pilM transcripts were subjected to ligand, a ligand-binding-sensitive window for transcript lengths spanning nucleotides 77 to 102 was observed. Interestingly, while pilM^{77} did not show a change in the NMR-spectra in the absence or presence of ligand, a K_D of 19 µM was measured via ITC (Table 2). The complete aptamer can form at a length of 84 nucleotides with a K_D of 0.44 µM.

When the ligand is added, the 1D NMR spectra of pilM^{84} display a downfield shift of U53 at 14.8 ppm and a distinct shift at 11 ppm (Figure 12A). The distinct shift corresponds to a structural rearrangement in the GU base pair closest to the binding pocket. The U15 and G38 signals are shifted upfield. All studied constructions with lengths between 79 and 101 nucleotides could be reliably seen to undergo these and multiple other changes. These ligand-binding-related chemical shift changes were no longer discernible for pilM^{102}. A dramatic rearrangement of the RNA fold before the ligand is added was found for pilM^{109} and longer constructs. The absence of the two reporter peaks resonant at 14.8 ppm and 11 ppm, respectively, attests to the loss of *apo* fold (Figure 12A).

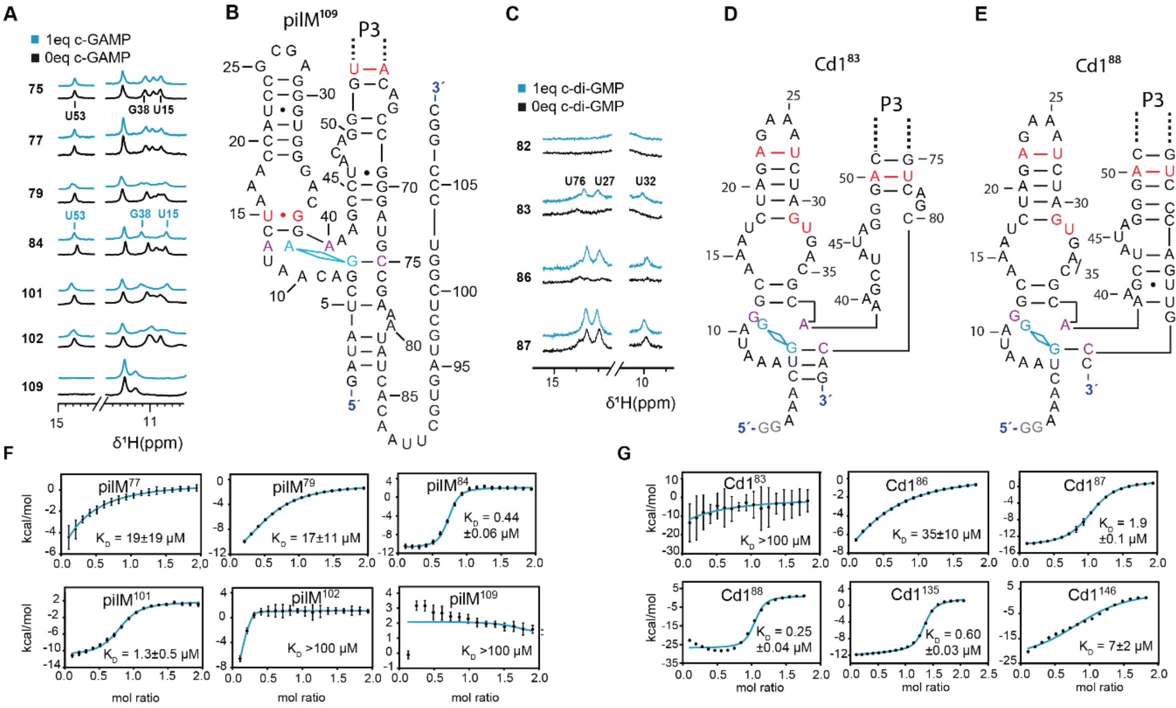


Figure 12 (A) and (C) show the NMR reporter signal regions of a ^{1}H-NMR titration for c-GAMP-binding to pilM-RNAs ending on nucleotides 75, 77, 79, 84, 101, 102, and 109 and for c-di-GMP binding to Cd1-RNAs ending on nucleotides 82, 83, 86, and 87. Indicated in blue are the spectra of metabolite-containing samples and in black samples without. Complete NMR spectra of the shown datasets and all other lengths measured for pilM are shown in SI Figure 2. The full peak lists can be found in SI Table 2 and Landgraf *et al*. (1). (B), (D), and (E) show the secondary structure models for ligand recognition. Indicated in violet are the key nucleotides of the binding pocket, and highlighted in red are the bases contributing to the reporter signals. Sections (F) and (G) show examples of the ITC titration curves of key constructs in outlining the ligand recognition window. Values and errors in the K_D are derived from triplicate measurements. Error bars indicated in the ITC plots refer to the individual fit. Figure as published in Landgraf *et al.* (1).

The affinity of the c-GAMP pilM interaction showed a wave-shaped trend with increasing construct length. Initially, the affinity drastically and stepwise increased starting at 19 µM at pilM^{77} and increasing until pilM88 with 0.25 µM affinity. The affinity subsequently decreased until it reached a local minimum at pilM96 at 2.9 µM. The maximum affinity (0.25 µM) was reached again at pilM^{99} and dropped to 1.3 µM at pilM101. Longer constructs were no longer responsive to c-GAMP addition. Overall, the difference in affinity between the lowest and highest K_D determined was 76-fold. k_{off} and k_{on} rates were obtained from the ITC titration curves using the kinITC method (Table 2) (72). The obtained values ranged from 4000 $M^{-1}s^{-1}$ to 32000 $M^{-1}s^{-1}$ for k_{on} and 0.0037 s-1 to 0.032 s^{-1} for k_{off}. While the difference between the highest and lowest value of k_{on} and k_{off} was 8- to 9-fold, the standard distribution around their corresponding means was substantially smaller for the k_{on} values. The mean k_{off} rate was 0.0125 s^{-1} with a standard deviation of 0.0110 $s^{-1,}$ and the mean k_{on} rate was 17500 $M^{-1}s^{-1}$ with a standard deviation of 7900 $M^{-1}s^{-1}$. Additionally, changes in K_D were strongly correlated with changes in the k_{off} rate. A similar correlation could not be observed for the k_{on} rates.

Table 2 K_D determination by ITC for ligand binding to riboswitch RNA transcripts of different lengths. Values were measured as triplicates. Errors represent the standard deviation of the determined triplicate values. Sequence-aligned constructs are shown in the same line. *: values determined after the publication of Landgraf *et al.* (1).

Construct	K_D/µM for c-GAMP	k_{on}/$M^{-1}s^{-1}$ for c-GAMP	k_{off}/s^{-1} for c-GAMP	Construct	K_D/µM for c-di-GMP	k_{on}/$M^{-1}s^{-1}$ for c-di-GMP	k_{off}/s^{-1} for c-di-GMP
				$Cd1^{83}$	>100		
				$Cd1^{86}$	35±10		
$pilM^{75}$	>100			$Cd1^{87}$	1.9±0.1	6300±800	0.0119±0.0015
$pilM^{14-75}$	12±5						
$pilM^{77}$	19±19			$Cd1^{88}$	0.25±0.04	21000±300	0.0045±0.0005
$pilM^{79}$	17±11						
$pilM^{80}$	8.5±1.4	4000±1700*	0.030±0.014*				
$pilM^{82}$	2.12±0.17	18000±8000	0.028±0.012				
$pilM^{83}$	1.67±0.09	9800±1700*	0.013±0.002*				
$pilM^{84}$	0.44±0.06	28000±7000	0.0037±0.0009				
$pilM^{88}$	0.25±0.04	17000±3000	0.0042±0.0005				
$pilM^{91}$	0.32±0.05	18000±3000	0.0040±0.006				
$pilM^{93}$	0.38±0.06	13900±1700	0.0037±0.0005	$Cd1^{135}$	0.59±0,04	11900±1600	0.0084±0.0011
$pilM^{95}$	1.6±0.6	32000±14000*	0.032±0.014*				
$pilM^{96}$	2.9±1.5	15000±4000*	0.021±0.005*				
$pilM^{97}$	1.6±0.7	8000±900*	0.0071±0.0008*				
$pilM^{99}$	0.25±0.04	18000±3000*	0.0043±0.0006*				
$pilM^{100}$	0.5±0.3	23000±5000*	0.0049±0.0008*	$Cd1^{146}$	7±2		
$pilM^{101}$	1.3±0.5	10000±1500*	0.0072±0.0010*				
$pilM^{102}$	>100						
$pilM^{103}$	>100						
$pilM^{105}$	>100						
$pilM^{106}$	>100						
$pilM^{109}$	>100						

2.2.3 Modeling of the cotranscriptional conformation landscape of CDN riboswitches

Based on the experimental results presented above, a Markov model with kinetic rates derived by ITC measurements and standard base closure rates, as reported in Fürtig et al. 2007, was created to derive a prediction model for riboswitch function (Figure 13) (185). The model in this work advanced the Markov model used to study the guanine-sensing riboswitch (31). It uses three parallel macrostates linked by transition rates for each length of RNA transcript (SI Figure 5).

The three states of pilM are an *apo*/antiterminator fold without ligand binding, a competing terminator fold, and a *holo*/antiterminator fold with ligand binding. The pilM equivalent states of Cd1 are a ligand-free *apo*/terminator fold, a competing antiterminator fold, and a ligand-bound *holo*/terminator fold. The rate-limiting steps enabling cotranscriptional refolding between the terminator and antiterminator are linked to the number of base pairs required to break before they can subsequently form base pairs in the alternative fold. ITC was used to determine the association and dissociation rate constants of ligand binding. These rates are the rates of the *apo holo* transitions of the model. The numerically lower rate, corresponding to the higher reaction barrier, was employed for state transitions requiring a binding event (k_{on} or k_{off}) and a refolding event. A single rate approximates the underlying dynamics to describe these transitions. The errors introduced by this single-rate approach were assessed and deemed nonsignificant. Base pair breaking is not required for the initial formation of the *apo* folds and the first transitions of the aptamer competing folds because the secondary structure does not contain already formed stems. The model accounts for this by starting with a 100% *apo* population and a transition rate to the aptamer competing fold of 0 since these structures can only form at greater lengths of RNA. For the first transitions to the aptamer competing fold, a base pair closing rate is introduced for states that can form without breaking base pairs. These base-pair closure events are known to be extremely fast, typically occurring in the ns time range (186). The lower numerical boundary for the base closing rate was tested, and as shown in Figure 13D and E, the model yields close identical results for base closing rates between 100 and 1000 nt/s for Cd1 and pilM. Therefore, base pair closing rates slower than 10 ms would impact the model. Since previous works showed base pair closing to occur orders of magnitude faster, the base closing rate was set to 400 nt/s (2.5 ms) for all further calculations, making the observation independent from this variable.

Ligand-dependence

To simulate the switching behavior of the model for Cd1 and pilM, changes in c-di-GMP and c-GAMP concentrations between 100 nM and 100 µM were simulated. The Cd1 simulation yields a spectrum for <5% terminator fold at low concentrations and a maximum of 90-95% at high c-di-GMP concentrations, in line with the regulation of an OFF switch (Figure 13A). In the ON switch pilM, low c-GAMP concentration resulted in >95% terminator fold, and high c-GAMP concentration resulted in 20-25%. The *apo* state populations decreased readily at all concentrations but faster at higher ligand concentrations.

This ligand-dependent behavior is expected for a metastable *apo* state where ligand binding transitions are possible at earlier stages of transcription than the refolding into *apo* competing folds. Similarly, the Cd1 antiterminator population and the pilM *holo* population continuously increased. The populations were at their highest at the end of the simulated time frames. Transient maxima were observed in the Cd1 *holo* and pilM terminator populations.

For Cd1, the transient maximum is reached after 2.5 s for concentrations of 40 µM and above of ligand. The pilM maximum is reached after 1 s and appears only at concentrations above 40 µM.

These maxima and the subsequent fall in population indicate that these states are readily populated at high ligand concentrations and subsequently lose population density to their competing folds. The long transcriptional window until the Cd1 antiterminator can be formed allows the saturation of the *holo* state. Ligand concentrations of 60 µM and higher result in a >90% population. Increased distances between the contour lines also indicate the saturation at these concentrations. The slight drop in the Cd1 *holo* population after 2.5 s is due to a population shift towards the antiterminator, favored when the *holo* state is highly populated.

In the case of pilM, the terminator starts forming while the *holo* population is still readily increasing. In the early phase of the terminator transcription, only the initial base pairs are formed, and the terminator only contains two newly formed base pairs, placing both folds in a similar stability range. As a result, transitions between pilM *holo* and terminator can occur. At high concentrations (>40 µM), this favors the *holo* state of pilM. *Holo* is only transiently favored, and as previously stated, we observed that full-length pilM is posttranscriptionally in the terminator fold and did not bind ligand. The pilM terminator state is also favored in high nt/s transcription speed simulation after extended transcription times (Figure 13C). These simulation time points are equivalent to a prolonged presence of full-length riboswitch and show the expected behavior of long-term favoring terminator fold.

Transcription speed dependence

The *in vivo* transcription rates of *C. difficile* and *G. metallireducens* are unknown. Therefore, a synthesis rate of 20 nt/s was chosen for the model as an average value for transcription rates described in the literature (26–29). Synthesis rates ranging from 1 to 100 nt/s were simulated to determine the limits of this assumption (Figure 13B and C).

Under low CDN conditions, an increasingly fast buildup of 100% antiterminator fold for Cd1 and terminator for pilM can be observed. After a timeframe that allowed all transcriptions independent of speed to transcribe their corresponding riboswitches fully, Cd1 still shows a population of 100% antiterminator after 60 s while pilM decreases to 90-95% after 30 s.

The equivalent simulation under high c-di-GMP conditions results in a maximum of 85-90% Cd1 terminator populations for transcription rates around 20 nt/s. Increased or decreased transcription speeds result in a higher antiterminator fold population, reaching an equal population of the terminator and antiterminator states at rates over 90 nt/s. At low c-GAMP concentration, pilM showed a population of 90-95% terminator fold and the absence of the *holo* state over the full range of simulated transcription speeds (Figure 13B). Under high c-GAMP concentration, the ratio between the *holo* and terminator is strongly influenced by transcription speed (Figure 13C). Starting with 10% terminator and 90% *holo* fold at low transcription rates, the populations of terminator and *holo* states begin to equalize with rising transcription speeds. Terminator and *holo* are equally populated at transcription rates of about 50 nt/s. At higher transcription speeds exceeding 90 nt/s, the terminator population approaches 70% and the *holo* population 30%.

While both riboswitches demonstrate that they can carry out their biological function over a wide range of inducer ligand concentrations, transcription speed strongly impacts the populations of the functionally essential states for regulation. The pilM riboswitch is significantly more sensitive to the transcription speed of the RNA-polymerase than the Cd1 riboswitch.

Optimal transcription speeds for the persistence of terminator fold at low CDN concentrations

Independent of ligand concentration, a higher transcription speed results in a faster formation of RNA lengths that allow refolding. The effect of transcription speed is best observed in the *apo* and antiterminator population of Cd1, shown in Figure 13B. A very sharp transition between states can be observed because the base pair closing rate is faster than the highest transcription rate analyzed, 400 nt/s, compared to 100 nt/s. These fast closing rates are present for a stretch of 12 nucleotides between G134 and U145 of Cd1.

A second trend could be observed for the low-concentration simulation of pilM (Figure 4B). The second trend represents a population drifting towards a new equilibrium after the initially complete transition from *apo* to the terminator population. The highly populated terminator state refolds back into the *apo* state. The latest time point of this population change could be observed at a speed of 40 nt/s where the <95% threshold of the terminator is only surpassed after 20 seconds. Faster or slower transcription speeds resulted in a quicker drop below 95%, indicating the terminator was either maximally populated or more persistent at this transcription speed. Since the riboswitch function relies on the persistence of the terminator, the optimum of 40 nt/s, where the best riboswitch performance occurs, is close to the 20 nt/s, representing common *in vivo* transcription rates. The formation of a thermodynamic equilibrium for Cd1 in the timeframe of the simulation is likely not observed because the readily formed antiterminator is very persistent to refolding due to the ten base pairs formed in its formation.

Optimal transcription speeds for the persistence of holo fold

Similarly, transcription speeds that maximized the *holo* population were observed for Cd1 and pilM (Figure 13C). Cd1 showed an ideal speed for the population of the *holo* state at 20 nt/s, where a population of over 90% persists for 55 s. Faster or slower transcription speeds resulted in a quicker drop below 90% and a faster population of the antiterminator state. pilM showed an ideal speed for the population of the *holo* state at 5-10 nt/s, where the *holo* population remains above 95% for 15 s. Faster or slower transcription speeds resulted in a quicker drop below this level and a faster population of the terminator state. The model can not differentiate between more populated or persistent macrostates at different transcription speeds. pilM showed a reversal of the *holo* terminator equilibrium favoring the transition to the *holo* state at a transcription speed higher than 40 nt/s. The change in equilibrium can be observed *holo* state by contour lines transitioning from a downward slope to an upward slope at increased transcription speeds. An inversion of this equilibrium was not observed for Cd1 within the range of analyzed transcription speeds.

A transient saturation of the Cd1 *holo* state of over 95% can be observed for very slow transcription speeds. The contour lines in this area show an inversely proportional behavior toward the transcription speed, and these lines are matched in the corresponding antiterminator contour plot. They indicate that these speeds are so slow that Cd1 can fully transition into the *holo* state before a length is transcribed, which allows the formation of the antiterminator. Because Cd1 intermediately favors antiterminator formation, these slow speeds result in a more substantial decrease in *holo* state population than the optimal 20 nt/s speed. Very slow transcription speeds also allow an uncompeted pilM *holo* state formation. Yet, even at 1 nt/s, the *holo* state does not reach saturation and still increases after the terminator fold can be formed.

These optimal transcription speeds for Cd1 and pilM *holo* state population combine, maximizing the time of *holo* formation before aptamer competing fold is possible and minimizing the time available for forming aptamer competing fold. The optimal speed depends on the size of the underlying transcription windows, the difference in stability between the folds at each length, and the corresponding transition rates and, as a result, requires the deterministic simulation of the system.

Effect of pausing

A possible putative stop site between nucleotides 141 and 145 can be found in the sequence of the Cd1 riboswitch. A simulation of 10 s pausing at high ligand concentration shows that the maximum of the terminator persistence shifts to a transcription speed of 40 nt/s (Figure 13F). Also, the transcription speed window with the same ratio of terminator population widens with the addition of pausing, resulting in the same riboswitch behavior of the Cd1 riboswitch over an extensive range of transcription speed, only deviating for transcription speed below 15 nt/s. Therefore, pausing at the stop site reduces the transcription speed dependence of Cd1.

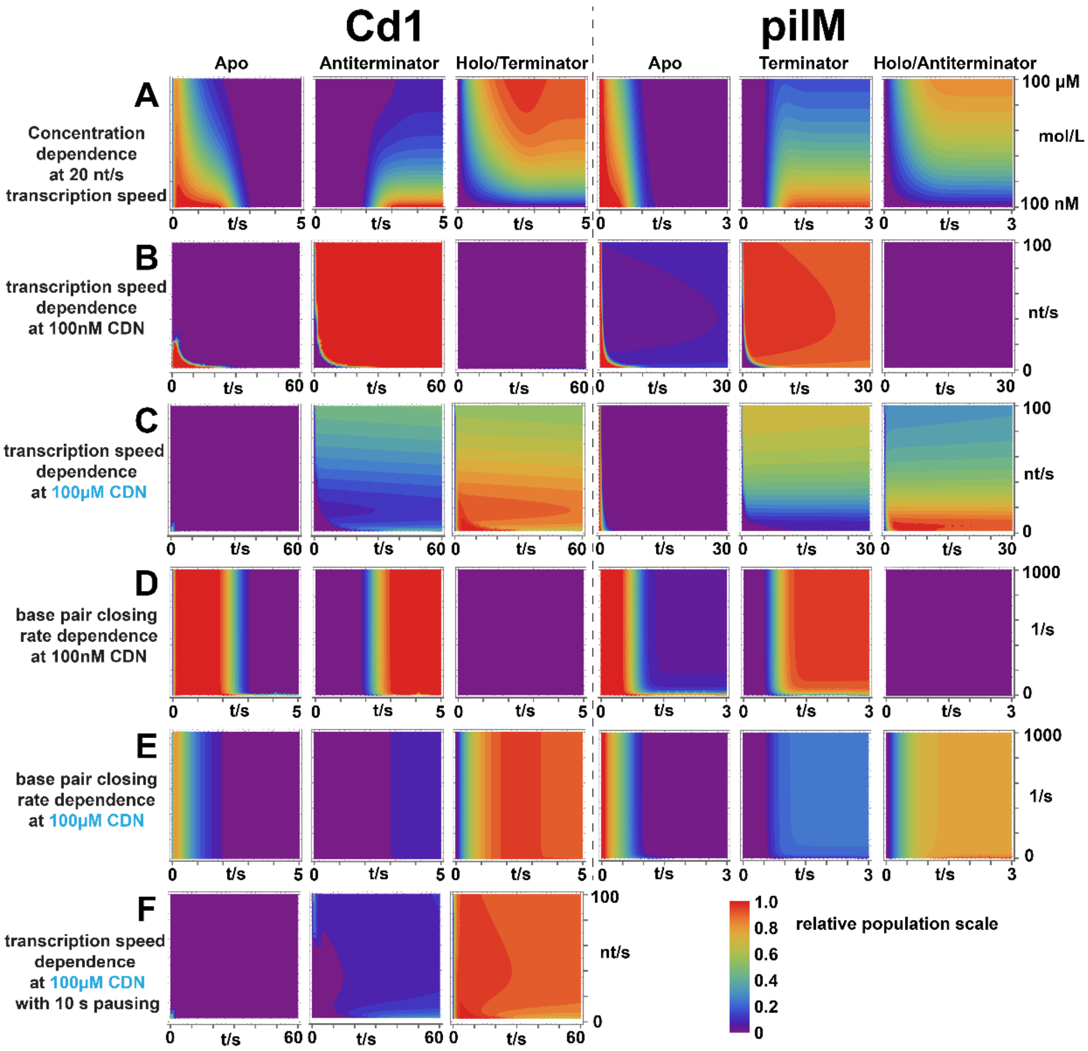


Figure 13 Contour level plots of cotranscriptional folding states for the Cd1 and pilM riboswitch. The population of the three functional relevant states over time is depicted in 5% levels of relative population. The population of the macrostates is the sum of all contributing RNA lengths in the corresponding state. The conditions of the corresponding simulation were 20 nt/s transcription speed, no pausing, 400 s^{-1} base pair closing rate, and either 100 nM or 100 µM cyclic-di-nucleotide (CDN), c-di-GMP for Cd1 and c-GAMP for pilM. The second variables in the contour level plots were (A) the concentration of CDN, (B) the transcription speed at low ligand concentration, (C) the transcription speed at high ligand concentration (simulated from 2 to 100 nt/s),(D) the base pair closing rate at low ligand concentration (simulated from 2 to 100 nt/s),(E) the base pair closing rate at high ligand concentration and (F) the transcription speed at high ligand concentration with 10 s pausing at the possible pause site at nucleotides 141-145 (simulated from 3 to 100 nt/s) (F). Figure as published in Landgraf et al. (1).

2.3 Discussion

Due to their ability to control changes in bacterial lifestyle, the Cd1 and pilM riboswitches have essential biological functions. Knowledge of the internal and external variables that result in the expression of genes linked to a particular lifestyle change requires a comprehensive understanding of the metabolic cycles in which riboswitches play regulatory roles. The signaling pathways and the enzymes that control ligand synthesis and degradation are examples of these external stimuli. Important factors are specific riboswitch-ligand interactions that stabilize the structures and drive the transitions relevant to transcription regulation. This work employed solution NMR, ITC, and computer modeling to examine the dynamics of transcription regulation of the pilM riboswitch. It then compared this data to the Cd1 riboswitch using a Markov model of cotranscriptional folding.

2.3.1 Determination of key transcriptional intermediates of the pilM riboswitch

Single nucleotide extension transcription intermediates were used in NMR investigations to examine the functional conformational transitions of pilM (Figure 12). The base pairing RNA nucleotide for c-GAMP in pilM is known from the crystal structure to be C75 (42). It was determined that PilM^{75} is ligand-binding incompetent (Table 2). A K_D of 18 µM was obtained for pilM^{77} by extending pilM^{75} by just two nucleotides. An extra 5′-end truncation was produced for the construct pilM$^{14\text{-}75}$. K_D measurements of pilM$^{14\text{-}75}$ reveal an affinity that is comparable to pilM^{77}. These findings suggest that pilM^{75}'s binding incapacity is most likely caused by the 5′-end creating an alternate fold that prevents the assembly of the binding pocket. Without a P1 stem, the pilM aptamer can bind the inducer by offering a partial pocket with the Watson-Crick base pairing residue. The pilM aptamer affinity is dramatically increased into the nanomolar range by additional P1 residues. Compared to pilM^{77}, the complete aptamer pilM^{84} has a 43-times higher affinity (Table 2).

Within the binding competent window between pilM^{77} and pilM^{101}, K_D increases with additional nucleotides associated with the P1 stem. Until pilM^{84,} the transcript affinity increases, reaching a high-affinity regime of around 0.4 µM. The transcripts remain in this high-affinity regime until pilM^{95}. The residues associated with the range between 84nt and 95nt are found in the Terminator loop and are not expected to contribute base-pairing to the aptamer or the Terminator fold. Very little influence on the transcripts' affinity is in line with the expected influence based on the secondary structure. The remaining six transcripts, until the transcripts are observed as no longer binding competent, can be observed in 2 different affinity regimes. PilM^{99} and pilM^{100} are high-affinity binding like pilM^{84} and contain a fully formed P1 stem based on their affinities. While pilM^{95}, pilM^{96}, pilM^{97}, and pilM^{101} show a reduced affinity around 1.6 µM, equivalent to pilM^{83,} this could indicate destabilization of the P1 fold due to a formation of truncated Terminator fold. These observations were also utilized in computational modeling, allowing all transitions between the aptamer and truncated Terminator with the transition rates associated with the corresponding amount of refolding base pairs.

PilM^{102} is the initial RNA transcription length beyond which ligand binding is no longer possible. The terminator sequence is present and is of a length that enables the formation of base pairs necessary to maintain the terminator conformation. The terminator also contains C75, the residue necessary for the base pairing of the ligand. The intramolecular interaction between C75 and G102 outperforms the intermolecular contact between the ligand and the aptamer. Helming *et al.* researched the deoxyguanosine-sensing riboswitch, which suggested a broad window of several nucleotide lengths of gradually diminishing binding capability. In contrast, more recent work by Binas *et al.* showed a very sharp transition for transcript lengths differing by a single nucleotide from high to no binding capability (53, 187). This study found a sharp change in the binding capacity for pilM^{101} to pilM^{102} by a single nucleotide extension, which is consistent with the findings of Binas *et al*.

2.3.2 The importance of the P1 helix (P1A)

A conserved three-stem junction is postulated as the binding aptamer by the consensus sequence for GEMM I riboswitches. A conserved sequence, including a fully formed P1, is expected to be necessary for binding (41).

There were no P1 helix-supporting NMR signals found for Cd1 (1). Since nucleotide C87 is the site of c-di-GMP binding, no residues in the secondary structure can form a base-paired P1 helix close to the binding pocket. The K_D drops drastically between Cd1^{87}, 1.9 µM, and Cd1^{88}, 0.25 µM (Table 2). Cd1^{88} reaches an affinity level typical for riboswitches and is in line with other CDN riboswitches (41, 80, 188). Contrary to Cd1, adding a single nucleotide to the transcribed RNA does not significantly improve the affinity of pilM^{77}(Table 2). The affinity continuously increases for 12 nucleotides and reaches a maximum at pilM^{88}, confirming that with the addition of c-GAMP, these pilM RNAs form a stable P1 stem. The P1 stem is only formed through ligand binding. There is no evidence of a formed P1 stem when c-GAMP is absent (Figure 11D). Based on the secondary structure, a complete formation can occur at the length of pilM^{84}. The further decrease in K_D from 0.44 µM to 0.25 µM could be caused by additional nucleotides stacking onto the end of the P1 stem. Neither sequence nor x-ray structure analyses indicate this P1 instability or the stark difference in P1's relative stability compared to P2 and P3. This research matches previous reports of the stability of the P2-P3 and the instability of P1 required for the allosteric change in riboswitches (73, 81, 189).

In literature, the translational OFF switch Vc2 is the most prominent example of a GEMM I switch (190). The stability of the P1 helix, which is essential for achieving tight ligand binding, is the determining element in Vc2 regulation. Without the P1 stem, Vc2 cannot bind its ligand, c-di-GMP (81). Smith *et al.* observed the Vc2 riboswitch increased its affinities to ligand from 3.4 µM for Vc2^{93} to under 0.5 nM for Vc2^{100}. The Vc2 behavior shares the steep increase in affinity observed for Cd1 while occurring over a comparable length of nucleotides to pilM, 7 compared to 12. Given that full-length wildtype has an affinity of 0.011 nM, Vc2^{100} is probably not the highest affinity intermediate, and the increasing affinity trend could continue beyond 7 nucleotides.

In CDN riboswitches, a highly conserved aptamer scaffold that functions as the binding pocket is represented by the P2 and P3 stems. So conserved in comparison to the P1 that the initial description where a riboswitch function was suggested exclusively featured the P2 and P3 stem (48). Further affinity fine-tuning appears to be a possibility through changes in the P1. Given that the CDN concentration is regulated by a network of several CDN-forming and CDN-hydrolyzing proteins, the matching of riboswitch sensitivity to the predominant CDN concentration in the associated organism would be possible (75). The affinity tuning through P1 stability is more impactful for riboswitches that approach saturation during transcription, operate under thermodynamic control like Cd1, and could be adjusted over orders of magnitude. Changes to the P1 are generally only associated with minor changes in k_{on} rate (76). P1 changes are, therefore, less influential to riboswitches under kinetic control but would allow fine adjustments. Similarly, pausing strongly increases transcription and, in turn, strongly influences the affinity range of transcriptional riboswitches. A fine adjustment of the regulatory range could be accomplished through changes to the length of the binding competent window by increasing the sequence length (52).

2.3.3 Cotranscriptional modeling parameters and regulatory limits of kinetic and thermodynamic control

The three-state Markov model provides a detailed insight into the gene regulatory function of the riboswitches and allows comparing an ON- or an OFF-switch despite their opposite switching logic. The physiological concentration associated with the bacterial gene regulation of the GEMM motif matched the model's sensitivity to CDNs (191–194). The pilM and the Cd1 riboswitches exhibited a very low *holo* population for low ligand concentration (100 nM) and a *holo* population of 80% for pilM and 95% for Cd1 for high ligand concentration (100 µM). These findings confirm that both riboswitches operate within the timeframes offered at typical transcription speeds. Additional variables allowed the codependent modeling of transcription speed, pause sites, the rate of closing base pairs and an assessment of their impact on riboswitch regulation. Transcription speed strongly affected both riboswitches, always resulting in increased populations of the ligand-bound states at low speeds. A lower speed allows a longer presence of binding-competent states, increasing the probability of forming a stable *holo* complex. The stable ligand-bound RNA complexes will not reach equilibrium for 220 s (Cd1[88]) and 270 s (pilM[84]), as determined by the inverse ITC-derived k_{off} rates. Due to the stability of these complexes, ligand release is unlikely to occur within the corresponding transcription time frame. Consequently, the riboswitches can not reach thermodynamic equilibrium.

Yet whether the riboswitches exert effectively thermodynamic or kinetic control requires further discussion. The basis for this discussion is the definition of kinetic and thermodynamic control regimes introduced by Wickiser *et al.* (57). In high regulatory ligand concentration, orders of magnitude higher than the K_D, and with fast association rate constants, riboswitches can approach the near total population of the *holo* state at time scales shorter than the inverse k_{off} rate. This near-total population would also form under thermodynamic control. At this limit, k_{off} is a poor descriptor of the control regimes and riboswitch regulation. The regulatory concentration range is better understood as dependent on the ligand association time ($t_{on}=(k_{on}*c_{Lig})^{-1}$) relative to the transcription time. Riboswitch regulation can be assigned to kinetic or thermodynamic control based on the relative fraction of *holo* formed during the ligand-binding competent time windows relative to the equilibrium population of *holo*. If less than 75% *holo* is formed during transcription, the riboswitch is kinetically controlled and shows a linear response to changes in ligand concentration. Over 95% is classified as thermodynamic control, and the 75%-95% range is called mixed control. Given the *holo* population of 80% for pilM and 95% for Cd1 for high ligand concentration, pilM would be classified as mixed control near the upper limit of kinetic control, and Cd1 would be classified as thermodynamically controlled.

2.3.4 Pausing reduces ligand dependence of Cd1 and indicates thermodynamic control with a regulatory window at lower concentrations.

For both riboswitches, switching efficiency decreased at high transcription speeds (Figure 13A). The high sensitivity of the pilM riboswitch regulatory function to changes in transcription speed is observed, while Cd1 does not show this degree of sensitivity. With a 30% ON state at 100 nt/s transcription speed, the pilM riboswitch's efficiency decreases more rapidly than the Cd1 riboswitch. The sensitivity of the riboswitch to variations in transcription speed suggests the potential for higher-level control through speed adjustments. This sensitivity is also a hallmark of kinetic riboswitch control, and our transcription speed-dependent contour plots at high ligand concentrations (Figure 13c) allow a qualitative assessment of the control character.

Compared to pilM, the Cd1 riboswitch is less influenced by changes in transcription speed due to its longer expression platform ahead of the downstream genetic decision point (Figure 10). With an additional 10 s pausing delay, the switching efficiency of Cd1 increases to 95% for a transcription speed range of 5-100 nt/s, presumably remaining high beyond 100 nt/s. In this instance, the riboswitch sensitivity should be independent of the transcription rate, a hallmark of thermodynamic control. Given the Cd1[88] riboswitch aptamer affinity of 0.25 µM, a confirmed pause site would suggest a regulatory window at lower concentrations than modeled in our analysis based on concentration ranges typically associated with lifestyle changes.

A pause site analysis using native *C. difficile* RNA polymerase similar to studies by Steinert *et al.* and Helmling *et al.* on their respective riboswitches could resolve this question (31, 53, 54). Local *in vivo* concentrations of CDNs remain elusive and are regulated by a complex network (195). A better understanding of lifestyle-associated local CDN concentrations could establish upper and lower boundaries of riboswitch regulation and improve future models. A caveat in this regard is that CDN concentration obtained from cell lysates might not represent concentrations found at the site of transcription due to the known strong localization of cyclases and phosphodiesterases at bacterial poles (196).

2.3.5 Median response concentrations and regulatory limits of kinetic and thermodynamic control

We observed the k_{on} rates of pilM aptamers as essentially constant throughout the riboswitch elongation (Table 2). Given constant k_{on} rates and the linear codependence of k_{on} rate, ligand concentration, and transcription time in kinetic control regimes (167), we can establish regulatory ligand concentration ranges around median response values and thresholds for the transition from kinetic to thermodynamic control. Kinetic control assumes a vast excess of ligand concentration compared to the aptamer and short binding windows that make a k_{off} rate contribution negligibly small. For ease of comparison, we omitted the formation of aptamer competing fold, which could interfere with ligand binding.

We determined the concentrations leading to a median response (T50), described as 50% binding to the aptamer. The transcription speed of 20 nt/s ($v_{\text{transcription}}$) yields transcription times ($t_{\text{transcription}}$) of 3 s (Cd1) and 1.2 s (pilM) of the ligand-binding component window. Using a k_{on} rate of 17500 $M^{-1}s^{-1}$ for both aptamers and applying these values in the formula,

$$P_{apo} = e^{-k_{on}*c_{Lig}*v_{\text{transcription}}*l_{binding}} = e^{-k_{on}*c_{Lig*}*t_{\text{transcription}}}$$

The equation can be solved to yield the concentration that would result in a median response,

$$\frac{ln(0.5)}{-k_{on} * t_{\text{transcription}}} = c_{Lig}$$

Therefore, a median response is expected at a concentration of around 13 µM CDN for Cd1 and 33 µM for pilM (Table 3). Adding 10 s of pausing would reduce the CDN concentration of the median response for Cd1 further to 3 µM. These trends continue for the kinetic control (less than 75% *holo*) and thermodynamic control (over 95% *holo*). These regulatory thresholds need to be compared to the regulatory concentrations assumed in our model (100 nM to 100 µM).

In the assumed regulatory concentrations, pilM operates in the transition zone between kinetic and thermodynamic control (Figure 14). The upper limit of 100 µM falls between the threshold of 66µM and 143 µM. In our Markov model, where terminator formation is not omitted, over 90 µM ligand is required to leave the kinetic control regime (Figure 13a). Given the absence of pause sites, we think it is very likely that the assumed regulatory ranges are present *in vivo* and required to drive the riboswitch. The riboswitch utilizes the full regulatory range and operates fast enough to accommodate just 1.2 s of transcription time. In the similar pubE riboswitch (also a transcriptional ON switch with a short expression platform), low *in vivo* ligand concentrations have been found to necessitate pausing, while hypothetical higher ligand concentration could eliminate the requirement for pausing (167). Under kinetic control, increases in ligand concentration result in a proportional increase of *holo* formation. The system operates under nonsaturating conditions, and a steeper ligand response curve is observed (57).

Without pausing, the Cd1 riboswitch operates under thermodynamic control (Figure 14). The range from 57 µM to the assumed upper limit of the regulatory ligand concentration 100 µM contributes less than 5% to gene suppression. A reduction of gene expression of 50% would be achieved at 13 µM ligand. A large part of the dynamic range of ligand concentration is allocated to suppressing the residual 50% of gene expression because the system approaches saturation. When pausing is introduced, the corresponding thresholds drop even lower. Signaling ligand concentration over 13 µM would result in increasingly diminishing regulatory effects. Absent an extreme need to reduce basal expression or a remarkably large dynamic range for ligand concentration, the conformation of a pause site would suggest that the Cd1 riboswitch operates at a lower regulatory concentration than modeled.

Table 3 Ligand concentration values associated with median ligand response, kinetic control, and thermodynamic control. Values assume no *holo*-complex dissociation, no formation of aptamer competing fold during the transcription time interval, and a large excess of ligand. A transcription speed of 20 nt/s and a k_{on} rate of 17500 $M^{-1}s^{-1}$ were assumed for both aptamers. Threshold values of kinetic and thermodynamic control were applied based on the limits Wickiser *et al*. set (57).

Transcription time	Median response (T50)/µM	Threshold for kinetic control/µM	Threshold for thermodynamic control/µM
pilM 1.2 s	33	<66	>143
Cd1 3 s	13	<26	>57
Cd1 3 s +10 s pausing	3	<6	>13

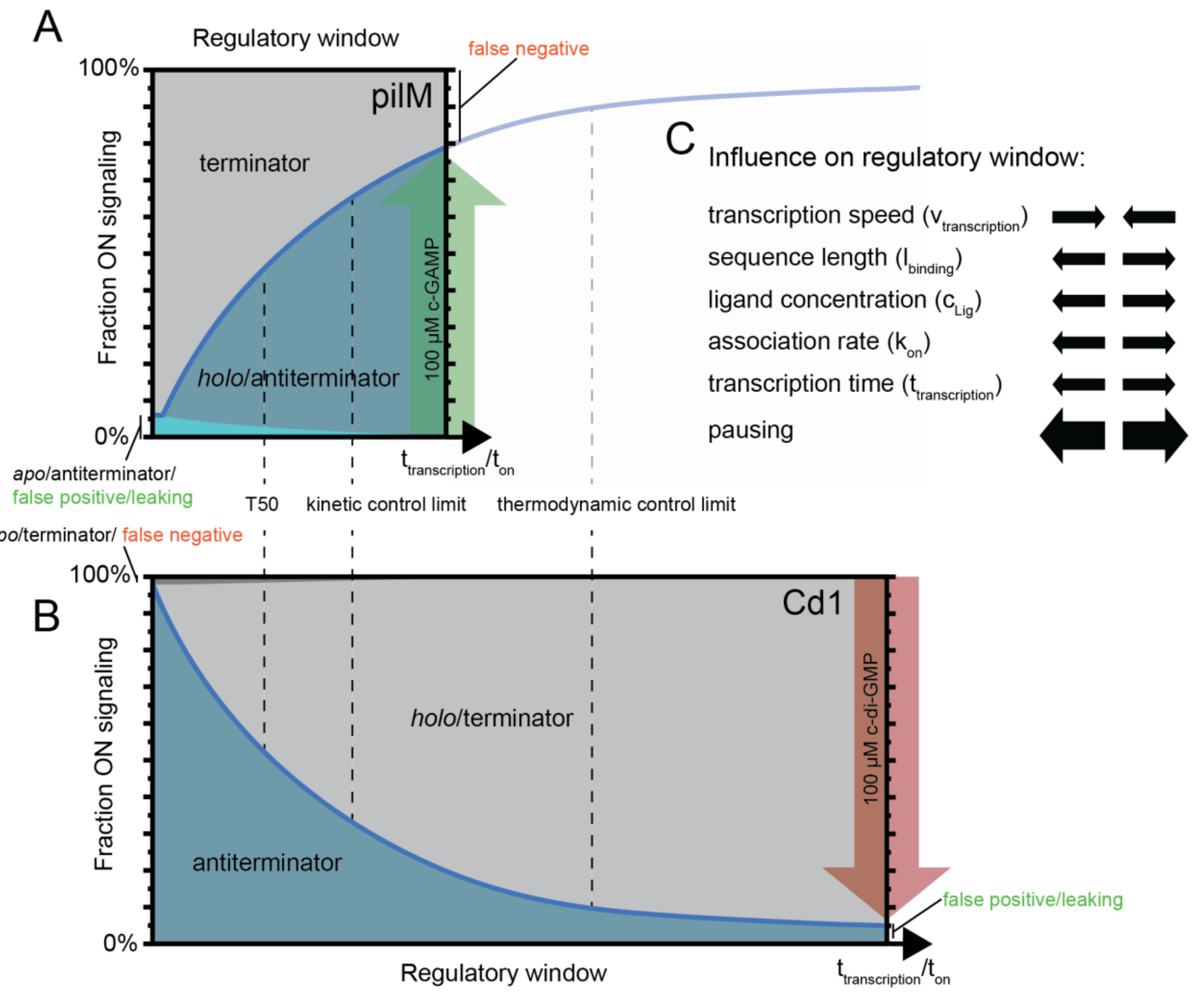


Figure 14 Kinetic model of binding to pilM (A) and Cd1 (B) during regulatory window assuming the same transcription speed (20 nt/s), association rate (17500 $M^{-1}s^{-1}$), and regulatory CDN concentration (100 µM). Fraction of ON signaling is indicated by blue line. The difference in the size of the regulatory window is caused by pilM having 60% fewer binding competent transcriptional intermediates. RNA states associated with ON signaling are marked in blue-shaded areas. RNA states associated with OFF signaling are marked in gray-shaded areas. Fractions of leaking, false positive, and false negative signaling are marked on the two y-axis representing the the lower and upper limit of regulating CDN concentration. The left limit represents $t_{transcription} \ll t_{on}$ and is equivalent to low CDN concentration (100 nM) OFF signaling for pilM and ON signaling for Cd1. The right limit represents $t_{transcription} > t_{on}$ for pilM and $t_{transcription} \gg t_{on}$ for Cd1, this is equivalant to high CDN concentration (100 µM) ON signaling for pilM and OFF signaling for Cd1. Dashed lines indicate median response (T50), upper limit of kinetic control character, and lower limit of thermodynamic control character as introduced by Wickiser et al. (55). (C) depicts the qualitative effect of regulatory variables on the size of the regulatory window. The size decreases with transcription speed, increases with sequence length, ligand concentration, association rate, and transcription time, and strongly increases with pausing delays.

2.3.6 Base pair closing rate as a major driver of regulatory transitions

The rate for base pair closure is by far the quickest conformational transition during RNA refolding of riboswitches, usually in the ns range (171, 186). Base pair closing drives the transition from the aptamer conformation towards the aptamer competing folds. For pilM, this is the terminator fold. The Markov model predicts that this base pair closure must occur faster than 20 ms for Cd1 or 200 ms for pilM to avoid non-regulatory riboswitch behavior. Speeds faster than these limits in the previously established ns range showed a constant effect on the population ratios evident through the vertical lines in the contour plots (Figure 13D). Therefore, a fast base pair closing rate is necessary for the riboswitch function. It is a major driver of transitions from apo to aptamer competing fold by quickly forming the top part of the terminator or antiterminator hairpin. Further speed increases have no effect because the *apo* state is fully transitioned to the aptamer competing fold. PilM requires a faster base pair closing rate to reach this population distribution because it contains a shorter stretch of only two residues, U93 and G94, that can close into two base pairs without requiring the melting of another base pair. The Cd1 riboswitch contains a long stretch of ten base pairs between G134 and U145 that can form base pairs in this way. The residues G138 and C139 do not contribute base pairs in the Cd1 antiterminator. This difference in preformed base pairs makes subsequent transitions of the Cd1 riboswitch highly unlikely due to the associated high barrier of unfolding ten base pairs.

To summarize, Cd1 riboswitch has a longer time frame to transition to aptamer competing fold (antiterminator) and higher barriers to leaving this aptamer competing state than pilM. This difference can also be observed in the contour plots of Figure 13B, where transitions back to the *apo* population only occur in the pilM contour plots. These back transitions result in significant ON signaling despite a low signaling CDN concentration associated with OFF signaling. This inverse signaling is commonly called leaking, and a regulator's propensity to allow it is called leakiness. *In vitro* single-round transcription experiments for pilM also showed low leakiness, around 10%, for no or near cognate ligands (41).

2.3.7 Key conformations and time intervals of ON and OFF regulation

The Markov model outlines key conformations and the timescales required for riboswitch-regulated gene transcription (Figure 15). To examine various structural arrangements, Helmling et al. and Steinert and Sochor et al. used the PATH representation(31, 53, 54). The four key strands, P, A, T, and H, assemble to form riboswitches in this model. The aptamer, represented as [PA], is formed when the 5′-aptamer strand P combines with an aptamer-stabilizing strand A. Downstream on the mRNA are the switching strand T and the terminator strand H. In OFF switches like Cd1, the T and H strand form the terminator [TH], and the ligand binding domain stabilizes [PA]. Because PilM's T strand is situated in front of its A strand, it stands out as a variation from other riboswitches.

The pilM riboswitch exhibits a binding-competent $P^{T}A$ structure in a time frame of 1.2 s, whereas the PA structure in the Cd1 riboswitch is present for a window of 3.0 s. In the folding step, pilM folds metastable PT[AH] terminator within 0.4 s. it takes Cd1 1.5 s to form the equivalent transiently stable P[AT] antiterminator fold. The transiently stable Cd1 antiterminator structures must remain stable throughout the transcription for the riboswitch to function. Contrary, the metastable pilM structures at this point of the transcriptional window readily refold by adding ligands. After prolonged periods, the transiently stable Cd1 antiterminator complexes will refold to form the terminator. In high CDN concentration, the two riboswitches reach different levels of *holo*-state populations. The holo states contain the same P^TA and PA structures, yet these structures are stable and persist with half-life times of multiple seconds. Before the antiterminator conformation is folded, the Cd1 riboswitch reaches a *holo* state population of 95% (PA) under the expected *in vivo* OFF signaling conditions. Under the same conditions, pilM reaches 55% *holo* population [P^TA] before the terminator fold can occur. Due to this, pilM has a contested *apo* state P^TA at the point of decision that folds into 20-25% terminator P[TAH] as opposed to Cd1's 5% antiterminator fold P[ATH].

2.3.8 Erroneous gene expression is minimized at optimal transcription speeds and by large dynamic range of signaling for OFF switches

Riboswitch switching efficiency is imperfect, and despite OFF/termination signaling, transcriptions of complete mRNA can occur and result in basal gene expression. This erroneous regulation resulting in gene expression is referred to as leakiness. Our model predicts similar amounts of leakiness of <6% for the pilM and 5-10% for the Cd1 riboswitch. The underlying cause of the leakiness differs between the two riboswitches.

Cd1 being an OFF switch, can only leak in high CDN concentration conditions when it would completely suppress gene transcription under perfect regulation (Figure 14). The equilibrium between the *holo* fold and antiterminator causes the leakiness. Under the conditions chosen for Figure 15, the Cd1 equilibrium stabilizes at 90-95% *holo* and 5-10% antiterminator. Assessing the dependency of the leakiness on CDN concentration, a decrease in CDN results in a higher amount of antiterminator formation. This reduction is equivalent to the inherent regulatory function of a riboswitch. Depending on the levels chosen to represent ON or OFF signaling, at some point, the leakiness turns into gene expression associated with normal riboswitch function.

Transcription speed impacts leakiness as well. The typical bacterial transcription speed of 20 nt/s was optimal for *holo*-fold persistence. It maximizes the holo fold's persistence and, in turn, minimizes leakiness for Cd1 (Figure 13C, antiterminator population). Further increases or decreases in transcription speed resulted in higher leakiness. The minimal leakiness of 5-10% of Cd1 is a result of the 20 nt/s transcription speed allowing for a nearly complete transition from *apo* to *holo* while being fast enough to reach the terminator strand H, thereby limiting the amount of refolding from *holo* into antiterminator.

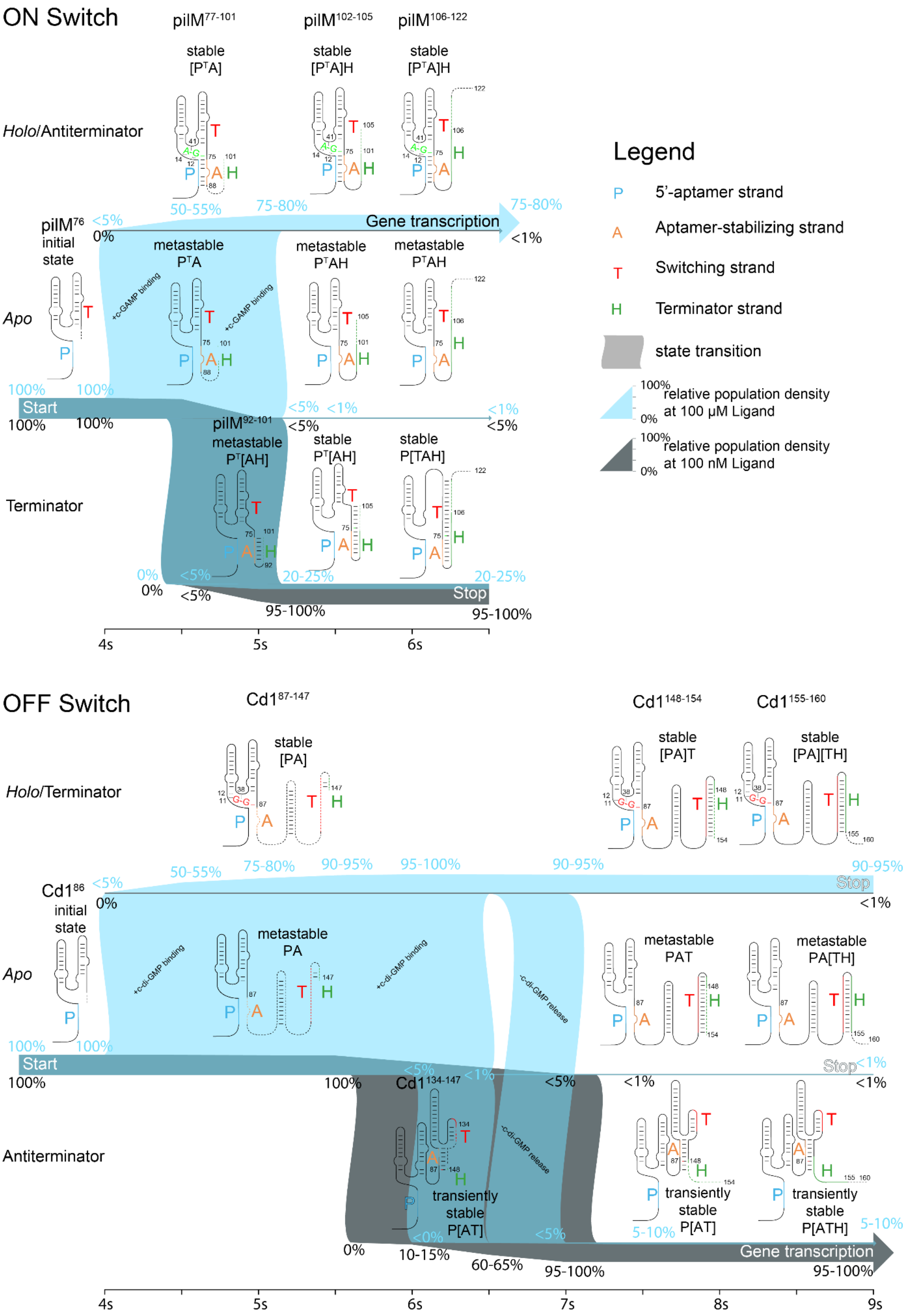

ON Switch
pilM77-101
pilM102-105
pilM106-122
stable [PTA]
stable [PTA]H
stable [PTA]H
Holo/Antiterminator
pilM76 initial state
Apo
Terminator
<5%
0%
50-55%
75-80%
Gene transcription
75-80%
<1%
metastable PTA
metastable PTAH
metastable PTAH
+c-GAMP binding
+c-GAMP binding
100%
100%
Start
100%
100%
pilM92-101 metastable PT[AH]
stable PT[AH]
stable P[TAH]
<5% <1%
<5%
<1%
<5%
0%
<5%
20-25%
20-25%
0%
<5%
Stop
95-100%
95-100%
4s
5s
6s
P
A
T
H
Legend
P 5'-aptamer strand
A Aptamer-stabilizing strand
T Switching strand
H Terminator strand
state transition
100% 0% relative population density at 100 µM Ligand
100% 0% relative population density at 100 nM Ligand
OFF Switch
Cd187-147
stable [PA]
Cd1148-154
stable [PA]T
Cd1155-160
stable [PA][TH]
Holo/Terminator
Cd186 initial state
Apo
Antiterminator
<5%
0%
50-55%
75-80%
90-95%
95-100%
90-95%
90-95%
Stop
<1%
metastable PA
metastable PAT
metastable PA[TH]
+c-di-GMP binding
+c-di-GMP binding
-c-di-GMP release
-c-di-GMP release
100%
100%
Start
100%
<5%
<1%
<5%
<1%
Stop
<1%
<1%
Cd1134-147
transiently stable P[AT]
transiently stable P[AT]
transiently stable P[ATH]
<0%
<5%
5-10%
5-10%
0%
10-15%
60-65%
95-100%
Gene transcription
95-100%
4s
5s
6s
7s
8s
9s

Figure 15 Markov model simulations of cotranscriptional folding state distribution over time for the pilM (ON) and Cd1 (OFF) riboswitch at 20 nt/s transcription speed with no pausing. State population densities are shown for 100 nM ligand concentration (gray, based on data shown in Figure 13B) and 100 µM ligand concentration (blue, based on data shown in Figure 13C). Areas of the Figure where population densities overlap are shown in a gray-blue color. The thickness of the bar is indicative of a relative population ranging from 5% to 100%. Population densities are also shown in percent at 0.5 s increments above and below the population bars if the changes by more than 5% in the timeframe. Population densities below 1% are marked explicitly. The aptamer domain is formed with nucleotides 76 and 86 when the ligand can first be bound. Upon reaching nucleotides 77 and 87, the aptamer domain can bind ligands and transition up into the holo state. Transcription intermediates of length from 77-101 nucleotides for pilM and 87-147 nucleotides for Cd1 are binding capable. The time frames available for these transitions are indicated as state transitions. PilM can fold the terminator from nucleotide 92 onward, and Cd1 can fold the antiterminator starting with nucleotide 134. Both RNAs transition between the three states until reaching a point of decision with transcription lengths 102 and 148 when the binding is impaired. Figure as shown in (1).

Leakiness can only occur in the pilM switch in low signaling molecule concentration environments since it is an ON switch (Figure 14). In these low-concentration environments, the riboswitch function is to terminate transcripts. Every transcript that is transcribed and passes this decision point results in gene expression. This leakiness of <6% is solely based on the transitions between the aptamer competing fold and *apo* since *holo* is present in insignificant amounts.

The pilM riboswitch also has a leakiness minimum at intermediary transcription speeds, in this case at 40 nt/s (Figure 13B, *apo* population). The minimum leakiness at 40 nt/s of pilM is a feature of the riboswitch independent from the ligand concentration. It is caused by the *apo* terminator equilibrium. A less dynamic equilibrium favoring terminator formation could further reduce the leakiness. Yet, this would likely be to the detriment of the fast regulatory transition required under kinetic control (chapters 2.3.2 and 2.3.5).

2.3.9 Riboswitches minimize gene leakiness under thermodynamic control for OFF switches and maximize ligand responsiveness under kinetic control for ON switches

Leaking cannot occur for ON signaling for the pilM riboswitch. The presence of erroneous OFF signaling does not result in gene expression. It reduces gene expression. At no point is mRNA transcribed that was not signaled for. However, the reduced gene expression from erroneous OFF signaling does reduce the absolute dynamic range of ON signaling. The impact on the biological system with imperfect signaling can be assumed to be low. The organisms could adapt to this nonlinear proportional behavior through, for example, increased transcription initiation.

At low ligand concentrations, the Cd1 riboswitch will signal <2% OFF. The ON signal condition folds RNA structures [PA][TH] and PA[TH]. In low ligand concentration of the ON signaling, the *holo* fold [PA][TH] is absent due to the lack of ligand. The equilibrium of the *apo* fold PA[TH] and the antiterminator of Cd1 are less in favor of the *apo* state than in the pilM case, <1%(Cd1) compared to <5%(pilM).

When high ligand concentrations are present, two thermodynamically favored states, in addition to *apo*, compete. These states are *holo* and aptamer competing fold. Under these conditions, pilM still shows 20-25% P[TAH] OFF-signaling, compared to 5-10% of Cd1 ON-signaling. In both cases, this represents a substantial fraction of reverse signaling compared to the signaling expected from a perfect switch under these conditions. For pilM, this is false negative signaling, and for Cd1, this is false positive signaling. The difference is that 20-25% for pilM corresponds to a reduction in the absolute dynamic range. In contrast, the 10-15% represent leakiness for Cd1. The additional transcription time required for Cd1 and the operation under thermodynamic control with transiently over 95% *holo* fold decreases false positive signaling and leakiness for Cd1 compared to pilM. Cd1 minimizes leakiness through thermodynamic control.

PilM operates between <6% to 75-80% gene expression to the edges of its regulatory regimes. Leakiness is inherently tied to *apo* terminator equilibrium. The system can not eliminate leakiness by approaching saturation. It can instead utilize the more abrupt ligand response curve associated with the kinetic control regime (57).

2.3.10 Theoretical analysis from single nucleotide extension

This work used NMR spectroscopy to identify the secondary structures of different transcript lengths. The binding affinities and kinetic rates of the transcript lengths were incorporated into a cotranscriptional model. This model predicts the regulatory outcomes as a function of time and a second variable, such as transcription rate, base pair closing rate, pausing, and changes in ligand concentration. Critical transcriptional intermediates and the temporal dynamics of the state population required for riboswitch function could be identified. The determined rates and the derived contour plots allow classifying pilM as kinetically controlled and Cd1 as thermodynamically controlled based on quantitative and qualitative parameters. We found inherent aspects of leakiness (undesired basal gene expression) favor kinetic control for ON switches and thermodynamic control for OFF switches and that both riboswitches operate in matching regulatory regimes.

These biophysical investigations serve as a reference point, allowing insights into the inner workings of riboswitches and how transcriptional intermediates couple their kinetics to the overall riboswitch mechanism.

Chapter 3 ^{19}F NMR-Based Fragment Screening of RNA

3.1 Introduction

3.1.1 Fragment-based screening for drug discovery

While most validated pharmacological targets are proteins, RNA has recently become a target class of interest. The most popular RNA drug target is the ribosome, a large RNA-protein complex. Most antibiotics halt protein synthesis by targeting the RNA-protein interaction of the ribosome (197). Beyond being a target for antibiotics, RNA has long been regarded as being a difficult target for drug discovery. Recent advances in the field are challenging this school of thought successfully (198–202). A thorough investigation of the non-coding RNA target space is motivated by the clinical efficiency of substances originally discovered as RNA binders (203). Prospective targets include RNA involved in inflammatory processes, cancer, and viral and bacterial infections, examples of a few fields with unmet medical needs and lots of room for innovation. The potential target space also increases with the discovery of new regulatory RNAs, such as riboswitches or sRNAs (204). To battle the multi-drug resistant (MDR) pathogens that constitute a serious health threat to society, riboswitches, in particular, have attracted attention (205, 206). MDRs infections are currently treated with last-resort antibiotics. These drugs mark the last line of defense, and new interventions are highly sought to prevent a leap back to times when today’s benign infections will become a death sentence again. Therefore, one crucial area of research is the creation of medications that target riboswitches (207).

X-ray crystallography and NMR spectroscopy provide essential structural information making them a fundamental methodology for rational drug discovery (208, 209). Computational tools are frequently used with target structure information to refine library design. The progression of medicinal chemistry from initial hits toward the production of blockbuster medications is guided by virtual screening and *in silico* docking of a compound library to known structures of targets (210). For drug development, high-throughput screening necessitates reliable binding detection of a compound library with millions of compounds screened. Usually, only a minimal number of prospective hits are obtained. Screening huge libraries demands many infrastructural and financial resources.

Fragment-based screening is a viable alternative to conventional methods. Fragments frequently are weak binders, and their binding specificity may be lower than usual for hits obtained using traditional methods (211). Therefore, fragment-based drug development necessitates extra chemical engineering to form the first hits into lead compounds, for example, expanding the compound to fit the necessary number of binding contacts with the target or linking two fragments that bind to neighboring binding sites. Recent hit-to-lead research strategies are based on fragment-based drug discovery (212). It has been established that FDA-approved therapeutics targeting proteins can be designed starting with fragment screenings(120, 212–217). Methods to screen RNA range from fluorescent-based assays (218, 219), mass spectrometry(220, 221), small molecule microarrays (SMM)(222, 223), microscale thermophoresis (MST)(224, 225) to NMR spectroscopy(226). The studies commonly focus on a single RNA target. (134, 139, 227). Further details regarding NMR-based methods for drug discovery, comparison of FBS and high throughput screening, experimental challenges of FBS, and targeting RNA with fragment-based screening are discussed in chapters 1.4 to 1.4.3

This fragment-based NMR screening, however, investigated 14 distinct RNAs. The primary technique used for hit identification of the 102 fragments used was ^{19}F-NMR. The screened RNAs were tRNAs (often used as control RNAs in screening), short stem-loop structures, aptamer domains of riboswitches, full-length riboswitches, terminators and antiterminators of riboswitches, and ribozymes (Figure 16).

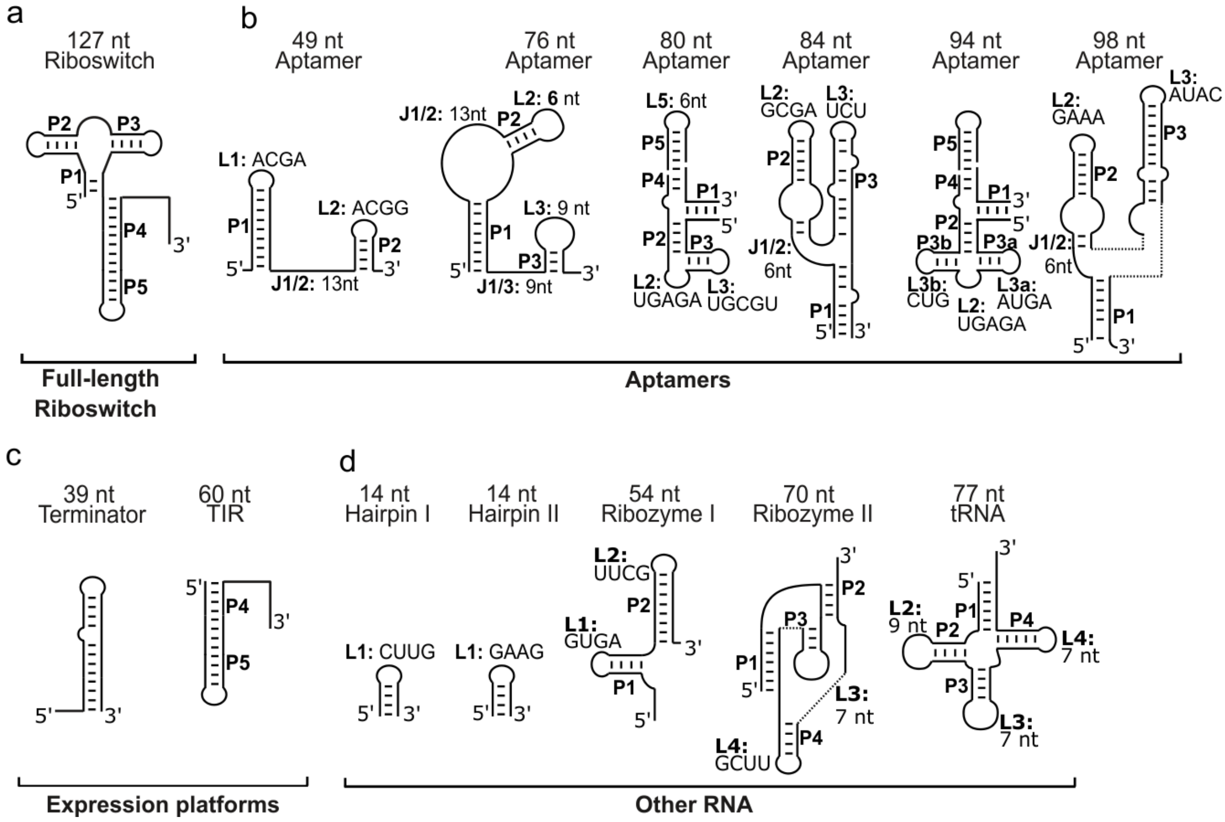

a
127 nt
Riboswitch
P2
P3
P1
5'
P4
3'
P5
Full-length Riboswitch
b
49 nt
Aptamer
L1: ACGA
P1
L2: ACGG
P2
5'
3'
J1/2: 13nt
76 nt
Aptamer
J1/2: 13nt
L2: 6 nt
P2
P1
L3: 9 nt
P3
5'
3'
J1/3: 9nt
80 nt
Aptamer
L5: 6nt
P5
P4
P1
3'
5'
P2
P3
L2: UGAGA
L3: UGCGU
84 nt
Aptamer
L2: GCGA
L3: UCU
P2
P3
J1/2: 6nt
P1
5'
3'
94 nt
Aptamer
P5
P4
P1
3'
P2
5'
P3b
P3a
L3b: CUG
L2: UGAGA
L3a: AUGA
98 nt
Aptamer
L3: AUAC
L2: GAAA
P3
P2
J1/2: 6nt
P1
5'
3'
Aptamers
c
39 nt
Terminator
5'
3'
60 nt
TIR
5'
P4
3'
P5
Expression platforms
d
14 nt
Hairpin I
L1: CUUG
5'
3'
14 nt
Hairpin II
L1: GAAG
5'
3'
54 nt
Ribozyme I
L2: UUCG
P2
L1: GUGA
3'
P1
5'
70 nt
Ribozyme II
3'
P2
P3
P1
5'
L3: 7 nt
L4: GCUU
P4
77 nt
tRNA
3'
5'
P1
L2: 9 nt
P2
P4
L4: 7 nt
P3
L3: 7 nt
Other RNA

Figure 16 Overview of RNA targets. Schematic secondary structures of the RNA targets investigated by 19F-FBS. Stems (P), loops (L), and junctions (J) are annotated, respectively. Tri-, tetra-, and pentaloop sequences are listed explicitly. Figure as published in Binas *et al.* (2)**.**

3.1.2 ^{19}F-CPMG-based screening for FBS

Initial fragment hits may be rapidly and precisely identified using NMR spectroscopy, which also allows the detection of weak binding in solution. Numerous NMR experiments are available to identify binding, including NOEs, chemical shift perturbations, saturation transfer differences, WaterLOGSY studies, and T2-relaxation spectroscopy (97, 228, 229). Depending on the experiment, these techniques can detect interactions with dissociation constants from 10 mM to low nM. The lower detection limit is now being further decreased using more advanced techniques of dynamic nuclear polarization (DNP) or hyperpolarization (230). NMR allows for the simultaneous observation of the target's interaction with multiple fragments, drastically reducing operational costs and time. Commonly, ^{1}H-detection is used in NMR screens where the number of fragments within a mixture that may be screened in a single experiment is limited by the substantial overlap of NMR signals. The large number of signals is caused by the high number of hydrogen atoms in fragment compounds and the additional peaks caused by couplings. ^{19}F-detection has significant advantages over ^{1}H detection (231). Compared to protons with a chemical shift dispersion of around 9 kHz (15 ppm), ^{19}F-NMR signals have a substantially greater chemical shift dispersion of about 50 kHz (83 ppm). Furthermore, each ^{19}F resonates at a single resonance frequency, assuming ^{1}H-decoupling is applied, allowing the detection of several fragments in a single mixture. ^{1}H-decoupling comes with additional requirements for the spectrometer, NMR probe head, and user expertise.

Figure 17 depicts the composition of fragment mixtures with 20 or 21 distinct fragments. A list of all 101 screened fragments is provided in SI Table 4.^{19}F transverse relaxation experiments were measured using CPMG pulse trains for various relaxation delays (229, 232). CPMG T_2 measurements detect the differing relaxation characteristics of fragments that are not bound compared to those transiently bound to the screened targets. While bound to a high-molecular-weight macromolecule (4 kDa-100 kDa), low-molecular-weight fragments will show significant relaxation/line broadening due to a decrease in rotational correlation times (τ_C) and a consequently faster T_2-relaxation. Even when the population of interacting fragments is as little as 1%, this line broadening is detected. This sensitivity is advantageous to other fragment-based strategies because it allows the detection of low-affinity interaction.

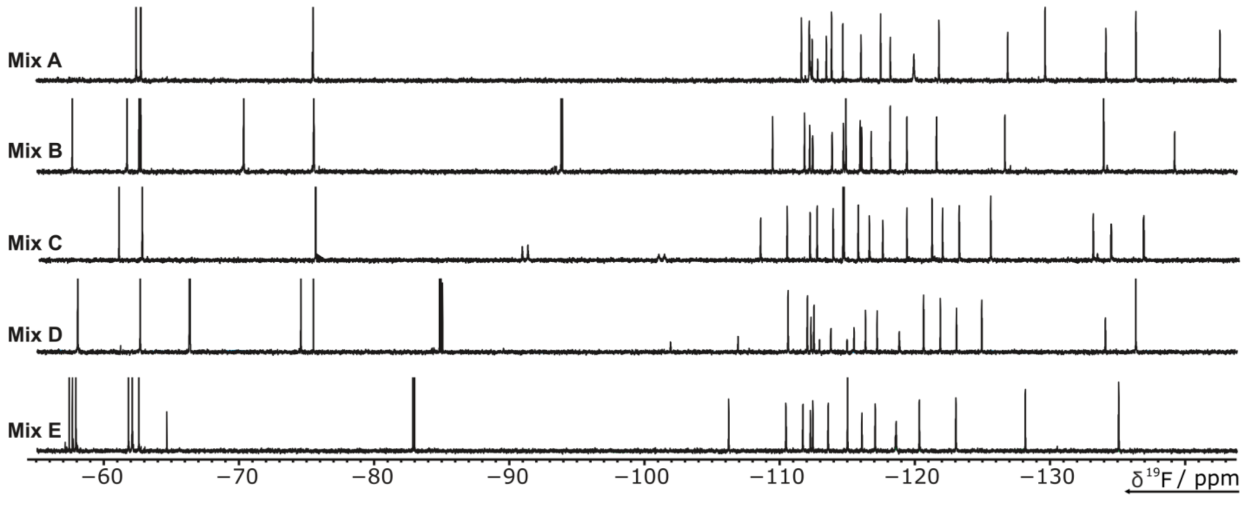


Figure 17 ^{19}F-1D NMR-spectra of the ^{19}F-library fragment mixtures. The ^{19}F-library contains 101 compounds (SI Table 4). Five mixtures of either 20 or 21 ligands were generated to avoid signal overlap. The spectra of the mixtures (A, B, C, D, E) in the screening buffer are displayed. Figure as published in Binas et al. (2).

3.1.3 Diversity of RNA structural elements – Targeting RNA

Target selection is a crucial phase in any screening process. Still, it's especially significant in multi-target approaches where finding a diverse range of physiologically relevant targets is essential to success. Showing not just the occurrence of hits but also selectivity between structure elements and classes of biomolecules is necessary to prove the feasibility of this fragment screening approach. Figure 16 and SI Table 3 list all RNA constructs that were examined. These structures represent major classes of structural motifs found for RNAs.

RNA structure elements – hairpins, bulges, internal loops, and pseudoknots

The small secondary structural motifs in RNA with the highest frequency are stem-loop/hairpin structures (233). Most loops are three to seven nucleotides long. However, tetraloops comprise more than half of all loops (234). Tetraloops are highly common and have high thermodynamic stability since interactions with hydrogen bonds and stacking frequently stabilize them. Further characterizations of the binding mode are required to distinguish between stem-binding (235) and loop-binding(223) ligands. The binding mode to physiologically relevant stem loops holds information on the structural basis of the ligand-induced changes, enabling the underlying biological function (236).

Rational drug design approaches created compounds that selectively bind to an RNA loop region (237). Two stem-loop structures represented this common secondary structural motif as targets of the screening. These structures contain 14 nt (nucleotides) and exhibit a GAAG- and a CUUG-tetraloop, respectively.

Helix-junction-helix (HJH) structure elements consist of two helices and an interconnecting junction. They can be divided into bulges or are named internal loops. While bulges contain only short single-stranded intersections on one side of an RNA structure element, loops contain opposing unpaired regions on both sides of the stem (238, 239).

The relative orientation of helixes to other helixes is enabled through bulges and internal loops. Variations in loops or bulges allow a multitude of conformations based on their size and shape. Beyond enabling a multitude of RNA structure arrangements, bulges and internal loops also govern conformational dynamics by acting as conformational hinges. Interhelix movements like dynamic nucleobase stacking or rearrangement of a junction are made possible by forming these structures (238). Low molecular weight homologs can be designed to target them, as demonstrated by the Tat-TAR interaction (240, 241). In this example, an arginamide ligand was imitated. Internal loops and bulges make great targets for drug screening due to their structural diversity, variability in topology, which provides binding pockets, and availability of possible hydrogen bridge interaction. This structural diversity increases the likelihood of specific ligand recognition and high-affinity binding. Especially compared to the stem structures of the helixes that are usually limited to pi-stacking interaction and are structurally identical in all stems with the same sequence. Virtual studies on the HIV-1 TAR revealed that sampling the full topological space led to an array of distinct conformations that could be independently targeted (236, 242, 243). The target pool contained several examples of small and large bulge regions, internal loops, and pseudoknots (Figure 16). Further aspects regarding RNA secondary and tertiary structure are outlined in chapter 1.1.2 1.1.1 .

Riboswitches

Riboswitches are RNA elements regulating gene expression through allosteric rearrangements of an expression platform element influenced by a small molecule binding to an RNA aptamer transcribed upstream of the expression platform. Most riboswitches respond to variations in the concentration of small molecules, commonly metabolites, which they bind with exceptional selectivity (244, 245). Figure 11 shows experiments for the detection of binding using homonuclear and heteronuclear 2D-NMR spectroscopy. Most aptamers create complex tertiary structures to accomplish the high-affinity binding necessary for optimum sensitivity, combined with adequate discrimination against non-cognate ligands. These binding pockets have a chemical space that is precisely defined and is often well-understood due to the availability of solved structures (SI Table 3). The diverse and specific chemical environments of riboswitch binding pockets make synthesizing derivates of the cognate ligand a feasible approach for finding novel ligands. This process can be aided through fragment-based drug discovery methods, which readily yield epitopes that can bind to parts of the binding pocket or novel interaction sites and allow drastic improvements in affinity or specificity when linked to the cognate ligand scaffold. Examples of new epitopes found for riboswitch aptamer domains were previously described (246, 247). Of the 14 RNAs examined, eight are generated from riboswitches with aptamer domains that naturally bind ligands.

The aptamer domains of riboswitches from the classes that sense second messengers, guanidinium, purines, and thiamin pyrophosphate (TPP) were screened (SI Table 3). The most prevalent riboswitch found in many prokaryotes and eukaryotes is the TPP-sensing riboswitch. The Schwalbe group extensively reported on the mechanism of full-length riboswitch function for purine-sensing riboswitches and their mechanisms acting on the transcriptional or translational level (24, 54). The guanidine-sensing riboswitch was added because it is an example of a riboswitch with hairpin structures. For purine-sensing riboswitches that are a member of the RNA targets, loop-loop interactions are a well-characterized element with a stabilizing function (248).

Other highly structured RNAs

To increase conformational space coverage and to access possible non-specific binding, we added five more RNAs to the pool of target RNAs, ranging in length from 14 nt to 77 nt, containing a diverse spectrum of RNA structure elements but lacking sites for cognate ligand binding. tRNA$_{\text{fMet}}$ was also used in this screening due to its previous use in high-throughput screens of RNA molecules. Using a widely available reference target also allows for increased comparability of the findings and easier reproducibility by other researchers. The RNAs further included an RNA G-quadruplex (139) and the transactivation response element (TAR) RNA (249).

Non-RNA targets – counter screen with other biomolecule classes

Further, five proteins with molecular weights ranging from 18 to 100 kDa and five DNAs were tested in the screening (2). Four of them were G-quadruplexes. The DNAs can rule out binders not specific to RNA by binding to them. G-quadruplexes are non-helical structure elements and have a high degree of accessible and, as a result, targetable residues.

3.2 Results

3.2.1 ^{19}F-CPMG screening of RNA and other macromolecules

Various fluorine-containing ligands exhibit binding to nearly all RNAs. With a hit rate of up to 26%, screening the fragment library against RNA targets produced several hits. These hits are expected to bind to their target RNA with a low millimolar K_D. These estimates are based on the finding that most fragments did not exhibit significant chemical changes in subsequent 1H-15N-correlation experiments. Low affinity is assumed, while it is unclear if changes in chemical shifts are strictly correlated with changes in the chemical environment.

Following the initial ^{19}F-screening, all hits binding to riboswitch aptamers were verified, a K_D value was determined as a proof of principle, and further insights into features of the fragment RNA interaction were gained. Samples containing a single fragment that showed binding interaction in initial screening mixtures were prepared to confirm hits. We observed the same altered T2-relaxation behavior for all hits, confirming their RNA binding capability and the feasibility of the pooled mixture screening (SI Figure 6 and (2)).

Mapping binding to a specific site in RNAs usually relies on chemical shift perturbations (CSPs) caused by fragment binding impacting RNA resonances of nucleobases near the binding epitope. The experimental setup must allow for the distinction of direct binding-induced CSPs from external factors from sample effects. Most of the RNAs in this screening were riboswitches, and they inherently bind to their specific low-molecular-weight metabolites with an affinity several orders of magnitude greater than the affinity anticipated for the fluorinated fragments. Fragments that are orthosterically bound can be identified in competition experiments. When added to an RNA sample, the native ligand will compete for the RNA binding site and drive out any lower-affinity binding fragment. This competition is experimentally observed by recovering signal intensity of fragment signals when the natural ligand is added to a CPMG experiment that showed T2-relaxation at an extended mixing time. Even though this signal recovery is a causal effect of fragment-ligand binding competition, this competition does not universally necessitate sharing the same binding site. There are examples where this effect is observed on allosteric binding sites. Here, a structural rearrangement caused by cognate ligand binding obstructs alternative binding sites previously accessible to fragments.

The fragment signal integrals were determined as part of the screening analyses, and the ratios between 200 ms CPMG and 0 ms CPMG measurements were calculated. The quotient Q^{bind} of the intensity ratios

$$Q^{bind} = \frac{Intensity\ Ratio^{\ +Target}}{Intensity\ Ratio^{\ -Target}}$$

with intensity ratio defined as

$$Intensity\ Ratio = \frac{Peak\ Integral^{\ CPMG(200\ ms)}}{Peak\ Integral^{\ CPMG(0\ ms)}}$$

is used to categorize the ligand-target interaction into no binding, weak, or strong binding (Figure 18). These three binding categories are a purely qualitative assessment of binding. The underlying increases in T2-relaxation observed in CPMG measurement are dependent on the overall rotational tumbling time τc of the target molecule and are dependent on its molecular weight, size, and shape. The analyses also did not account for differences between aromatic and aliphatic fluorine substituents. The quantitative assessment of all these effects is, while possible, not necessary within the scope of the initial fragment screening.

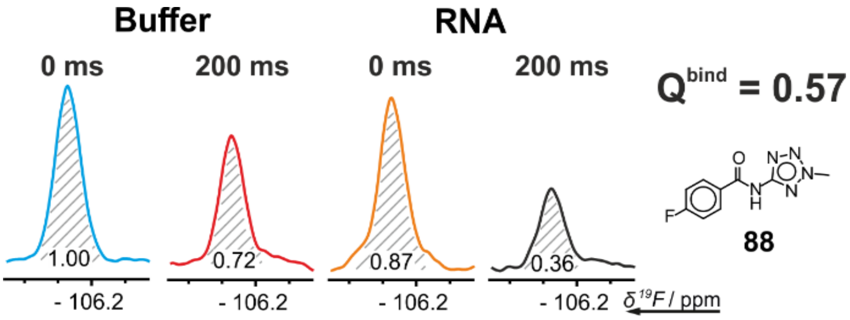


Figure 18 Determination of Q^{bind}. Four ^{19}F-CPMG experiments are recorded to determine the binding factor Q^{bind} from peak integrals, as discussed in the main text. The relaxation loss at 200 ms relaxation dephasing time relative to 0 ms dephasing for the ^{19}F signal of the ligand is recorded in the presence and absence of a biomolecular target. Figure as published in Binas et al. (2).

Figure 19 summarizes the complete screening into binding categories. SI Figure 6 shows the spectral regions containing the raw data of all hits. All targets tested in the screening had hits. Their respective biological hosts and available structural data are listed in SI Table 3. The screening yielded high hit rates for riboswitch RNAs, ranging from 7 to 26 hits per riboswitch. Aptamer domains seen in riboswitches bind molecules in the same size range as the fragments. For all other RNAs, only up to five hits were identified. Even though the CPMG measurements as relaxation-based experiments are target size dependent, the differences in hit rate between riboswitch targets compared to the slightly smaller 77 nt tRNA are striking.

Duplex and G-quadruplex, two DNA forms used to survey other nucleic acid structures, had a broad range of binding behaviors. While only one fragment demonstrated binding for the duplex, between 12 and 20 hits were found for G-quadruplexes, with some overlap to hits that also bind to riboswitch RNAs. Four of the five proteins under investigation displayed a great number of hits, ranging from 16 to 55.

Only four hits were found for the 18 kDa phosphatase MptpA, which is consistent with phosphatases' challenging drugability. On average, 5% of the data for 101 fragments screened across 24 different biomolecular targets containing DNA, RNA, or proteins could not be analyzed. This results from the requirement to adjust buffer conditions, particularly for proteins. For this subset of ligands, the various buffer conditions can cause problems with solubility and chemical stability. The remaining fragments exhibit a wide range of target selectivity, ranging from fragments (fragment 100) binding only one target to highly promiscuous binding behavior (fragment 57).

In addition to this large initial screening, hit lead fragments were followed up with cross-validation of binding to additional targets along with cheminformatic-based searches for comparable ligands that are commercially accessible.

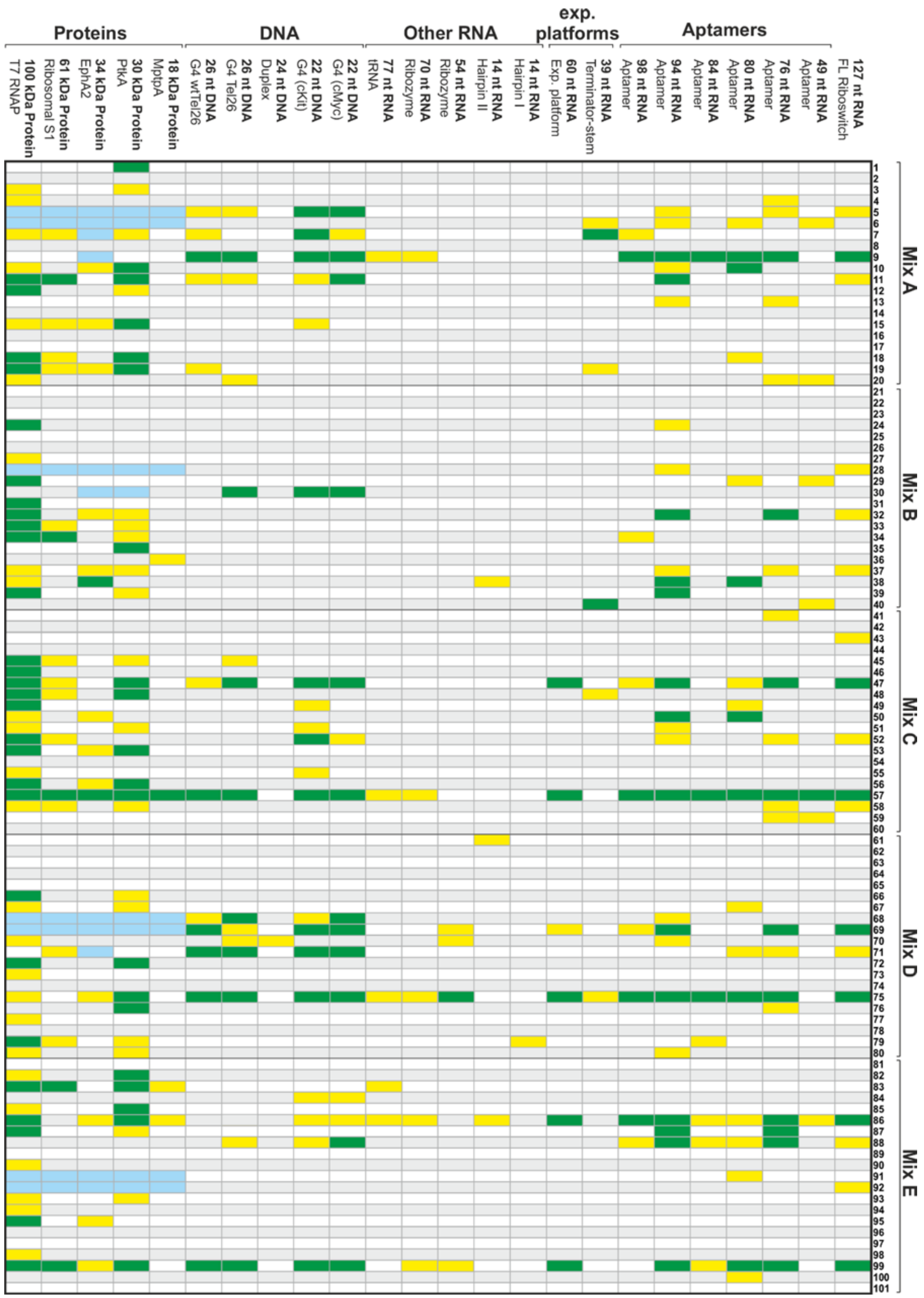


Figure 19 Interaction table of all fragments and biological targets screened. Hits were classified into no binding (Qbind > 0.67, alternating gray and white), weak (Qbind=0.66-0.33, yellow), or strong binding (Qbind < 0.32, green) in 19F-CPMG experiments. For protein screens, hits for ~5% of the ligands could not unambiguously be assigned (light blue). Figure as published in Binas et al. (2).

Hit conformation for riboswitch targets

Aptamer domains of the secondary-messenger-sensing riboswitches containing 76 nt, 84 nt, and 98 nt were selected for further study from the pool of biological targets. Fragments with weak hits were omitted. Individual fragments with strong hits were retested to confirm binding and rule out possible effects of fragment mixtures. The same strong binding was observed in the mixes for each studied fragment and RNA combination. A detailed analysis of fragment 75 addition to the 76 nt riboswitch, to confirm the effects seen and identify the binding epitope, was conducted by Oliver Binas (Figure 20a,b, and c). 15N-correlated 2D spectroscopy was performed on imino hydrogens using 15N-isotopically labeled RNA (Figure 20a) to look for any potential chemical shift disturbance of RNA signals introduced by adding the fragment. Only helical imino hydrogen signals, which notably shift in the case of helix groove binding fragments, are seen in these spectra. Tiny changes in the signals and occasionally additional, small signals were seen. Aromatic hydrogen signals in 1H,1H-TOCSY spectra exhibited substantial changes (Figure 20b). On the H5-H6 cross peaks of the pyrimidine residues, distinct signal changes above 10 Hz were measured. Only the strongest H5-H6 peaks are detectable at concentrations suitable for screening. Three signals in the compound 75 example (Figure 20b) showed dose-dependent chemical shift perturbation. Paired with observed minor alteration 15N-correlation data, we can conclude that the fragment binding site is positioned in a flexible region of the RNA. A sample of fragment 75 and the 76 nt secondary-messenger-sensing riboswitch displayed a 400 µM K_D in an NMR-titration experiment that measured the 1H-1D-chemical shift perturbations (Figure 20c) (250).

Competitive binding experiments provided more details on the binding site and mode. The impact of the native ligand competitive addition on the T2-modulated signal was assessed. The sample contained the investigated fragment and the RNA target. As in the screening of fragment mixtures, adding the RNA entirely suppresses the binding fragments signal. Upon introduction of the native ligand, a 15% recovery was observed. The partial recovery of fragment signals suggests that secondary non-specific binding occurs in addition to the fragment hit orthosterically binding to the corresponding ligand's binding site. A smaller population binds orthosterically to the same binding site as the cognate ligand, but the majority binds allosterically.

After adding the native ligand, a fragment signal recovery of up to 83% could be observed (Figure 20d). The higher the amount of recovered signal, the higher the displacement of fragments due to native binding. Since fluorine’s most abundant isotope is NMR active and can be processed to show as a single peak signal, observing 19F fragment signals is an efficient technique to collect affinity data by NMR.

Since 19F signals, in general, are sensitive to changes in their chemical environment, the 19F will likely exhibit a significant CSP following ligand addition. In comparison, most aromatic hydrogens have a smaller chemical shift dispersion, and the highest CSPs in 1H RNA signals are often just

5-8 Hz. Affinity constants in the high µM range, such as 400 µM for fragment 75, were obtained through ^{19}F CSP data analyses (251).

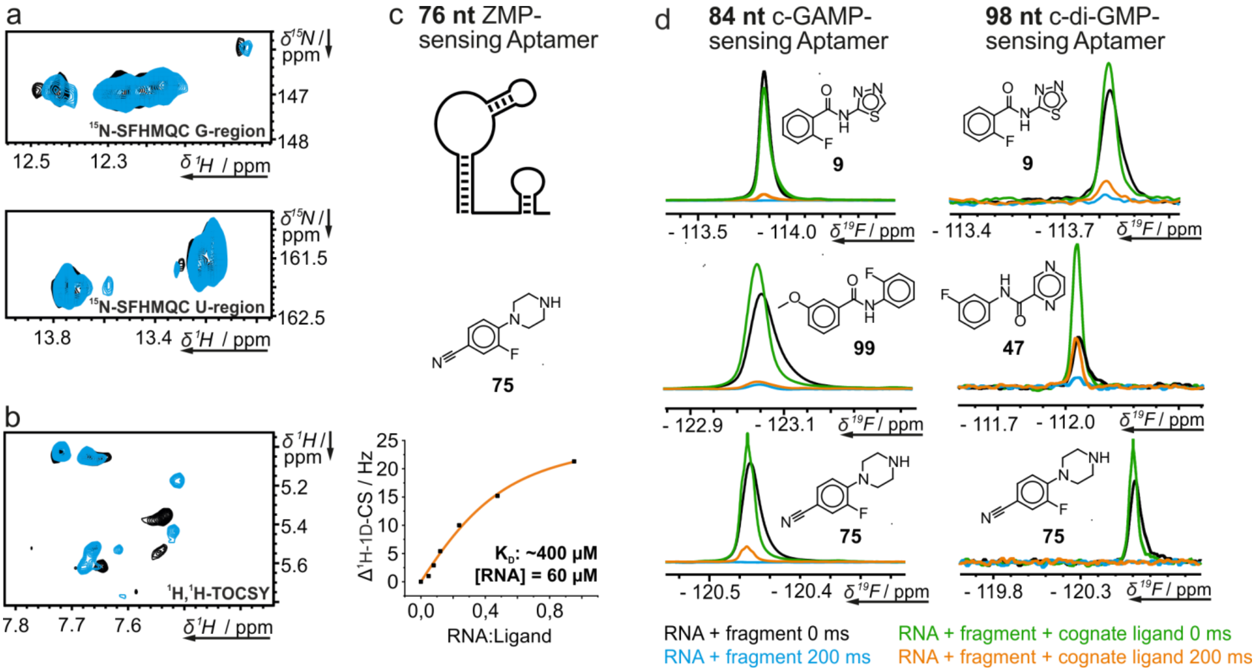


Figure 20 Hit validation and competition experiments: Validation of ^{19}F-CPMG screening hits for the aptamer domains of the three secondary messenger-sensing riboswitches. a) Spectral regions with signals from guanosine (top) and uridine (bottom) residues of the ^{1}H, ^{15}N-correlation experiment of the 76 nt riboswitch with (blue) and without 75 (black) shown under c. b) ^{1}H,^{1}H-TOCSY spectrum with (blue) and without 75 (black). c) ^{1}H-1D-NMR-titration of 75 with the RNA. K_D was determined, according to Williamson (250). d) (Partially) competitive binding of fragments to the 84 nt and 98 nt riboswitch observed in T_2-modulated 1D-^{19}F experiments. Figure as published in Binas et al. (2).

3.2.2 Cheminformatic analysis of hit data

Numerous biomolecular targets were screened, and the results show that the fragment library has the highest hit rate for proteins, followed by RNA and DNA. RNAs with loop regions, bulges, and internal loops are likelier to yield hits. The library contains some fragments that promiscuously bind to all three biomolecular target classes. There was around 20% overlap between hits that bound to proteins and RNA. Each class of biomolecular targets was selectively hit by specific targets, a very intriguing finding (Figure 21).

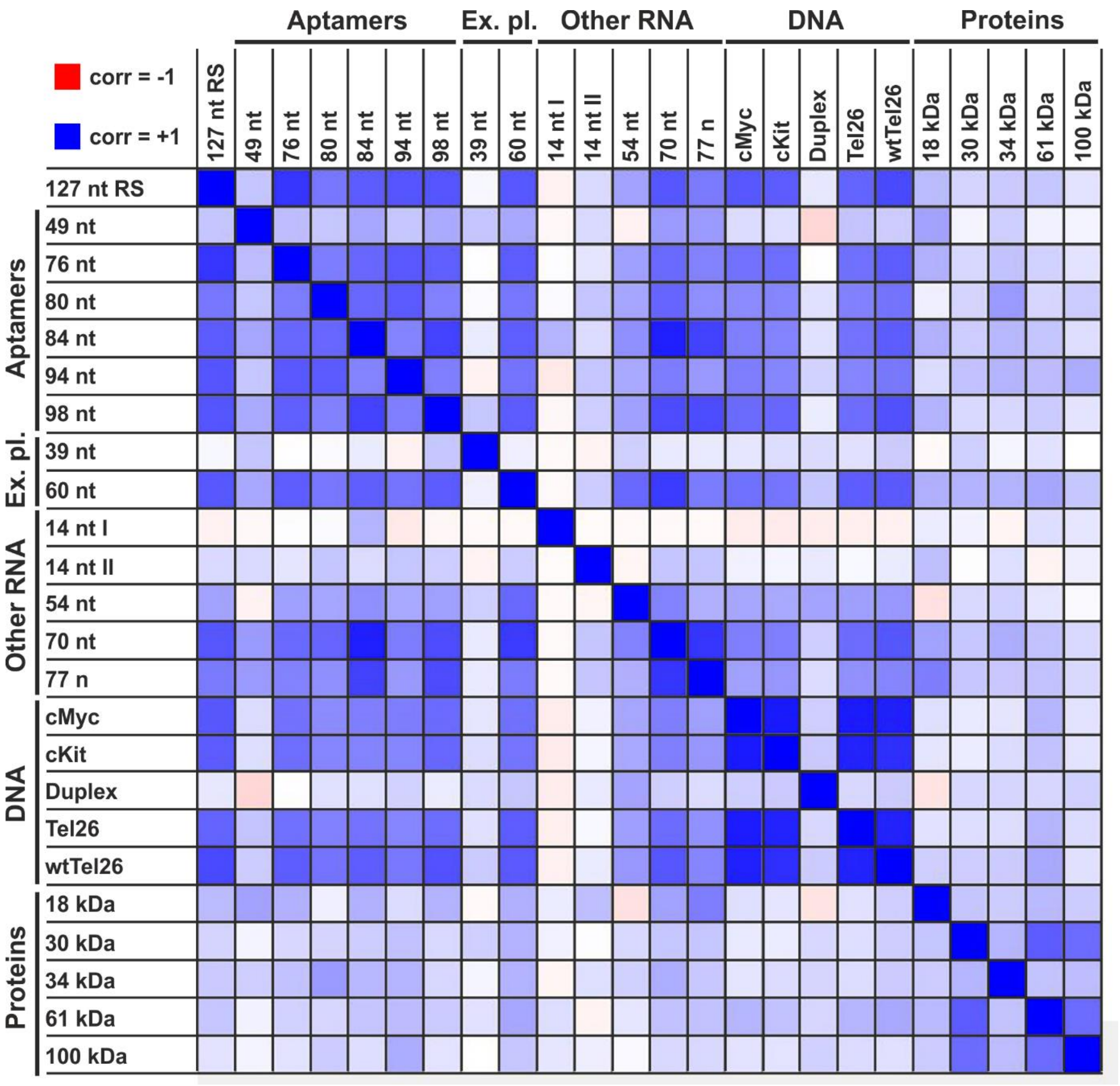


Figure 21 Correlation matrix of hit clusters, displaying hit correlation between different targets screened by ^{19}F-FBS. Figure as published in Binas et al. (2).

Using Hierarchical Clustering (DistMatrix, Morgan fingerprint, distance threshold 0.6, Knime software 4.0.2), Kamal Azzaoui and Marcel Blommers grouped the 69 fragment library hits into 38 singletons and 4 chemical families of a similar scaffold. Five members of the biggest cluster were binders for targets of DNA/RNA, proteins, and DNA/RNA/proteins. None of the target families evaluated seem to be associated with any particular cluster. Molecular descriptors related to shape, electrostatic, and hydrophobic interactions were developed to investigate correlations between chemical structures and the number of targets that bind to them. Based on the correlation matrix assessment, there were no significant correlations between the number of target hits and molecular descriptors. The highest correlation was found for the number of aromatic atoms (R = +0.27). In contrast, the lowest correlation was found for the SP3 descriptor, the ratio of SP3 carbon atom count over total carbon atom count (R

= -0.23). The statistical analysis of the number of aromatic atoms for each category of binders shows higher mean values of this descriptor, yet the standard deviation remained too high to differentiate between the categories. While increasing the library size could reduce the errors and result in values with highly significant differences, it is also possible that the low degree of underlying correlation represents the ground truth, and there is no preferred number of aromatic atoms for any of the targets in this study. There was no significant difference between the category for hits according to the SP3 descriptor and other molecular descriptors (Figure 22).

Substructure counting of popular and frequent motifs of organic compounds, as shown in Figure 22a, did not reveal any significant enrichment in various categories of binders. Due to the limited size of the fragments in the ^{19}F-library, there was no privileged class of compounds and no relevant physicochemical properties that showed specificity to a family of biological targets. The ^{19}F-fragment library has a composition suitable to produce starting points for further screening of RNA, DNA, and protein targets. Additionally, a remarkable clustering of hits between riboswitches and aptamers, DNA, and proteins, respectively, was observed in correlation analysis (Figure 21 and Figure 22c).

a Gaussion distribution for aromatic atoms count per class value

	DNA	RNA	Proteins	DNA/RNA	DNA/Proteins	RNA/Proteins	DNA/RNA/ Proteins
Count	2	9	23	5	4	18	13
Mean	5.5	7.2	6.9	9.2	8.2	6.9	8.3
Standard Deviation	0.7	2.4	2.6	5.2	4.5	2.2	3.9
Rate	3%	12%	31%	7%	5%	24%	18%

Gaussion distribution for SP3 descriptor per class value

	DNA	RNA	Proteins	DNA/RNA	DNA/Proteins	RNA/Proteins	DNA/RNA/ Proteins
Count	3	6	23	5	4	18	13
Mean	0.16	0.13	0.11	0.07	0.12	0.11	0.11
Standard Deviation	0.04	0.05	0.06	0.11	0.06	0.05	0.08
Rate	3%	12%	31%	12%	5%	24%	18%

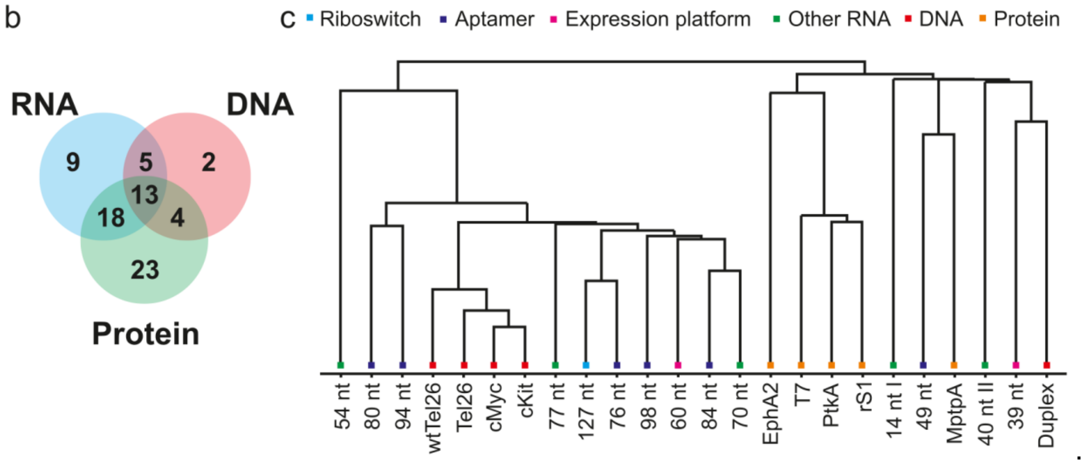


Figure 22 Cheminformatic analysis of hit data for all RNA, DNA, and Protein biomolecules. a) Gaussian distributions for aromatic atoms and SP3 descriptor over categories of biomolecules. SP3 descriptor (sp3 carbon atom count/total carbon atom count) reflects the flatness of the fragment molecules. b) Visualization of classes in a Venn diagram. c) Euclidian distribution of hits to the target biomolecules. (modified (added a percent sign) Figure as published in Binas et al. (2).

3.2.3 Follow-up chemistry

As a proof of concept and a follow-up to the fragment-based screening, binding fragments were combined to increase binding affinity. Linking a binding fragment to a known RNA binder can result in such improvements, as demonstrated for a Neomycin-Acridine compound(252).

The aim was to assess the feasibility of targeting the terminator and antiterminator structural elements found in riboswitches. The formation of the terminator and antiterminator stem are required steps of riboswitch-based gene regulation. The terminator and antiterminator stem, with sizes between 38 nt and 51 nt, were targeted for this prospective study since they represent the smallest RNA structures in the screening containing a drugable bulge motif. They are also a commonly found structure in transcription-regulating mechanisms.

The research targeted a benzamide (P2D11) for further modification, a near homolog of binding fragment 48 of ^{19}F-screening. P2D11 was chosen due to better chemical accessibility, commercial availability, and affordability of precursors. P2D11 was coupled to an acridine moiety, a known intercalator facilitating quick follow-up chemistry(253, 254). A three-step synthesis generates an acridine moiety and links it to the fragment. The intrinsic fluorescence of acridine made fluorescence-based binding experiments possible. Fluorescence titrations were conducted using the coupled derivative and the RNA targets (Figure 23).

The K_D interaction of acridine with the SAM-Antiterminator was 59 µM. The linked fragment acridine-P2D11 showed a 54-fold stronger affinity of 1.1 µM. Improved affinity was also observed for the A-terminator, which improved its affinity 3-fold, from an acridine only 4 µM K_D to acridine-P2D11 K_D of 1.5 µM. A decreased affinity was observed for the 2′dG-Terminator where acridine binds with 4 µM affinity and the acridine-P2D11 linked derivate binds with an affinity of 15 µM.

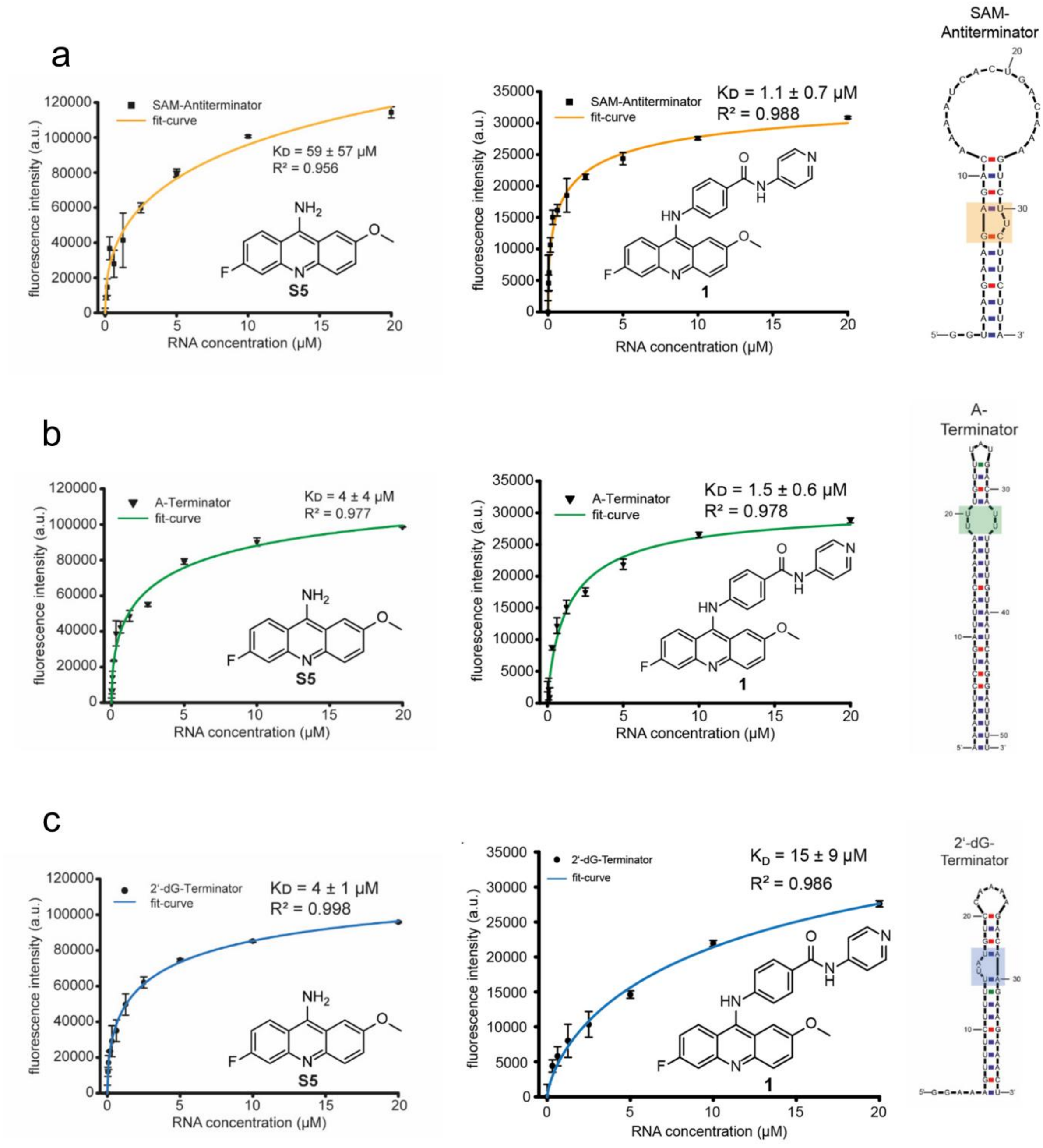


Figure 23 Fluorescence-based determination of affinity of acridine and coupled acridine-benzamide (P2D11) derivative to a) SAM-Antiterminator, b) A-terminator, and c) 2′dG-Terminator. Figure modified from Binas *et al.* (2).

3.3 Discussion

In summary, the screening of ^{19}F-containing libraries to 14 different RNA targets was successfully conducted. Commercially available fragment libraries were used to identify binding low molecular weight fragments. A general versatility of the used poised library could be shown and allowed cheap, fast, and simple follow-up chemistry to increase binding affinity to as low as 1.1 µM.

3.3.1 Feasabitly ^{19}F-CPMG screening of RNA and other macromolecules

The fragment-based ^{19}F-CPMG screening proved the applicability of the iNEXT library to our chosen RNA target space. Every target interacted with the fragment library. The number of hits correlated with the RNAs' size and structural diversity. It is expected for an FBS library to yield more hits for targets with a higher structural diversity since the fragments probe the possible interaction space, and higher structural diversity provides more different possible interactions. Striking was how much more hits were observed for aptamers than similar-sized tRNA. The larger interaction space indicated by more hits could be caused by an increased overall surface increase by the inside of an empty binding pocket or increased structural flexibility due to an empty binding pocket. Future studies could include ligand-stabilized aptamers to investigate this effect and to bias hit results toward allosteric binding. Based on the absence of CSPs and from the titration of one hit fragment, we know that the affinity of binders is in the low millimolar range and 0.4 mM in one case. From the vantage point of classical high throughput drug discovery, this would resemble a complete failure. But not so in FBS, fragments represent small and prototypic interaction surfaces, the maximum affinity strength is limited and millimolar affinities represent high-quality contacts devoid of flaws like steric hindrance or charge repulsion. Affinity increases are achieved down the line through fragment growth or linking.

Hit patterns were distinctly different from DNA or Protein interaction patterns, and mixture hits could be confirmed in single compound control experiments. The high reproducibility of hits confirmed the feasibility of the 25-fragment mixture approach and the high quality of the obtained hit data devoid of false positives. The low false positivity rate is partly caused by using intensity ratios, which correct for ligand-only effects. The analysis also benefited from the chosen 200 ms relaxation delay. In initial experiments, we observed that a 400 ms delay results in too much relaxation, increased artifacts, and strong hits, losing all signals, making their intensity ratio incomparable. The selected 200 ms delay was more appropriate for the affinity range of the fragments in this screening and exclusively used when the target pool was extended to its final size.

The classification into strong, weak, and non-binders allowed readily perceiving interaction affinity from Figure 9 and comparing different fragments binding to similar-sized aptamers. Due to rotational tumbling interactions, the small targets might cause less signal through relaxation, resulting in misclassification of strong as weak binders. Future analyses should integrate a rotation correlational time correction term if hit interaction strength comparisons over large target size ranges are required. The analysis of interaction strength is secondary to the quality of the discovered interaction hits. Hits should be clustered based on physical properties and the geometry of their likely interacting contact surfaces before selecting based on affinity. Novel interacting epitopes are novel, independent of their interaction strength.

About 5% of the fragment signals during protein screening could not be unambiguously assigned, likely due to signal shifts caused by different buffer conditions. This significant number of unassignable signals highlights the robustness of ^{19}F signals to buffer changes. It could get corrected in follow-up studies providing reference measurements in the identical buffer and would likely yield a completely assignable dataset. An advantage of RNA research is that the studied targets usually tolerate similar buffer conditions and do not participate out of solution allowing identical buffer conditions through a large target pool.

Beyond establishing the affinity range of hits and confirming hits in single compound experiments, our follow-up analysis could show that fragment 75 binds the ZMP aptamer in a flexible region of a helical groove (Figure 20a). As discussed in chapter 1.1.2 RNA helices have a deep minor groove that is hard to target and offers limited sequence-specific interaction, while the minor groove is more accessible. Fragment 75 is a prolific binder interacting with 11 RNAs in the screen and strongly interacting with 8 of them (Figure 19). While this degree of nonspecificity suggests sequence independence, it is not necessarily caused by minor groove binding. Instead, the interaction with H5-H6 signals suggests major groove interactions, and the fragment size might be too small to accommodate sequence specificity beyond three interacting residues. Further analysis would be required to resolve details about the binding mode unambiguously.

The competition experiments shown in figure Figure 20d indicate the fragments can interact with multiple binding sights on the same target, including the cognate ligand binding pocket. The interaction with the binding site shows the potential to disrupt the ligand-binding associated regulation mechanisms of riboswitches. At this time, the fragments of this screen lack specificity regarding the target and binding site, but this can be mitigated in the subsequent workflow of fragment-based drug discovery. Fragments that show a high degree of specificity are a good candidate for linking to other fragments. Less specific fragments can be incrementally extended or grown, increasing the possibility of constructive interactions. The CPMG measurements can easily be tuned to accommodate a correspondingly increased affinity range. This incremental approach allows a detailed understanding of which minute changes benefit affinity or selectivity and which targets penalize the induced changes. Fragments can be grown in a directed manner with a low likelihood of introducing negative interactions. Grown fragments can be added to the fragment pool, and subsequent computational analyses can refine the screening methodology. In a refined methodology, promiscuous fragments representing established interaction patterns can act as probes to establish which grown fragment pool should be applied to which target, reducing measurement time and cost.

3.3.2 Correlational analysis and limits of molecular descriptors

Cheminformatics allowed the delineation of features within the fragment pool specific to each biomolecular target class. The hits allow the assessment of the RNA targets' general drugability (141, 255). The overlap in interactions of promiscuous fragments allowed a detailed Euclidian distribution analysis of the targets based on correlations in binding patterns (Figure 22c). Already a 100-fragment library allows refining the epitope space towards specific RNA classes. Building libraries specifically suitable for RNA is a current field of study. The repeating structure of the RNA backbone, the higher charges, and RNA dynamics are core ideas considered in this library design (256, 257). Although the

^{19}F-fragment library demonstrated here was suitable for proteins and RNA, these recent advancements may enhance existing fragment libraries into a more RNA-focused library.

The computational analysis of molecular descriptors found no significant correlations between the number of target hits and molecular descriptors (Figure 22a). As stated, this might result from the data set size or the limited number of molecular descriptors chosen. Increasing the dataset and the number of descriptors analyzed bears the risk of finding significant trends by chance. Further refinement of the fragment pool towards specific RNA classes could yield more descriptive descriptors like molecular homologs named after chemical structure motifs. An example of this is the beta-lactam motif in beta-lactam antibiotics. A more targeted fragment pool would require fragments to share structural motifs and is not feasible in a small-scale initial screen, as conducted here.

3.3.3 Feasibility of drugs discovery follow up

As a proof of concept, Jason Martins demonstrated convenient follow-up chemistry, linking an RNA-binding fragment with the intercalator acridine to produce low micromolar RNA binders with greater than 15-fold selectivity for distinct RNAs. The fragment linking increased the affinity of acridine, a known binder, by 59-fold and showed a selectivity. Given the small sample size, this is a remarkable success and attests to the potential of this screening approach.

Future studies could link fragments where both parts showed a degree of specificity toward the target. Linker length and chemical makeup can influence the affinity of linked fragments and were not optimized in this initial test. Ideally, the linker design allows fragments to interact with their respective binding sites with minimal steric hindrance. Additional affinity could be gained through fragment growth. Growing a fragment before linking improves affinity, and more synthetically accessible homologs could be found through screening before linking to other fragments.

3.3.4 ^{19}F-CPMG NMR as a powerful tool for fragments-based drug discovery

The ^{19}F-NMR screening method presented in this work has several intrinsic advantages over competing methods and specific advantages when applied to RNA targets.

^{19}F-CPMG NMR is a ligand-observed method. Ligands are used at low and equimolar amounts compared to the macromolecule. Beyond providing cost savings, this prevents issues like low solubility or interference from unbound large excess fragments common in STD screening methods. The ^{19}F-CPMG experiments are optimized to the fragment mixture, and the only contributing factor of the macromolecule is its tumbling speed. This tumbling speed can be readily estimated, determined, and mathematically corrected in the downstream analysis. The method is inherently macromolecule agnostic. One macromolecule-free reference measurement in the shared buffer can be used for every target. The absence of target-specific measurements, like ON and OFF resonance pulses required in STD, halves the measurement time required when a large number of targets is investigated. Targets do not have to fall within the size limit constraints of NMR. No labeling of the macromolecule, common in competing methods, is required. The absence of labeling reduces preparative steps, cost, and room for error, be it from faulty preparation or introduced by interfering labels. If a target can be provided at a concentration that allows the detection of fragment signals, it can be screened.

The NMR data contain one signal per ^{19}F atom, and mixtures are designed to optimize the distribution of fragment signals throughout the detected spectrum under the specific buffer condition. Signal changes in mixtures can be directly attributed to fragments without deconvolution or further analytical workup. Signal intensity changes can be easily quantified and carry information on the strength of the underlying interaction. Multi-component measurements of 25 fragments were possible without false positives. Mixtures of fragments should be optimized for maximal measurement throughput. Distributing signals across the ^{19}F spectrum is an excellent example of this type of mixture design. The upper limit of the number of fragments that can be screened in one mixture was not tested and could lie a lot higher. Standard procedures from competing screening methods like orthogonal screening could increase efficiency. An ideal fragment pool would contain a preselected subset of fragments that have been shown to interact with similarly structured RNAs while still providing an epitope diversity that allows an unbiased screening of a substantial part of the chemical interaction space.

Beyond all this method's advantages and future potential, a core aspect of innovation is the general and broad adoption of a new approach. In this regard, we are delighted to report that this work convinced our coworkers to apply ^{19}F-CPMG to explore the drugbindingability of conserved RNA regulatory elements in the SARS-CoV-2 genome (136) and that our corresponding publication was recognized as highly downloaded by its journal.

Chapter 4 Concluding remarks

The first project discussed in Chapter 2 provides a detailed understanding of the cotranscriptional changes of the pilM riboswitch through single nucleotide extension. The analysis of transcriptional intermediates of pilM gave us a deep understanding of key transcriptional intermediates, P1 stem-associated affinity changes, and cotranscriptional changes to association rate constants. Our analysis of pilM aptamer showed that the truncated version of the pilM aptamer remained binding competent until the Watson-Crick-base pairing residue C75 was removed from the binding pocket either by truncation or by formation of the terminator. Therefore, the 25 nt long binding window ranges from the initial availability of C75 until cotranscriptional folding can refold it into competing structures. Constructs stayed binding competent while the initial bases of the terminator stem were included in the transcript and transitioned sharply to non-binding with the addition of nucleotide 102 (Figure 24B and D). The crystal structures and consensus sequences of GEMM I riboswitch suggest that a three-stem junction structure with a P1 stem is necessary for ligand binding. A stable P2 and P3 stem were observed for all constructs, while P1 stem formation required ligand addition. While a dynamic P1 has been described previously for other riboswitches, this is now also confirmed for CDN riboswitches. Ligand addition caused several changes in the imino spectrum that matched homolog crystal structures of the *holo* state. The matching binding competent windows were confirmed in ITC and NMR measurements. PilM affinities increase with P1 stem length from 19 µM to 0.25 µM, spanning two orders of magnitude. The k_{on} rates remained near constant during elongation. The differences in K_D originated from differences in k_{off} rate. Similar rates and behavior were determined for other riboswitches using stopped-flow fluorescence. These findings are now also observed in ITC line shape analysis we conducted. The K_D, k_{on}, and k_{off} rates of the highest affinity riboswitch aptamers were 0.25±0.04 µM, 17000±3000 $M^{-1}s^{-1}$, and 0.0042±0.0005 s^{-1} for pilM and 0.25±0.04 µM, 21000±300 $M^{-1}s^{-1}$, and 0.0045±0.0005 s^{-1} for Cd1, respectively.

The comparison with the Cd1 riboswitch allowed the development of a Markov model capable of reproducing ON and OFF switch regulation. Our modeling approach allowed us to couple the time-dependent conformational changes to a second variable and show both variables' impact in the contour plots. The observed dependence on ligand concentration, transcription speed, pausing, and base pair closing rate provided novel insights. We found that fast closing base pairs drive state transitions towards aptamer competing folds. The model reproduced the biological outputs in all high and low cognate ligand concentration conditions for ON and OFF switches. The terminator population of pilM decreased from over 95% to 20-25% at increasing CDN concentrations. The equivalent change in concentration increased the terminator population from under 5% to 90-95% for Cd1. The comparison highlighted that pilM *holo* and terminator formation compete at high ligand concentrations, indicated by their overlapping formation, while Cd1 cotranscriptionally saturates the *holo* population before the antiterminator can form. Our model contributes several advances in the simulation of cotranscriptional folding. It describes all three expected ground states of the riboswitch conformations, *apo, holo*, and the aptamer competing fold. It allows the observation of transitions from the initially populated *apo* state towards more stable states as a function of time and a second variable. The time-dependent influence of CDN concentration between 100 nM and 100 µM, with transcription speed between 1 and 100 nt/s, can be analyzed and understood (Figure 24A and E).

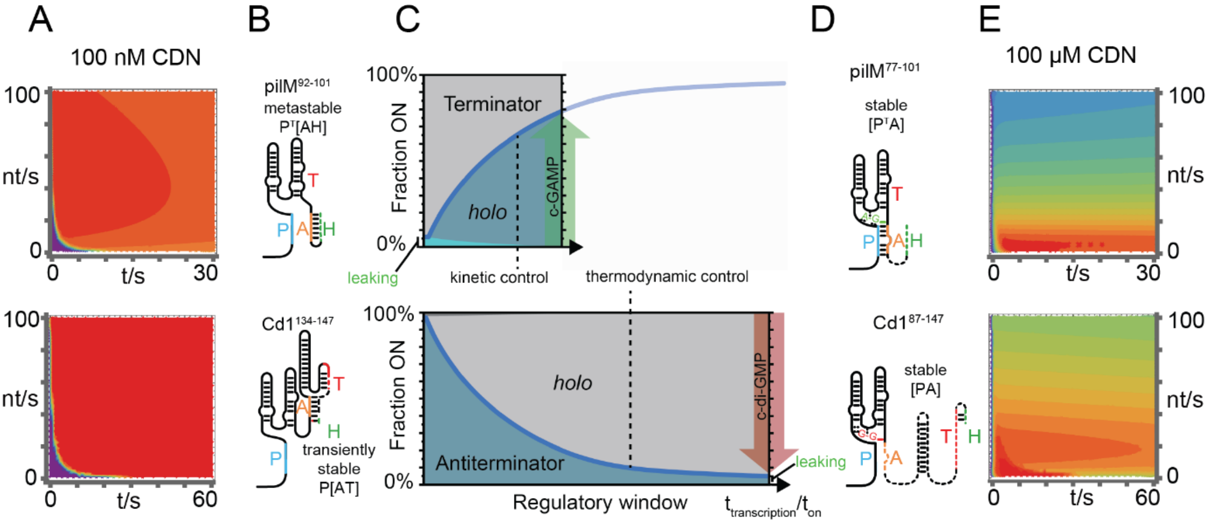


Figure 24 Regulatory understanding gained by the analysis of the regulatory window of CDN riboswitches: Contour plots of transcription speed dependence of the predominate state of pilM and Cd1 at 100 nM (A) and 100 µM CDN (E). Contour lines indicate 5% levels of relative population increasing from violet, blue, green, yellow, to red. At low CDN concentrations, pilM forms terminator and Cd1 forms antiterminator, and at high CDN concentrations, the *holo* state is formed, inverting the switch behavior during the regulatory window as shown in (C). PilM surpasses the limit for pure kinetic control, and Cd1 operates at near saturation with thermodynamic control character. Leaking/false positive signaling occurs only during OFF signaling and is indicated in green. The secondary structures of the predominate states are depicted in (B) and (D). Nucleotides that are transcribed during the regulatory window are marked with dashed lines. Figure modified from Landgraf et al. (1) and used in Figure 13, Figure 14, and Figure 15.

The observed differences between the riboswitches could be linked to optimized conditions for high dynamic range and minimal gene leakage. ON switches maximize regulatory responsiveness under kinetic control, and OFF switches minimize leakiness under thermodynamic-like control (Figure 24C). Our results show that if a Cd1 pause site would be confirmed, it would likely regulate at lower ligand concentrations than expected. We argue that riboswitch behavior at a concentration substantially higher than the K_D is well described through the ratio of ligand association time to transcription time. Based on their k_{off} rates, pilM and Cd1 operate under kinetic control. However, based on our analysis, Cd1 approaches saturation of its aptamer due to the longer transcription time and operates at near thermodynamic control. The approach of thermodynamic control caused a reduced transcription speed dependence of Cd1, which further decreased with pausing. PilM utilizes the linear range of the ligand response curve associated with kinetic control but does not require to highly populate the aptamer to minimize gene leaking. On the contrary, Cd1 cotranscriptionally reaches near full *holo* population, likely to minimize gene leaking.

The second project discussed in Chapter 3 optimized and deployed a ^{19}F fragment library in an NMR screen against multiple RNAs and macromolecular controls. Fourteen RNA targets were successfully screened and compared to five DNAs and five proteins to assess the general applicability of the screening and the targetability of riboswitches through FBS.

The method proved highly capable of finding interacting fragments. ^{19}F-NMR FBS approach has inherent advantages like direct detection of interaction, easy signal assignment, and signals carrying interaction strength information. Signal position optimized 20 fragment mixtures allowed unambiguous hit detection. The method had no false positives, based on the conducted hit conformation of a subset of strong binders. Competition experiments with cognate riboswitch ligands showed fragments interact with multiple binding sites on the same target, including the cognate ligand binding pocket (Figure 25C). The affinity of a fragment towards the ZMP aptamer was determined as submillimolar, 0.4 mM.

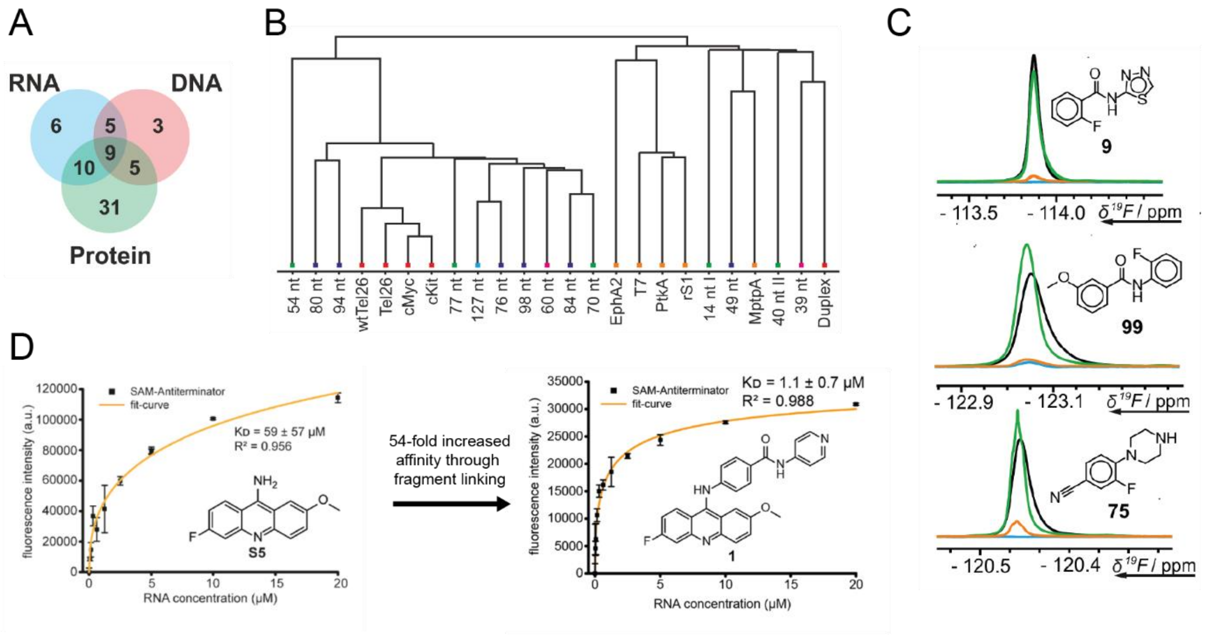


Figure 25 Computational analyses, competition experiments, and follow-up chemistry enabled by ^{19}F-NMR FBS**:** Cheminformatic analysis of hit data for all RNA, DNA, and Protein biomolecules. (A) Visualization of major classes in a Venn diagram. (B) Euclidian distribution of hits by target. (C) (Partially) competitive binding of fragments 9, 99, and 75 to the 84 nt riboswitch aptamer observed in T_2-modulated 1D-^{19}F experiments. Measurements with target and fragment are shown as black (0 ms) and blue (200 ms) lines and after the addition of c-GAMP with green (0 ms) and orange (200 ms) lines. (D) Fluorescence-based titration of acridine and coupled acridine-benzamide (P2D11) derivative to SAM-Antiterminator. Figure modified from Binas et al. (2) and used in Figure 20, Figure 22, and Figure 23

Computational analyses showed that macro molecules could be clustered based on hits, with RNA aptamers appearing as particularly targetable, highlighting the need for RNA-specific FBS libraries (Figure 25A and B). RNAs could be clustered into small, large, and aptamer RNAs. Small RNA yielded a limited amount of hits, appeared challenging to target, and exposed a limited chemical interaction space. Hit rates correspondingly increased with target size. Riboswitch aptamer RNAs were especially targetable compared to other RNAs of similar size. This increased targetability is likely a result of the larger diversity of structural interaction found in ligand binding pockets.

A follow-up experiment showed the feasibility of progressing a hit into a tighter binding lead compound. A 54-fold affinity increase to a K_D of 1.1 µM towards the SAM-Antiterminator was achieved (Figure 25D). Given the small scope of this first attempt, this is a very promising indicator of the potential of FBS.

4.1.1 Outlook on targeting kinetically controlled transcriptional riboswitches with FBS.

Considerations on kinetic riboswitch control and FBS on riboswitches lead to the question of how to target transcriptional riboswitches with FBS. A fragment can only influence cotranscriptional gene regulation if it interacts with the transcriptional riboswitch binding as fast or faster than the cognate ligand or interferes with the mechanism through allosteric binding at a similar timescale as the transcription speed of the riboswitch. The association time depends on the k_{on} rate and the concentration of the binder compound. A short transcription time and low *in vivo* concentrations of a possible future riboswitch-regulating compound might require extremely fast k_{on} rates.

Future screening efforts must be adjusted accordingly. Target should be filtered by the transcription time and the k_{on} rate of their native ligand. Our findings show that CDN nucleotide riboswitches represent challenging targets due to their fast association times. FBS approaches similarly would have to be adjusted to screen k_{on} rates rather than affinities. A possible approach to screen for k_{on} could be using mutated targets that cannot refold into stable aptamer complexes and are more representative of the initial encounter complex of ligand and target.

Regarding linking fragments to yield stronger binders, the cotranscriptional nature of transcriptional riboswitches suggests that the target space is not limited to the RNA but also includes the RNA polymerase, which will be present when regulatory decisions are made. The binding to RNA polymerase would occur under thermodynamic control and could be optimized using classical means. This way of targeting can be imagined as binding a riboswitch regulator to the RNA polymerase that will regulate the riboswitch when transcribed. The effective *in vivo* concentration would drastically increase through this localization.

These considerations would allow the pursuit of the challenging task of targeting transcriptional riboswitches in future research efforts.

References

1. Landgraf,T., Völklein,A.E., Fürtig,B. and Schwalbe,H. (2022) The cotranscriptional folding landscape for two cyclic di-nucleotide-sensing riboswitches with highly homologous aptamer domains acting either as ON- or OFF-switches. *Nucleic Acids Res.*, **50**, 6639–6655.

2. Binas,O., Jesus,V., Landgraf,T., Völklein,A.E., Martins,J., Hymon,D., Kaur Bains,J., Berg,H., Biedenbänder,T., Fürtig,B., *et al.* (2021) 19 F NMR-Based Fragment Screening for 14 Different Biologically Active RNAs and 10 DNA and Protein Counter-Screens. *ChemBioChem*, **22**, 423–433.

3. Eigen,M., Gardiner,W., Schuster,P. and Winkler-Oswatitsch,R. (1981) The Origin of Genetic Information. *Sci. Am.*, **244**, 88–118.

4. Gilbert,W. (1986) Origin of life: The RNA world. *Nature*, **319**, 618–618.

5. Bernhardt,H.S. (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biol. Direct*, **7**, 1–10.

6. Watson,J.D., Baker,T.A., Bell,S.P., Gann,A., Levine,M. and Losick,R. (2014) Molecular Biology of the Gene 7th ed. Pearson.

7. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with Small Molecules. *Chem. Rev.*, **108**, 1171–1224.

8. Brion,P. and Westhof,E. (1997) HIERARCHY AND DYNAMICS OF RNA FOLDING. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.

9. Tinoco,I. and Bustamante,C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.

10. Shelton,V.M., Sosnick,T.R. and Pan,T. (1999) Applicability of Urea in the Thermodynamic Analysis of Secondary and Tertiary RNA Folding. *Biochemistry*, **38**, 16831–16839.

11. Frugier,M. and Schimmel,P. (1997) Subtle atomic group discrimination in the RNA minor groove. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 11291–11294.

12. Zaman,G.J.R., Michiels,P.J.A. and Van Boeckel,C.A.A. (2003) Targeting RNA: New opportunities to address drugless targets. *Drug Discov. Today*, **8**, 297–306.

13. Wolters,J. (1992) The nature of preferred hairpin structures in 16S-like rRNA variable regions. *Nucleic Acids Res.*, **20**, 1843–1850.

14. Bottaro,S. and Lindorff-Larsen,K. (2017) Mapping the Universe of RNA Tetraloop Folds. *Biophys. J.*, **113**, 257–267.

15. Tuerk,C., Gauss,P., Thermes,C., Groebe,D.R., Gayle,M., Guild,N., Stormo,G., d'Aubenton-Carafa,Y., Uhlenbeck,O.C. and Tinoco,I. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 1364–1368.

16. Cheong,C., Varani,G. and Tinoco,I. (1990) Solution structure of an unusually stable RNA hairpin, 5GGAC(UUCG)GUCC. *Nature*, **346**, 680–682.

17. Staple,D.W. and Butcher,S.E. (2005) Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol.*, **3**, e213.

18. Bina-Stein,M., Crothers,D.M., Hilbers,C.W. and Shulman,R.G. (1976) Physical studies of denatured tRNA2Glu from Escherichia coli. *Proc. Natl. Acad. Sci.*, **73**, 2216–2220.

19. Nelson,J.A. and Uhlenbeck,O.C. (2006) When to Believe What You See. *Mol. Cell*, **23**, 447–450.

20. Mccown,P.J., Corbino,K.A., Stav,S., Sherlock,M.E. and Breaker,R.R. (2017) Riboswitch diversity and distribution. *Rna*, **23**, 995–1011.

21. Kavita,K. and Breaker,R.R. (2023) Discovering riboswitches: the past and the future. *Trends Biochem. Sci.*, **48**, 119–141.

22. Wu,M. and Tinoco,I. (1998) RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci.*, **95**, 11555–11560.

23. Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2014) Hierarchy of RNA Functional Dynamics. *Annu. Rev. Biochem.*, **83**, 441–466.

24. Reining,A., Nozinovic,S., Schlepckow,K., Buhr,F., Fürtig,B. and Schwalbe,H. (2013) Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature*, **499**, 355–359.

25. Voet,D. and Voet,J.G. (2011) Biochemistry Wiley.

26. Schafer,D.A., Gelles,J., Sheetz,M.P. and Landick,R. (1991) Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*, **352**, 444–448.

27. John,R., Davenport, Wuite,G.J.L., Landick,R. and Bustamante,C. (2000) Single-Molecule Study of Transcriptional Pausing and Arrest by E. coli RNA Polymerase. *Science (80-.).*, **287**, 2497–2500.

28. Vogel,U. and Jensen,K.F. (1994) The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J. Bacteriol.*, **176**, 2807–2813.

29. Adelman,K., La Porta,A., Santangelo,T.J., Lis,J.T., Roberts,J.W. and Wang,M.D. (2002) Single molecule analysis of rna polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 13538–13543.

30. Perdrizet,G.A., Artsimovitch,I., Furman,R., Sosnick,T.R. and Pan,T. (2012) Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl. Acad. Sci.*, **109**, 3323–3328.

31. Steinert,H., Sochor,F., Wacker,A., Buck,J., Helmling,C., Hiller,F., Keyhani,S., Noeske,J., Grimm,S., Rudolph,M.M., *et al.* (2017) Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *Elife*, **6**, 1–18.

32. Kramer,F.R. and Mills,D.R. (1981) Secondary structure formation during RNA synthesis. *Nucleic Acids Res.*, **9**, 5109–5124.

33. Lai,D., Proctor,J.R. and Meyer,I.M. (2013) On the importance of cotranscriptional RNA structure formation. *Rna*, **19**, 1461–1473.

34. Schärfen,L. and Neugebauer,K.M. (2021) Transcription Regulation Through Nascent RNA Folding. *J. Mol. Biol.*, **433**.

35. Ray-Soni,A., Bellecourt,M.J. and Landick,R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.*, **85**, 319–347.

36. Römling,U., Galperin,M.Y. and Gomelsky,M. (2013) Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol. Mol. Biol. Rev.*, **77**, 1–52.

37. Ross,P., Weinhouse,H., Aloni,Y., Michaeli,D., Weinberger-Ohana,P., Mayer,R., Braun,S., de Vroom,E., van der Marel,G.A., van Boom,J.H., *et al.* (1987) Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature*, **325**, 279–281.

38. Witte,G., Hartung,S., Büttner,K. and Hopfner,K.-P. (2008) Structural Biochemistry of a Bacterial Checkpoint Protein Reveals Diadenylate Cyclase Activity Regulated by DNA Recombination Intermediates. *Mol. Cell*, **30**, 167–178.

39. Davies,B.W., Bogard,R.W., Young,T.S. and Mekalanos,J.J. (2012) Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. *Cell*, **149**, 358–370.

40. Kellenberger,C.A., Wilson,S.C., Hickey,S.F., Gonzalez,T.L., Su,Y., Hallberg,Z.F., Brewer,T.F., Iavarone,A.T., Carlson,H.K., Hsieh,Y.-F.F., *et al.* (2015) GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 5383–5388.

41. Nelson,J.W., Sudarsan,N., Phillips,G.E., Stav,S., Lünse,C.E., McCown,P.J. and Breaker,R.R. (2015) Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc. Natl. Acad. Sci.*, **112**, 5389–5394.

42. Ren,A., Wang,X.C., Kellenberger,C.A., Rajashankar,K.R., Jones,R.A., Hammond,M.C. and Patel,D.J. (2015) Structural basis for molecular discrimination by a 3’,3’-cGAMP sensing riboswitch. *Cell Rep.*, **11**, 1–12.

43. Lotz,T.S. and Suess,B. (2018) Small-Molecule-Binding Riboswitches. **2002**.

44. Sudarsan,N., Lee,E.R., Weinberg,Z., Moy,R.H., Kim,J.N., Link,K.H. and Breaker,R.R. (2008) Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science (80-.).*, **321**, 411–413.

45. Winkler,W.C. and Breaker,R.R. (2003) Genetic control by metabolite-binding riboswitches. *ChemBioChem*, **4**, 1024–1032.

46. Mandal,M. and Breaker,R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451–463.

47. Tucker,B.J. and Breaker,R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin.*

Struct. Biol., **15**, 342–348.

48. Weinberg,Z., Barrick,J.E., Yao,Z., Roth,A., Kim,J.N., Gore,J., Wang,J.X., Lee,E.R., Block,K.F., Sudarsan,N., *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.*, **35**, 4809–4819.

49. Pursley,B.R., Maiden,M.M., Hsieh,M.L., Fernandez,N.L., Severin,G.B. and Waters,C.M. (2018) Cyclic di-GMP regulates TfoY in Vibrio cholerae to control motility by both transcriptional and posttranscriptional mechanisms. *J. Bacteriol.*, **200**, 1–19.

50. Tang,Q., Yin,K., Qian,H., Zhao,Y., Wang,W., Chou,S.H., Fu,Y. and He,J. (2016) Cyclic di-GMP contributes to adaption and virulence of Bacillus thuringiensis through a riboswitch-regulated collagen adhesion protein. *Sci. Rep.*, **6**, 1–12.

51. Bordeleau,E. and Burrus,V. (2015) Cyclic-di-GMP signaling in the Gram-positive pathogen Clostridium difficile. *Curr. Genet.*, **61**, 497–502.

52. Breaker,R.R. (2012) Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.*, **4**, a003566–a003566.

53. Helmling,C., Wacker,A., Wolfinger,M.T., Hofacker,I.L., Hengesbach,M., Fürtig,B. and Schwalbe,H. (2017) NMR Structural Profiling of Transcriptional Intermediates Reveals Riboswitch Regulation by Metastable RNA Conformations. *J. Am. Chem. Soc.*, **139**, 2647–2656.

54. Helmling,C., Klötzner,D.-P., Sochor,F., Mooney,R.A., Wacker,A., Landick,R., Fürtig,B., Heckel,A. and Schwalbe,H. (2018) Life times of metastable states guide regulatory signaling in transcriptional riboswitches. *Nat. Commun.*, **9**, 944.

55. Wickiser,J.K., Winkler,W.C., Breaker,R.R. and Crothers,D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell*, **18**, 49–60.

56. Wacker,A., Weigand,J.E., Akabayov,S.R., Altincekic,N., Bains,J.K., Banijamali,E., Binas,O., Castillo-Martinez,J., Cetiner,E., Ceylan,B., *et al.* (2020) Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.*, **48**, 12415–12435.

57. Wickiser,J.K., Cheah,M.T., Breaker,R.R. and Crothers,D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**, 13404–13414.

58. Haller,A., Soulière,M.F. and Micura,R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.*, **44**, 1339–1348.

59. Al-Hashimi,H.M. and Walter,N.G. (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.*, **18**, 321–329.

60. Scull,C.E., Dandpat,S.S., Romero,R.A. and Walter,N.G. (2021) Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control. *Front. Mol. Biosci.*, **7**, 1–10.

61. Sudarsan,N., Hammond,M.C., Block,K.F., Welz,R., Barrick,J.E., Roth,A. and Breaker,R.R. (2006) Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science (80-.).*,

314, 300–304.

62. Sherlock,M.E., Higgs,G., Yu,D., Widner,D.L., White,N.A., Sudarsan,N., Sadeeshkumar,H., Perkins,K.R., Mirihana Arachchilage,G., Malkowski,S.N., *et al.* (2022) Architectures and complex functions of tandem riboswitches. *RNA Biol.*, **19**, 1059–1076.

63. Averianova,L.A., Balabanova,L.A., Son,O.M., Podvolotskaya,A.B. and Tekutyeva,L.A. (2020) Production of Vitamin B2 (Riboflavin) by Microorganisms: An Overview. *Front. Bioeng. Biotechnol.*, **8**.

64. Higashitsuji,Y., Angerer,A., Berghaus,S., Hobl,B. and Mack,M. (2007) RibR, a possible regulator of the Bacillus subtilis riboflavin biosynthetic operon, in vivo interacts with the 5'-untranslated leader of rib mRNA. *FEMS Microbiol. Lett.*, **274**, 48–54.

65. Pedrolli,D.B., Kühm,C., Sévin,D.C., Vockenhuber,M.P., Sauer,U., Suess,B. and Mack,M. (2015) A dual control mechanism synchronizes riboflavin and sulphur metabolism in Bacillus subtilis. *Proc. Natl. Acad. Sci.*, **112**, 14054–14059.

66. Pursley,B.R., Fernandez,N.L., Severin,G.B. and Waters,C.M. (2019) The Vc2 Cyclic di-GMP-Dependent Riboswitch of Vibrio cholerae Regulates Expression of an Upstream Putative Small RNA by Controlling RNA Stability. *J. Bacteriol.*, **201**, 1–17.

67. Garst,A.D., Edwards,A.L. and Batey,R.T. (2011) Riboswitches: Structures and mechanisms. *Cold Spring Harb. Perspect. Biol.*, **3**, 1–13.

68. Batey,R.T. (2012) Structure and mechanism of purine-binding riboswitches. *Q. Rev. Biophys.*, **45**, 345–381.

69. Coppins,R.L., Hall,K.B. and Groisman,E.A. (2007) The intricate world of riboswitches. *Curr. Opin. Microbiol.*, **10**, 176–181.

70. Sun,T., Zhao,C. and Chen,S.-J. (2018) Predicting Cotranscriptional Folding Kinetics For Riboswitch. *J. Phys. Chem. B*, **122**, 7484–7496.

71. Widom,J.R., Nedialkov,Y.A., Rai,V., Hayes,R.L., Brooks,C.L., Artsimovitch,I. and Walter,N.G. (2018) Ligand Modulates Cross-Coupling between Riboswitch Folding and Transcriptional Pausing. *Mol. Cell*, **72**, 541-552.e6.

72. Burnouf,D., Ennifar,E., Guedich,S., Puffer,B., Hoffmann,G., Bec,G., Disdier,F., Baltzinger,M. and Dumas,P. (2012) KinITC: A new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *J. Am. Chem. Soc.*, **134**, 559–565.

73. Stagno,J.R., Liu,Y., Bhandari,Y.R., Conrad,C.E., Panja,S., Swain,M., Fan,L., Nelson,G., Li,C., Wendel,D.R., *et al.* (2017) Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature*, **541**, 242–246.

74. Guedich,S., Puffer-Enders,B., Baltzinger,M., Hoffmann,G., Da Veiga,C., Jossinet,F., Thore,S., Bec,G., Ennifar,E., Burnouf,D., *et al.* (2016) Quantitative and predictive model of kinetic regulation by E. coli TPP riboswitches. *RNA Biol.*, **13**, 373–390.

75. Beisel,C.L. and Smolke,C.D. (2009) Design Principles for Riboswitch Function. *PLoS Comput. Biol.*, **5**, e1000363.

76. Rode,A.B., Endoh,T. and Sugimoto,N. (2015) Tuning riboswitch-mediated gene regulation by rational control of aptamer ligand binding properties. *Angew. Chemie - Int. Ed.*, **54**, 905–909.

77. Lang,K., Rieder,R. and Micura,R. (2007) Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res.*, **35**, 5370–5378.

78. Greenleaf,W.J., Frieda,K.L., Foster,D.A.N., Woodside,M.T. and Block,S.M. (2008) Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science (80-.).*, **319**, 630–633.

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

80. Smith,K.D., Lipchock,S. V., Ames,T.D., Wang,J., Breaker,R.R. and Strobel,S.A. (2009) Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat. Struct. Mol. Biol.*, **16**, 1218–1223.

81. Smith,K.D., Lipchock,S. V., Livingston,A.L., Shanahan,C.A. and Strobel,S.A. (2010) Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry*, **49**, 7351–7359.

82. Zweckstetter,M., Holak,T.A. and Schwalbe,M. (2022) Magnetische Resonanzspektroskopie von Biomolekülen. In Kurreck,J., Engels,J.W., Lottspeich,F. (eds), *Bioanalytik*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 489–526.

83. Fürtig,B., Richter,C., Wöhnert,J. and Schwalbe,H. (2003) NMR spectroscopy of RNA. *ChemBioChem*, **4**, 936–962.

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

85. Sakamoto,T., Otsu,M. and Kawai,G. (2018) NMR Studies on RNA. In *Experimental Approaches of NMR Spectroscopy*. Springer Singapore, Singapore, pp. 439–459.

86. Plateau,P. and Gueron,M. (1982) Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**, 7310–7311.

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

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

89. Lundquist,K.P., Panchal,V., Gotfredsen,C.H., Brenk,R. and Clausen,M.H. (2021) Fragment-Based Drug Discovery for RNA Targets. *ChemMedChem*, **16**, 2588–2603.

90. Latham,M.P., Brown,D.J., McCallum,S.A. and Pardi,A. (2005) NMR methods for studying the

structure and dynamics of RNA. *ChemBioChem*, **6**, 1492–1505.

91. Zartler,E. and Shapiro,M. (2007) Protein NMR-Based Screening in Drug Discovery. *Curr. Pharm. Des.*, **12**, 3963–3972.

92. Emwas,A., Szczepski,K., Poulson,B.G., Chandra,K., McKay,R.T., Dhahri,M., Alahmari,F., Jaremko,L., Lachowicz,J.I. and Jaremko,M. (2020) NMR as a “Gold Standard” Method in Drug Design and Discovery. *Molecules*, **25**, 4597.

93. Diethelm-Varela,B. (2021) Using NMR Spectroscopy in the Fragment-Based Drug Discovery of Small-Molecule Anticancer Targeted Therapies. *ChemMedChem*, **16**, 725–742.

94. Haniff,H.S., Knerr,L., Chen,J.L., Disney,M.D. and Lightfoot,H.L. (2020) Target-Directed Approaches for Screening Small Molecules against RNA Targets. *SLAS Discov.*, **25**, 869–894.

95. Moumné,R., Catala,M., Larue,V., Micouin,L. and Tisné,C. (2012) Fragment-based design of small RNA binders: Promising developments and contribution of NMR. *Biochimie*, **94**, 1607–1619.

96. Johnson,E.C., Feher,V.A., Peng,J.W., Moore,J.M. and Williamson,J.R. (2003) Application of NMR SHAPES Screening to an RNA Target. *J. Am. Chem. Soc.*, **125**, 15724–15725.

97. Dalvit,C., Fogliatto,G., Stewart,A., Veronesi,M. and Stockman,B. (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR*, **21**, 349–359.

98. Yu,L., Oost,T.K., Schkeryantz,J.M., Yang,J., Janowick,D. and Fesik,S.W. (2003) Discovery of Aminoglycoside Mimetics by NMR-Based Screening of Escherichia c oli A-site RNA. *J. Am. Chem. Soc.*, **125**, 4444–4450.

99. Qin,P.Z. and Dieckmann,T. (2004) Application of NMR and EPR methods to the study of RNA. *Curr. Opin. Struct. Biol.*, **14**, 350–359.

100. Zídek,L., Štefl,R. and Sklenář,V. (2001) NMR methodology for the study of nucleic acids. *Curr. Opin. Struct. Biol.*, **11**, 275–281.

101. Dalvit,C., Pevarello,P., Tatò,M., Veronesi,M., Vulpetti,A., Sundström,M., Tato,M., Veronesi,M., Vulpetti,A., Sundström,M., *et al.* (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water*. *J. Biomol. NMR*, **18**, 65–68.

102. Becker,W., Bhattiprolu,K.C., Gubensäk,N. and Zangger,K. (2018) Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem*, **19**, 895–906.

103. Di Carluccio,C., Forgione,M.C., Martini,S., Berti,F., Molinaro,A., Marchetti,R. and Silipo,A. (2021) Investigation of protein-ligand complexes by ligand-based NMR methods. *Carbohydr. Res.*, **503**, 108313.

104. Clore,G.. and Gronenborn,A.. (1982) Theory and applications of the transferred nuclear overhauser effect to the study of the conformations of small ligands bound to proteins. *J. Magn. Reson.*, **48**, 402–417.

105. Albrand,J.P., Birdsall,B., Feeney,J., Roberts,G.C.K. and Burgen,A.S.V. (1979) The use of transferred nuclear Overhauser effects in the study of the conformations of small molecules bound to proteins. *Int. J. Biol. Macromol.*, **1**, 37–41.

106. Li,D., DeRose,E.F. and London,R.E. (1999) The inter-ligand Overhauser effect: A powerful new NMR approach for mapping structural relationships of macromolecular ligands. *J. Biomol. NMR*, **15**, 71–76.

107. Cala,O., Guillière,F. and Krimm,I. (2014) NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.*, **406**, 943–956.

108. Balaram,P., Bothner-By,A.A. and Dadok,J. (1972) Negative nuclear Overhuaser effects as probes of macromolecular structure. *J. Am. Chem. Soc.*, **94**, 4015–4017.

109. Anglister,J., Srivastava,G. and Naider,F. (2016) Detection of intermolecular NOE interactions in large protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.*, **97**, 40–56.

110. Kreishman,G.P., Chan,S.I. and Bauer,W. (1971) Proton magnetic resonance study of the interaction of ethidium bromide with several uracil residues, uridylyl (3′ → 5′) uridine and polyuridylic acid. *J. Mol. Biol.*, **61**, 45–58.

111. Sánchez-Pedregal,V.M., Reese,M., Meiler,J., Blommers,M.J.J., Griesinger,C. and Carlomagno,T. (2005) The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. *Angew. Chemie - Int. Ed.*, **44**, 4172–4175.

112. Mayer,M. and Meyer,B. (1999) Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.*, **38**, 1784–1788.

113. Antanasijevic,A., Ramirez,B. and Caffrey,M. (2014) Comparison of the sensitivities of WaterLOGSY and saturation transfer difference NMR experiments. *J. Biomol. NMR*, **60**, 37–44.

114. Shortridge,M.D., Hage,D.S., Harbison,G.S. and Powers,R. (2008) Estimating Protein−Ligand Binding Affinity Using High-Throughput Screening by NMR. *J. Comb. Chem.*, **10**, 948–958.

115. Aretz,J. and Rademacher,C. (2019) Ranking hits from saturation transfer difference nuclear magnetic resonance-based fragment screening. *Front. Chem.*, **7**, 1–8.

116. Norton,R., Leung,E., Chandrashekaran,I. and MacRaild,C. (2016) Applications of 19F-NMR in Fragment-Based Drug Discovery. *Molecules*, **21**, 860.

117. Morris,G.A. (2009) Diffusion-Ordered Spectroscopy. In *Encyclopedia of Magnetic Resonance*. John Wiley & Sons, Ltd, Chichester, UK, Vol. 2009, pp. 1–13.

118. Iwahara,J., Tang,C. and Marius Clore,G. (2007) Practical aspects of 1H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.*, **184**, 185–195.

119. Price,A.J., Howard,S. and Cons,B.D. (2017) Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem.*, **61**, 475–484.

120. Shuker,S.B., Hajduk,P.J., Meadows,R.P. and Fesik,S.W. (1996) Discovering High-Affinity

Ligands for Proteins: SAR by NMR. *Science (80-.).*, **274**, 1531–1534.

121. Congreve,M., Chessari,G., Tisi,D. and Woodhead,A.J. (2008) Recent developments in fragment-based drug discovery. *J. Med. Chem.*, **51**, 3661–3680.

122. Fink,T., Bruggesser,H. and Reymond,J.-L. (2005) Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. *Angew. Chemie Int. Ed.*, **44**, 1504–1508.

123. Hann,M.M., Leach,A.R. and Harper,G. (2001) Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.*, **41**, 856–864.

124. Leach,A.R., Hann,M.M., Burrows,J.N. and Griffen,E.J. (2006) Fragment screening: An introduction. *Mol. Biosyst.*, **2**, 430–446.

125. Congreve,M., Carr,R., Murray,C. and Jhoti,H. (2003) A ‘Rule of Three’ for fragment-based lead discovery? *Drug Discov. Today*, **8**, 876–877.

126. Nitsche,C. and Otting,G. (2018) NMR studies of ligand binding. *Curr. Opin. Struct. Biol.*, **48**, 16–22.

127. Erlanson,D.A., Fesik,S.W., Hubbard,R.E., Jahnke,W. and Jhoti,H. (2016) Twenty years on: The impact of fragments on drug discovery. *Nat. Rev. Drug Discov.*, **15**, 605–619.

128. Tenson,T. and Mankin,A. (2006) Antibiotics and the ribosome. *Mol. Microbiol.*, **59**, 1664–1677.

129. Poehlsgaard,J. and Douthwaite,S. (2005) The bacterial ribosome as a target for antibiotics. *Nat. Rev. Microbiol.*, **3**, 870–881.

130. Westhof,E. (2006) The ribosomal decoding site and antibiotics. *Biochimie*, **88**, 931–933.

131. Weeks,K.M. and Crothers,D.M. (1993) Major Groove Accessibility of RNA. *Science (80-.).*, **261**, 1574–1577.

132. Lee,M.-K., Bottini,A., Kim,M., Bardaro,M.F., Zhang,Z., Pellecchia,M., Choi,B.-S. and Varani,G. (2014) A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.*, **50**, 368–370.

133. Davidson,A., Begley,D.W., Lau,C. and Varani,G. (2011) A Small-Molecule Probe Induces a Conformation in HIV TAR RNA Capable of Binding Drug-Like Fragments. *J. Mol. Biol.*, **410**, 984–996.

134. Cressina,E., Chen,L., Moulin,M., Leeper,F.J., Abell,C. and Smith,A.G. (2011) Identification of novel ligands for thiamine pyrophosphate (TPP) riboswitches. *Biochem. Soc. Trans.*, **39**, 652–657.

135. Warner,K.D., Homan,P., Weeks,K.M., Smith,A.G., Abell,C. and Ferré-D’Amaré,A.R. (2014) Validating fragment-based drug discovery for biological RNAs: Lead fragments bind and remodel the TPP riboswitch specifically. *Chem. Biol.*, **21**, 591–595.

136. Sreeramulu,S., Richter,C., Berg,H., Wirtz Martin,M.A., Ceylan,B., Matzel,T., Adam,J., Altincekic,N., Azzaoui,K., Bains,J.K., *et al.* (2021) Exploring the Druggability of Conserved RNA Regulatory Elements in the SARS-CoV-2 Genome. *Angew. Chemie - Int. Ed.*, **60**, 19191–19200.

137. Shortridge,M.D. and Varani,G. (2021) Efficient NMR Screening Approach to Discover Small Molecule Fragments Binding Structured RNA. *ACS Med. Chem. Lett.*, **12**, 1253–1260.

138. Tam,B., Sherf,D., Cohen,S., Eisdorfer,S.A., Perez,M., Soffer,A., Vilenchik,D., Akabayov,S.R., Wagner,G. and Akabayov,B. (2019) Discovery of small-molecule inhibitors targeting the ribosomal peptidyl transferase center (PTC) of M. tuberculosis. *Chem. Sci.*, **10**, 8764–8767.

139. Garavís,M., López-Méndez,B., Somoza,A., Oyarzabal,J., Dalvit,C., Villasante,A., Campos-Olivas,R. and González,C. (2014) Discovery of Selective Ligands for Telomeric RNA G-quadruplexes (TERRA) through 19 F-NMR Based Fragment Screening. *ACS Chem. Biol.*, **9**, 1559–1566.

140. Chung,F., Tisné,C., Lecourt,T., Dardel,F. and Micouin,L. (2007) NMR-Guided Fragment-Based Approach for the Design of tRNALys3 Ligands. *Angew. Chemie Int. Ed.*, **46**, 4489–4491.

141. Chen,L., Cressina,E., Leeper,F.J., Smith,A.G. and Abell,C. (2010) A Fragment-Based Approach to Identifying Ligands for Riboswitches. *ACS Chem. Biol.*, **5**, 355–358.

142. Marousis,K.D., Tsika,A.C., Birkou,M., Matsoukas,M.T. and Spyroulias,G.A. (2018) Lead identification through the synergistic action of biomolecular NMR and in silico methodologies. In *Methods in Molecular Biology*.Vol. 1824, pp. 299–316.

143. Kurreck,J., Engels,J.W. and Lottspeich,F. eds. (2022) Bioanalytik Springer Berlin Heidelberg, Berlin, Heidelberg.

144. Blume,A. (2022) Mikrokalorimetrie. In *Bioanalytik*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 59–76.

145. Song,C., Zhang,S. and Huang,H. (2015) Choosing a suitable method for the identification of replication origins in microbial genomes. *Front. Microbiol.*, **6**, 1–18.

146. Sturtevant,J.M. (1937) The Calculation of Second Order Reaction Velocity Constants. *J. Am. Chem. Soc.*, **59**, 699–702.

147. Bianconi,M.L. (2007) Calorimetry of enzyme-catalyzed reactions. *Biophys. Chem.*, **126**, 59–64.

148. Zuker,M. (1989) On Finding All Suboptimal Foldings of an RNA Molecule. *Science (80-.).*, **244**, 48–52.

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

150. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.

151. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.*, **101**, 7287–7292.

152. Badelt,S., Hammer,S., Flamm,C. and Hofacker,I.L. (2015) Thermodynamic and Kinetic Folding of Riboswitches. In *Methods in Enzymology*. Elsevier Inc., Vol. 553, pp. 193–213.

153. Meyer,I.M. (2017) In silico methods for co-transcriptional RNA secondary structure prediction and for investigating alternative RNA structure expression. *Methods*, **120**, 3–16.

154. Levinthal,C. (1968) Are there pathways for protein folding? *J. Chim. Phys.*, **65**, 44–45.

155. Levinthal,C. (1969) How to fold graciously. In *Mössbaun Spectroscopy in Biological Systems Proceedings*.Vol. 67, pp. 22–24.

156. Dill,K.A. and MacCallum,J.L. (2012) The Protein-Folding Problem, 50 Years On. *Science (80-.).*, **338**, 1042–1046.

157. Wolfinger,M.T., Flamm,C. and Hofacker,I.L. (2018) Efficient computation of co-transcriptional RNA-ligand interaction dynamics. *Methods*, **143**, 70–76.

158. Proctor,J.R. and Meyer,I.M. (2013) CoFold: An RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Res.*, **41**, 1–11.

159. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.

160. DANILOVA,L. V., PERVOUCHINE,D.D., FAVOROV,A. V. and MIRONOV,A.A. (2006) RNAKINETICS: A WEB SERVER THAT MODELS SECONDARY STRUCTURE KINETICS OF AN ELONGATING RNA. *J. Bioinform. Comput. Biol.*, **04**, 589–596.

161. Xayaphoummine,A., Bucher,T. and Isambert,H. (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.*, **33**, W605–W610.

162. Gong,S., Wang,Y., Wang,Z., Wang,Y. and Zhang,W. (2018) Genetic regulation mechanism of the yjdF riboswitch. *J. Theor. Biol.*, **439**, 152–159.

163. Fu,L., Cao,Y., Wu,J., Peng,Q., Nie,Q. and Xie,X. (2022) UFold: fast and accurate RNA secondary structure prediction with deep learning. *Nucleic Acids Res.*, **50**, e14–e14.

164. Lin,J.-C. and Thirumalai,D. (2012) Gene Regulation by Riboswitches with and without Negative Feedback Loop. *Biophys. J.*, **103**, 2320–2330.

165. Lin,J.-C., Yoon,J., Hyeon,C. and Thirumalai,D. (2015) Using Simulations and Kinetic Network Models to Reveal the Dynamics and Functions of Riboswitches. In *Physiology & behavior*.Vol. 176, pp. 235–258.

166. Parra-Rojas,C., Fürtig,B., Schwalbe,H. and Hernandez-Vargas,E.A. (2020) Quantitative modeling of the function of kinetically driven transcriptional riboswitches. *J. Theor. Biol.*, **506**, 110406.

167. Gong,S., Wang,Y. and Zhang,W. (2015) Kinetic regulation mechanism of pbuE riboswitch. *J. Chem. Phys.*, **142**.

168. Gong,S., Wang,Y., Wang,Z., Wang,Y. and Zhang,W. (2016) Reversible-Switch Mechanism of the SAM-III Riboswitch. *J. Phys. Chem. B*, **120**, 12305–12311.

169. Gong,S., Wang,Y., Wang,Z. and Zhang,W. (2017) Co-Transcriptional Folding and Regulation

Mechanisms of Riboswitches. *Molecules*, **22**, 1169.

170. Zhao,P., Zhang,W.B. and Chen,S.J. (2010) Predicting secondary structural folding kinetics for nucleic acids. *Biophys. J.*, **98**, 1617–1625.

171. Zhang,W. and Chen,S.-J. (2006) Exploring the Complex Folding Kinetics of RNA Hairpins: I. General Folding Kinetics Analysis. *Biophys. J.*, **90**, 765–777.

172. Purcell,E.B. and Tamayo,R. (2016) Cyclic diguanylate signaling in Gram-positive bacteria. *FEMS Microbiol. Rev.*, **40**, 753–773.

173. McKee,R.W., Harvest,C.K. and Tamayo,R. (2018) Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in Clostridium difficile . *mSphere*, **3**, 1–15.

174. Dapa,T. and Unnikrishnan,M. (2013) Biofilm formation by Clostridium difficile. *Gut Microbes*, **4**, 397–402.

175. Dapa,T., Leuzzi,R., Ng,Y.K., Baban,S.T., Adamo,R., Kuehne,S.A., Scarselli,M., Minton,N.P., Serruto,D. and Unnikrishnan,M. (2013) Multiple factors modulate biofilm formation by the anaerobic pathogen Clostridium difficile. *J. Bacteriol.*, **195**, 545–555.

176. Semenyuk,E.G., Laning,M.L., Foley,J., Johnston,P.F., Knight,K.L., Gerding,D.N. and Driks,A. (2014) Spore Formation and Toxin Production in Clostridium difficile Biofilms. *PLoS One*, **9**, e87757.

177. Purcell,E.B., McKee,R.W., Bordeleau,E., Burrus,V. and Tamayo,R. (2016) Regulation of Type IV Pili contributes to surface behaviors of historical and epidemic strains of Clostridium difficile. *J. Bacteriol.*, **198**, 565–577.

178. Keller,H., Weickhmann,A.K., Bock,T. and Wöhnert,J. (2018) Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. *Rna*, **24**, 1390–1402.

179. Reguera,G., McCarthy,K.D., Mehta,T., Nicoll,J.S., Tuominen,M.T. and Lovley,D.R. (2005) Extracellular electron transfer via microbial nanowires. *Nature*, **435**, 1098–1101.

180. Lovley,D.R., Holmes,D.E. and Nevin,K.P. (2004) Dissimilatory Fe(III) and Mn(IV) Reduction. In *Advances in Microbial Physiology*.Vol. 49, pp. 219–286.

181. Hua,B., Panja,S., Wang,Y., Woodson,S.A. and Ha,T. (2018) Mimicking Co-Transcriptional RNA Folding Using a Superhelicase. *J. Am. Chem. Soc.*, **140**, 10067–10070.

182. Frieda,K.L. and Block,S.M. (2012) Direct Observation of Cotranscriptional Folding in an Adenine Riboswitch. *Science (80-.).*, **338**, 397–400.

183. Strobel,E.J., Watters,K.E., Nedialkov,Y., Artsimovitch,I. and Lucks,J.B. (2017) Distributed biotin–streptavidin transcription roadblocks for mapping cotranscriptional RNA folding. *Nucleic Acids Res.*, **45**, e109–e109.

184. Boyle,J., Robillard,G.T. and Kim,S.H. (1980) Sequential folding of transfer RNA. A nuclear magnetic resonance study of successively longer tRNA fragments with a common 5′ end. *J. Mol. Biol.*, **139**, 601–625.

185. Fürtig,B., Wenter,P., Reymond,L., Richter,C., Pitsch,S. and Schwalbe,H. (2007) Conformational Dynamics of Bistable RNAs Studied by Time-Resolved NMR Spectroscopy. *J. Am. Chem. Soc.*, **129**, 16222–16229.

186. Kuznetsov,S. V. and Ansari,A. (2012) A kinetic zipper model with intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys. J.*, **102**, 101–111.

187. Binas,O., Schamber,T. and Schwalbe,H. (2020) The conformational landscape of transcription intermediates involved in the regulation of the ZMP-sensing riboswitch from Thermosinus carboxydivorans. *Nucleic Acids Res.*, **48**, 6970–6979.

188. Furukawa,K., Gu,H., Sudarsan,N., Hayakawa,Y., Hyodo,M. and Breaker,R.R. (2012) Identification of ligand analogues that control c-di-GMP riboswitches. *ACS Chem. Biol.*, **7**, 1436–1443.

189. Nozinovic,S., Reining,A., Kim,Y.-B.B., Noeske,J., Schlepckow,K., Wöhnert,J., Schwalbe,H., Wöhnert,J., Schwalbe,H., Wöhnert,J., *et al.* (2014) The importance of helix P1 stability for structural pre-organization and ligand binding affinity of the adenine riboswitch aptamer domain. *RNA Biol.*, **11**, 82–83.

190. Inuzuka,S., Kakizawa,H., Nishimura,K.I., Naito,T., Miyazaki,K., Furuta,H., Matsumura,S. and Ikawa,Y. (2018) Recognition of cyclic-di-GMP by a riboswitch conducts translational repression through masking the ribosome-binding site distant from the aptamer domain. *Genes to Cells*, **23**, 435–447.

191. Barraud,N., Schleheck,D., Klebensberger,J., Webb,J.S., Hassett,D.J., Rice,S.A. and Kjelleberg,S. (2009) Nitric Oxide Signaling in Pseudomonas aeruginosa Biofilms Mediates Phosphodiesterase Activity, Decreased Cyclic Di-GMP Levels, and Enhanced Dispersal. *J. Bacteriol.*, **191**, 7333–7342.

192. Simm,R., Morr,M., Kader,A., Nimtz,M. and Römling,U. (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Mol. Microbiol.*, **53**, 1123–1134.

193. Kulasekara,B.R., Kamischke,C., Kulasekara,H.D., Christen,M., Wiggins,P.A. and Miller,S.I. (2013) c-di-GMP heterogeneity is generated by the chemotaxis machinery to regulate flagellar motility. *Elife*, **2**, 1–19.

194. Purcell,E.B., McKee,R.W., McBride,S.M., Waters,C.M. and Tamayo,R. (2012) Cyclic diguanylate inversely regulates motility and aggregation in clostridium difficile. *J. Bacteriol.*, **194**, 3307–3316.

195. Hengge,R. (2021) High-specificity local and global c-di-GMP signaling. *Trends Microbiol.*, **29**, 993–1003.

196. Paul,R., Weiser,S., Amiot,N.C., Chan,C., Schirmer,T., Giese,B. and Jenal,U. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.*, **18**, 715–727.

197. Vicens,Q. and Westhof,E. (2003) RNA as a drug target: The case of aminoglycosides.

ChemBioChem, **4**, 1018–1023.

198. Warner,K.D., Hajdin,C.E. and Weeks,K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.*, **17**, 547–558.

199. Garcia-Lopez,A., Tessaro,F., Jonker,H.R.A.A., Wacker,A., Richter,C., Comte,A., Berntenis,N., Schmucki,R., Hatje,K., Petermann,O., *et al.* (2018) Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes. *Nat. Commun.*, **9**, 1–12.

200. Sivaramakrishnan,M., McCarthy,K.D., Campagne,S., Huber,S., Meier,S., Augustin,A., Heckel,T., Meistermann,H., Hug,M.N., Birrer,P., *et al.* (2017) Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.*, **8**, 1–13.

201. Donlic,A. and Hargrove,A.E. (2018) Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip. Rev. RNA*, **9**, e1477.

202. Matsui,M. and Corey,D.R. (2017) Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.*, **16**, 167–179.

203. Palacino,J., Swalley,S.E., Song,C., Cheung,A.K., Shu,L., Zhang,X., Van Hoosear,M., Shin,Y., Chin,D.N., Keller,C.G., *et al.* (2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.*, **11**, 511–517.

204. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in Bacteria. *Cell*, **136**, 615–628.

205. van Duin,D. and Paterson,D.L. (2016) Multidrug-Resistant Bacteria in the Community: Trends and Lessons Learned. *Infect. Dis. Clin. North Am.*, **30**, 377–390.

206. Panchal,V. and Brenk,R. (2021) Riboswitches as drug targets for antibiotics. *Antibiotics*, **10**, 1–22.

207. S. Butler,M. and A. Cooper,M. (2012) Screening Strategies to Identify New Antibiotics. *Curr. Drug Targets*, **13**, 373–387.

208. Davies,D.R. (2014) Screening Ligands by X-ray Crystallography. In *Methods in molecular biology (Clifton, N.J.)*. Humana Press, New York, NY, Vol. 1140, pp. 315–323.

209. Sugiki,T., Furuita,K., Fujiwara,T. and Kojima,C. (2018) Current NMR techniques for structure-based drug discovery. *Molecules*, **23**, 1–27.

210. Liu,X., Shi,D., Zhou,S., Liu,H.H.H.H., Liu,H.H.H.H. and Yao,X. (2018) Molecular dynamics simulations and novel drug discovery. *Expert Opin. Drug Discov.*, **13**, 23–37.

211. Erlanson,D.A. (2011) Introduction to Fragment-Based Drug Discovery. In *Topics in current chemistry*.Vol. 317, pp. 1–32.

212. Erlanson,D.A., De Esch,I.J.P.P., Jahnke,W., Johnson,C.N. and Mortenson,P.N. (2020) Fragment-to-lead medicinal chemistry publications in 2018. *J. Med. Chem.*, **63**, 4430–4444.

213. Tsai,J., Lee,J.T., Wang,W., Zhang,J., Cho,H., Mamo,S., Bremer,R., Gillette,S., Kong,J., Haass,N.K., *et al.* (2008) Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci.*, **105**, 3041–3046.

214. Brough,P.A., Aherne,W., Barril,X., Borgognoni,J., Boxall,K., Cansfield,J.E., Cheung,K.M.J., Collins,I., Davies,N.G.M., Drysdale,M.J., *et al.* (2008) 4,5-Diarylisoxazole Hsp90 chaperone inhibitors: Potential therapeutic agents for the treatment of cancer. *J. Med. Chem.*, **51**, 196–218.

215. Howard,S., Berdini,V., Boulstridge,J.A., Carr,M.G., Cross,D.M., Curry,J., Devine,L.A., Early,T.R., Fazal,L., Gill,A.L., *et al.* (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J. Med. Chem.*, **52**, 379–388.

216. Park,C.-M., Bruncko,M., Adickes,J., Bauch,J., Ding,H., Kunzer,A., Marsh,K.C., Nimmer,P., Shoemaker,A.R., Song,X., *et al.* (2008) Discovery of an Orally Bioavailable Small Molecule Inhibitor of Prosurvival B-Cell Lymphoma 2 Proteins. *J. Med. Chem.*, **51**, 6902–6915.

217. Wang,Y. Sen, Strickland,C., Voigt,J.H., Kennedy,M.E., Beyer,B.M., Senior,M.M., Smith,E.M., Nechuta,T.L., Madison,V.S., Czarniecki,M., *et al.* (2010) Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel µM leads for the development of nM BACE-1 (β-site APP cleaving enzyme 1) inhibitors. *J. Med. Chem.*, **53**, 942–950.

218. Largy,E., Hamon,F. and Teulade-Fichou,M.P. (2011) Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.*, **400**, 3419–3427.

219. Asare-Okai,P.N. and Chow,C.S. (2011) A modified fluorescent intercalator displacement assay for RNA ligand discovery. *Anal. Biochem.*, **408**, 269–276.

220. Laughlin,S. and David Wilson,W. (2015) May the best molecule win: Competition ESI mass spectrometry. *Int. J. Mol. Sci.*, **16**, 24506–24531.

221. Shimizu,H., Jinno,F., Morohashi,A., Yamazaki,Y., Yamada,M., Kondo,T. and Asahi,S. (2012) Application of high-resolution ESI and MALDI mass spectrometry to metabolite profiling of small interfering RNA duplex. *J. Mass Spectrom.*, **47**, 1015–1022.

222. Sztuba-Solinska,J., Shenoy,S.R., Gareiss,P., Krumpe,L.R.H., Le Grice,S.F.J., O'Keefe,B.R. and Schneekloth,J.S. (2014) Identification of biologically active, HIV TAR RNA-binding small molecules using small molecule microarrays. *J. Am. Chem. Soc.*, **136**, 8402–8410.

223. Connelly,C.M., Boer,R.E., Moon,M.H., Gareiss,P. and Schneekloth,J.S. (2017) Discovery of Inhibitors of MicroRNA-21 Processing Using Small Molecule Microarrays. *ACS Chem. Biol.*, **12**, 435–443.

224. Moon,M.H., Hilimire,T.A., Sanders,A.M. and Schneekloth,J.S. (2018) Measuring RNA-Ligand Interactions with Microscale Thermophoresis. *Biochemistry*, **57**, 4638–4643.

225. Jerabek-Willemsen,M., André,T., Wanner,R., Roth,H.M., Duhr,S., Baaske,P. and Breitsprecher,D. (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.*, **1077**, 101–113.

226. Pellecchia,M., Bertini,I., Cowburn,D., Dalvit,C., Giralt,E., Jahnke,W., James,T.L., Homans,S.W., Kessler,H., Luchinat,C., *et al.* (2008) Perspectives on NMR in drug discovery: A technique

comes of age. *Nat. Rev. Drug Discov.*, **7**, 738–745.

227. Zeiger,M., Stark,S., Kalden,E., Ackermann,B., Ferner,J., Scheffer,U., Shoja-Bazargani,F., Erdel,V., Schwalbe,H. and Göbel,M.W. (2014) Fragment based search for small molecule inhibitors of HIV-1 Tat-TAR. *Bioorganic Med. Chem. Lett.*, **24**, 5576–5580.

228. Begley,D.W., Moen,S.O., Pierce,P.G. and Zartler,E.R. (2013) Saturation transfer difference NMR for fragment screening. *Curr. Protoc. Chem. Biol.*, **5**, 251–268.

229. Carr,H.Y. and Purcell,E.M. (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, **94**, 630–638.

230. Kim,Y. and Hilty,C. (2019) Applications of Dissolution-DNP for NMR Screening. In *Methods in Enzymology*.Vol. 615, pp. 501–526.

231. Dalvit,C., Fagerness,P.E., Hadden,D.T.A., Sarver,R.W. and Stockman,B.J. (2003) Fluorine-NMR experiments for high-throughput screening: Theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.*, **125**, 7696–7703.

232. Meiboom,S. and Gill,D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.*, **29**, 688–691.

233. Groebe,D.R. and Uhlenbeck,O.C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res.*, **16**, 11725–11735.

234. Sheehy,J.P., Davis,A.R. and Znosko,B.M. (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *Rna*, **16**, 417–429.

235. Childs-Disney,J.L., Hoskins,J., Rzuczek,S.G., Thornton,C.A. and Disney,M.D. (2012) Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol.*, **7**, 856–862.

236. Stelzer,A.C., Frank,A.T., Kratz,J.D., Swanson,M.D., Gonzalez-Hernandez,M.J., Lee,J., Andricioaei,I., Markovitz,D.M. and Al-Hashimi,H.M. (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.*, **7**, 553–559.

237. Hong,C., Hagihara,M. and Nakatani,K. (2011) Ligand-assisted complex formation of two DNA hairpin loops. *Angew. Chemie - Int. Ed.*, **50**, 4390–4393.

238. Bailor,M.H., Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2011) Topological constraints: Using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr. Opin. Struct. Biol.*, **21**, 296–305.

239. Bailor,M.H., Musselman,C., Hansen,A.L., Gulati,K., Patel,D.J. and Al-Hashimi,H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat. Protoc.*, **2**, 1536–1546.

240. Hamy,F., Brondani,V., Flörsheimer,A., Stark,W., Blommers,M.J.J. and Klimkait,T. (1998) A new class of HIV-1 Tat antagonist acting through Tat-TAR inhibition. *Biochemistry*, **37**, 5086–5095.

241. Brodsky,A.S. and Williamson,J.R. (1997) Solution structure of the HIV-2 TAR-argininamide

complex. *J. Mol. Biol.*, **267**, 624–639.

242. Frank,A.T., Stelzer,A.C., Al-Hashimi,H.M. and Andricioaei,I. (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: New insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.*, **37**, 3670–3679.

243. Patwardhan,N.N., Ganser,L.R., Kapral,G.J., Eubanks,C.S., Lee,J., Sathyamoorthy,B., Al-Hashimi,H.M. and Hargrove,A.E. (2017) Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *Medchemcomm*, **8**, 1022–1036.

244. Mandal,M., Boese,B., Barrick,J.E., Winkler,W.C. and Breaker,R.R. (2003) Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, **113**, 577–586.

245. Schwalbe,H., Buck,J., Fürtig,B., Noeske,J. and Wöhnert,J. (2007) Structures of RNA Switches: Insight into Molecular Recognition and Tertiary Structure. *Angew. Chemie Int. Ed.*, **46**, 1212–1219.

246. Gilbert,S.D., Mediatore,S.J. and Batey,R.T. (2006) Modified pyrimidines specifically bind the purine riboswitch. *J. Am. Chem. Soc.*, **128**, 14214–14215.

247. Lee,E.R., Blount,K.F. and Breaker,R.R. (2009) Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol.*, **6**, 187–194.

248. Marino,J.P., Gregorian,R.S., Csankovszki,G. and Crothers,D.M. (1995) Bent Helix Formation Between RNA Hairpins with Complementary Loops. *Science (80-.).*, **268**, 1448–1454.

249. Ganser,L.R., Lee,J., Rangadurai,A., Merriman,D.K., Kelly,M.L., Kansal,A.D., Sathyamoorthy,B. and Al-Hashimi,H.M. (2018) High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nat. Struct. Mol. Biol.*, **25**, 425–434.

250. Williamson,M.P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.*, **73**, 1–16.

251. Rüdisser,S.H., Goldberg,N., Ebert,M.O., Kovacs,H. and Gossert,A.D. (2020) Efficient affinity ranking of fluorinated ligands by 19F NMR: CSAR and FastCSAR. *J. Biomol. NMR*, **74**, 579–594.

252. Kirk,S.R., Luedtke,N.W. and Tor,Y. (2000) Neomycin-acridine conjugate: A potent inhibitor of Rev-RRE binding [16]. *J. Am. Chem. Soc.*, **122**, 980–981.

253. Prasher,P. and Sharma,M. (2018) Medicinal chemistry of acridine and its analogues. *Medchemcomm*, **9**, 1589–1618.

254. Arambula,J.F., Ramisetty,S.R., Baranger,A.M. and Zimmerman,S.C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl. Acad. Sci.*, **106**, 16068–16073.

255. Hajduk,P.J., Huth,J.R. and Tse,C. (2005) Predicting protein druggability. *Drug Discov. Today*, **10**, 1675–1682.

256. Rzuczek,S.G., Southern,M.R. and Disney,M.D. (2015) Studying a Drug-like, RNA-Focused Small

Molecule Library Identifies Compounds That Inhibit RNA Toxicity in Myotonic Dystrophy. *ACS Chem. Biol.*, **10**, 2706–2715.

257. Morgan,B.S., Forte,J.E., Culver,R.N., Zhang,Y. and Hargrove,A.E. (2017) Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chemie - Int. Ed.*, **56**, 13498–13502.

258. Kato,K., Ishii,R., Hirano,S., Ishitani,R. and Nureki,O. (2015) Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP-AMP synthase. *Structure*, **23**, 843–850.

259. Sklenar,V. (1995) Suppression of Radiation Damping in Multidimensional NMR Experiments Using Magnetic Field Gradients. *J. Magn. Reson. Ser. A*, **114**, 132–135.

260. Keller,S., Vargas,C., Zhao,H., Piszczek,G., Brautigam,C.A. and Schuck,P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, **84**, 5066–5073.

261. Zhao,H., Piszczek,G. and Schuck,P. (2015) SEDPHAT - A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods*, **76**, 137–148.

262. Prinz,J.-H., Wu,H., Sarich,M., Keller,B., Senne,M., Held,M., Chodera,J.D., Schütte,C. and Noé,F. (2011) Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.*, **134**, 174105.

263. Vassylyev,D.G., Vassylyeva,M.N., Perederina,A., Tahirov,T.H. and Artsimovitch,I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, **448**, 157–162.

264. Bains,J.K., Blechar,J., de Jesus,V., Meiser,N., Zetzsche,H., Fürtig,B., Schwalbe,H. and Hengesbach,M. (2019) Combined smFRET and NMR analysis of riboswitch structural dynamics. *Methods*, **153**, 22–34.

265. Hwang,T.L. and Shaka,A.J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. Ser. A*, **112**, 275–279.

266. Marius Clore,G., Kimber,B.J. and Gronenborn,A.M. (1983) The 1-1 hard pulse: A simple and effective method of water resonance suppression in FT 1H NMR. *J. Magn. Reson.*, **54**, 170–173.

267. Huang,L., Wang,J. and Lilley,D.M.J. (2017) The Structure of the Guanidine-II Riboswitch. *Cell Chem. Biol.*, **24**, 695-702.e2.

268. Ren,A., Rajashankar,K.R. and Patel,D.J. (2015) Global RNA Fold and Molecular Recognition for a pfl Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. *Structure*, **23**, 1375–1381.

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

270. Kulshina,N., Baird,N.J., Ferré-D’Amaré,A.R., Amaré,A.R.F. and Ferré-D’Amaré,A.R. (2009) Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat. Struct. Mol. Biol.*, **16**, 1212–1217.

271. Serganov,A., Yuan,Y.R., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A.T., Hobartner,C., Micura,R., Breaker,R.R. and Patel,D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, **11**, 1729–1741.

272. Zhang,J. and Ferré-D'Amaré,A.R. (2014) Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. *Structure*, **22**, 1363–1371.

273. Rieder,R., Lang,K., Graber,D. and Micura,R. (2007) Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control. *ChemBioChem*, **8**, 896–902.

274. Schnieders,R., Richter,C., Warhaut,S., de Jesus,V., Keyhani,S., Duchardt-Ferner,E., Keller,H., Wöhnert,J., Kuhn,L.T., Breeze,A.L., *et al.* (2017) Evaluation of 15N-detected H–N correlation experiments on increasingly large RNAs. *J. Biomol. NMR*, **69**, 31–44.

275. Okada,K., Takahashi,M., Sakamoto,T., Kawai,G., Nakamura,K. and Kanai,A. (2006) Solution structure of a GAAG tetraloop in helix 6 of SRP RNA from Pyrococcus furiosus. *Nucleosides, Nucleotides and Nucleic Acids*, **25**, 383–395.

276. Jucker,F.M. and Pardi,A. (1995) Solution Structure of the CUUG Hairpin Loop: A Novel RNA Tetraloop Motif. *Biochemistry*, **34**, 14416–14427.

277. Scott,W.G., Finch,J.T. and Klug,A. (1995) The crystal structure of an AII-RNAhammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991–1002.

278. Ferré-D'Amaré,A.R., Zhou,K. and Doudna,J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567–574.

279. Woo,N.H., Roe,B.A. and Rich,A. (1980) Three-dimensional structure of escherichia coli initiator tRNAfmet. *Nature*, **286**, 346–351.

280. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.

281. Dai,J., Carver,M., Hurley,L.H. and Yang,D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: Insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.*, **133**, 17673–17680.

282. Wei,D., Parkinson,G.N., Reszka,A.P. and Neidle,S. (2012) Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.*, **40**, 4691–4700.

283. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.

284. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, **78**, 2179–2183.

285. Dai,J., Punchihewa,C., Ambrus,A., Chen,D., Jones,R.A. and Yang,D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: A novel adenine triple formation. *Nucleic Acids Res.*, **35**, 2440–2450.

286. Liu,W., Zhong,Y.F., Liu,L.Y., Shen,C.T., Zeng,W., Wang,F., Yang,D. and Mao,Z.W. (2018) Solution structures of multiple G-quadruplex complexes induced by a platinum(II)-based tripod reveal dynamic binding. *Nat. Commun.*, **9**, 3496.

287. Wirmer-Bartoschek,J., Bendel,L.E., Jonker,H.R.A., Grün,J.T., Papi,F., Bazzicalupi,C., Messori,L., Gratteri,P. and Schwalbe,H. (2017) Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding. *Angew. Chemie - Int. Ed.*, **56**, 7102–7106.

288. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: Insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.

289. Madhurantakam,C., Rajakumara,E., Mazumdar,P.A., Saha,B., Mitra,D., Wiker,H.G., Sankaranarayanan,R. and Das,A.K. (2005) Crystal structure of low-molecular-weight protein tyrosine phosphatase from Mycobacterium tuberculosis at 1.9-Å resolution. *J. Bacteriol.*, **187**, 2175–2181.

290. Stehle,T., Sreeramulu,S., Löhr,F., Richter,C., Saxena,K., Jonker,H.R.A. and Schwalbe,H. (2012) The apo-structure of the low molecular weight protein-tyrosine phosphatase A (MptpA) from Mycobacterium tuberculosis allows for better target-specific drug development. *J. Biol. Chem.*, **287**, 34569–34582.

291. Niesteruk,A., Jonker,H.R.A., Richter,C., Linhard,V., Sreeramulu,S. and Schwalbe,H. (2018) The domain architecture of PtkA, the first tyrosine kinase from Mycobacterium tuberculosis, differs from the conventional kinase architecture. *J. Biol. Chem.*, **293**, 11823–11836.

292. Heinzlmeir,S., Kudlinzki,D., Sreeramulu,S., Klaeger,S., Gande,S.L., Linhard,V., Wilhelm,M., Qiao,H., Helm,D., Ruprecht,B., *et al.* (2016) Chemical Proteomics and Structural Biology Define EPHA2 Inhibition by Clinical Kinase Drugs. *ACS Chem. Biol.*, **11**, 3400–3411.

293. Giraud,P., Créchet,J.B., Uzan,M., Bontems,F. and Sizun,C. (2015) Resonance assignment of the ribosome binding domain of E. coli ribosomal protein S1. *Biomol. NMR Assign.*, **9**, 107–111.

294. Salah,P., Bisaglia,M., Aliprandi,P., Uzan,M., Sizun,C. and Bontems,F. (2009) Probing the relationship between gram-negative and gram-positive S1 proteins by sequence analysis. *Nucleic Acids Res.*, **37**, 5578–5588.

295. Yin,Y.W. and Steitz,T.A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science (80-.).*, **298**, 1387–1395.

Chapter 5 Appendix

5.1 Figure and table captions

Figure 1 Project 1 overview: (A) Secondary structures of pilM in ligand-bound conformation. The bound ligands are indicated in blue. The 5´-aptamer strand P pairs with an aptamer-stabilizing strand A to form the binding competent aptamer. The switching strand T and the terminator strand H form the expression platform. (B) and (C) show key constructs in outlining the ligand recognition window through ^{1}H-NMR and ITC titration with c-GAMP. Ligand-containing samples are indicated in blue. (D) Markov model simulations of cotranscriptional folding state distribution over time. State population densities are shown for 100 nM ligand concentration (gray, based on data shown in Figure 13B) and 100 µM ligand concentration (blue, based on data shown in Figure 13C). Areas of the figure where population densities overlap are shown in a gray-blue color. The bar thickness indicates a relative population ranging from 5% to 100%. Transcription intermediates of length from 77-101 nucleotides are binding capable. Figure modified from Landgraf *et al.* (1) and used in Figure 10, Figure 12, and Figure 15. 12

Figure 2 Project 2 overview: (A) Schematic secondary structures of the RNA aptamer targets investigated by 19F-FBS. Stems (P), loops (L), and junctions (J) are annotated, respectively. (B) 19F-1D NMR-spectra of the 19F-library fragment mixtures optimized to avoid signal overlap. (C) Interaction table of all fragments and biological targets screened. Hits were classified as no binding (Qbind > 0.67, alternating gray and white), weak (Qbind=0.66-0.33, yellow), or strong binding (Qbind < 0.32, green). For protein screens, hits for ~5% of the ligands could not be assigned (light blue). (D) Venn diagram of hit overlap between major target classes, RNA, DNA, and Proteins. 1H-1D-NMR-titration (E), 1H, 15N-correlation (F), and 1H,1H-TOCSY (G) of fragment 75 and 76 nt riboswitch aptamer as hit validation. Measurements with fragment are highlighted in blue. Figure modified from Binas et al. (2) and used in Figure 16, Figure 17, Figure 19, Figure 20, and Figure 22. 15

Figure 3 RNA secondary structure elements (12, 17). 27

Figure 4 Schematic representation of the ligand-dependent secondary structure rearrangement of riboswitches. Comparing the ligand bound state (*holo*) and ligand free state (*apo*) of (A) xpt guanine-sensing riboswitch from *B. Subtilis*, (B) 2′dG riboswitch from *M. florum* (favored of two *apo* structures is shown), (C) Cd1 c-di-GMP riboswitch from *C. difficile*, and (D) pilM c-GAMP sensing riboswitch from *G. metallireducens* (1, 31, 53). Structural rearrangements are described by four distinct sequence segments in the mRNA chain: 5′-aptamer strand (P), aptamer-stabilizing strand (A), switching strand (T), and terminator strand (H). Ligand binding junctions are indicated in blue. 33

Figure 5 Previously known tandem architectures for riboswitches and their established functions. (A) Cooperative riboswitch aptamers carry highly similar aptamer domains that bind chemically identical ligands and associate with a single expression platform. Examples of this riboswitch architecture demonstrate cooperative ligand binding and a steeper dose-response curve. (B) Interactive aptamer logic gates are formed by two adjacent aptamers that respond to different target ligands, here depicted as X and Y, and associate with a single expression platform. Ligand binding by one aptamer affects the function of the adjacent aptamer. (C) Pseudo-cooperative and bi-mechanism riboswitches involve the tandem arrangement of independently functioning riboswitches that respond to chemically identical ligands. For a bi-mechanism system, each riboswitch operates with a different regulatory mechanism (e.g. one transcriptional and one translational). (D) Dual riboswitch logic gates involve the tandem arrangement of independently functioning riboswitches that respond to different target ligands. (E) Allosteric ribozyme logic gates involve allosteric regulation by an aptamer for the function of a ribozyme that requires a second distinct compound for its activity. Figure was upscaled, rearranged, and initially published by Sherlock et al. under Creative Commons Attribution License (62). 35

Figure 6 Kinetic simulation of binding and genetic decision using kinetic constants determined at 25 °C. (A) A sample set of data. The y-axis represents the fraction of bound aptamer and the fraction of full-length transcript. The purely kinetic, mixed, and purely thermodynamic zones of character are marked by the light gray [1], white [2], and dark gray [3] background. The partition between zones was set by defining the mixed region from approximately 75-95% of the fraction bound; time domains below and above this zone are designated kinetic and thermodynamic, respectively. (B) Simulated dependence of the extent of adenine binding to the riboswitch at 25 °C using the kinetic parameters reported in Figure 4 of Wickiser *et al.* (57). A 1 nM concentration of RNA is assumed. The different curves correspond to different values of ΔtRNAP, expressed as factors of $1/k_{off}$, where k_{off} is the dissociation rate constant. The maximum value of ΔtRNAP assumed in the simulation is $30/k_{off}$, or about 200 s. Under these conditions binding is at equilibrium, and median response (B50) is equal to the KD. Deviations from equilibrium, particularly at low adenine concentrations, are apparent when $\Delta tRNAP=1/k_{off}$. As ΔtRNAP becomes progressively smaller, B50 moves to higher adenine concentrations. A noteworthy feature of the curves is their asymmetry and the abrupt transition to saturation binding when the system is under kinetic control. Figure reprinted adapted from Wickiser *et al.* (57).Copyright 2005 American Chemical Society. 37

Figure 7 Sensitivity of screening techniques: Detection limits of common screening methods and common hits affinities for 150–250 Da FBS fragments, HTS, and lead compounds. Based on Price *et al.* (119). 51

Figure 8 Basic principle of isothermal titration calorimetry. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment (upper right) with its evaluation (lower right). The titration thermogram is represented as heat per unit of time released after each injection of the ligand into the protein (black), as well as the dilution of ligand into buffer (blue). The dependence of released heat in each injection versus the ratio between total ligand concentration and total protein concentration is represented. Circles represent experimental data and the line corresponds to the best fitting to a model considering n identical and independent sites. Figure was upscaled and initially published by Song *et al.* under Creative Commons Attribution License (145)... 56

Figure 9 Rescuing riboswitch performance in the non-functional regime: (A) Riboswitch functioning through transcriptional termination general response relating ligand concentration [*L*] and regulated protein levels [*P*]. The aptamer is first transcribed in conformation [*B**] and can reversibly bind to form

[BL*] and release ligand before the terminator stem is transcribed (k_E). Terminator stem formation ($k_M=k_{MA}+k_{TA}=k_{MB}+k_{TB}$) occurs much faster than ligand release (k_2') and the progression from conformation *A* to *B* (k_1). Conformation *B* contains a formed aptamer that can bind ligand to form [*BL*]. Green arrow designates mRNA synthesis with biased transcriptional folding, red arrows designate species degradation, and blue arrow designates translation proportional to mRNA levels. Under transcriptional termination, riboswitches effectively choose between termination to form a truncated product [*T*] and extension to form the full-length mRNA [*M*]. (B) The competition between terminator stem formation (k_M) and the progression from conformation *B* to *A* (k_1') determines the dynamic range (η). (C) EC_{50} can be tuned independently from the dynamic range. The accessible range of EC_{50} values is bounded by the aptamer association constant ($K_2=k_2/k_2'$), the rate constant for the progression from conformation *B* to *A* (k_1'), and the rate constant for terminator stem formation (k_M). EC_{50} is tuned over this range by the rate constant representing the delay between aptamer formation and transcription of the terminator stem (k_E). Parameter values: $k_2=10^6\ M^{-1}s^{-1}$; $k_2'=3*10^{-3}\ s^{-1}$; $K_A=k_P*k_{MA}/k_M=10^{-3}\ s^{-1}$; $K_B=k_P*k_{MB}/k_M=10^{-2}\ s^{-1}$; $k_f=10^{-11}\ M^{-1}s^{-1}$; $k_{dP}=10^{-3}\ s^{-1}$; $k_{dM}=10^{-3}\ s^{-1}$; $k_1'+k_M=20\ s^{-1}$. Figure was upscaled, rearranged, and originally published by Beisel and Smolke under Creative Commons Attribution License (75). 62

Figure 10 Nucleotide sequence and secondary structures of full-length pilM and Cd1 riboswitches in their ligand-bound conformation with their cognate ligands c-di-GMP and c-GAMP, respectively. The bound ligands are indicated in blue, and stems are annotated with P1, P2, and P3. The nucleotides of the binding pocket are indicated in violet. The strands that contribute to the riboswitch function are color-coded in the PATH nomenclature in P(blue), A(orange), T(red), and H(green). The 5´-aptamer strand P pairs with an aptamer-stabilizing strand A to form the binding competent aptamer. The switching strand T and the terminator strand H are located downstream on the mRNA. For Cd1, the terminator conformation is formed through interactions of strands T+H and antiterminator through T+A+H. For pilM, the terminator conformation is formed through interactions of strands T+A+H and antiterminator through P+A. Additionally, the binding pockets are marked with a gray background. Figure as published in Landgraf *et al.* (1). 66

Figure 11 Imino region of ^{1}H/^{1}H NOESY spectrum for pilM^{84} in the *apo* state (A), *holo* state (B), and double truncated pilM$^{14-75}$ (C). (D) shows an overlay of the pilM^{101} *apo* (black) and a *holo* (blue) spectrum with four additional Us resulting from c-GAMP addition. (E) shows the pilM secondary structure indicating the key last residues and the start residue of pilM^{84}, pilM ^{101}, and pilM$^{14-75}$. The assignment walks of (A), (B), and (C) are indicated in the same color scheme as was used for residues in the secondary structures in (E). Figure as published in Landgraf *et al.* (1). 70

Figure 12 (A) and (C) show the NMR reporter signal regions of a ^{1}H-NMR titration for c-GAMP-binding to pilM-RNAs ending on nucleotides 75, 77, 79, 84, 101, 102, and 109 and for c-di-GMP binding to Cd1-RNAs ending on nucleotides 82, 83, 86, and 87. Indicated in blue are the spectra of metabolite-containing samples and in black samples without. Complete NMR spectra of the shown datasets and all other lengths measured for pilM are shown in SI Figure 2. The full peak lists can be found in SI Table 2 and Landgraf *et al*. (1). (B), (D), and (E) show the secondary structure models for ligand recognition. Indicated in violet are the key nucleotides of the binding pocket, and highlighted in red are the bases contributing to the reporter signals. Sections (F) and (G) show examples of the ITC titration curves of key constructs in outlining the ligand recognition window. Values and errors in the K_D are derived from triplicate measurements. Error bars indicated in the ITC plots refer to the individual fit. Figure as published in Landgraf *et al.* (1). 71

Figure 13 Contour level plots of cotranscriptional folding states for the Cd1 and pilM riboswitch. The population of the three functional relevant states over time is depicted in 5% levels of relative population. The population of the macrostates is the sum of all contributing RNA lengths in the corresponding state. The conditions of the corresponding simulation were 20 nt/s transcription speed, no pausing, 400 s^{-1} base pair closing rate, and either 100 nM or 100 µM cyclic-di-nucleotide (CDN), c-di-GMP for Cd1 and c-GAMP for pilM. The second variables in the contour level plots were (A) the concentration of CDN, (B) the transcription speed at low ligand concentration, (C) the transcription speed at high ligand concentration (simulated from 2 to 100 nt/s),(D) the base pair closing rate at low ligand concentration (simulated from 2 to 100 nt/s),(E) the base pair closing rate at high ligand concentration and (F) the transcription speed at high ligand concentration with 10 s pausing at the possible pause site at nucleotides 141-145 (simulated from 3 to 100 nt/s) (F). Figure as published in Landgraf et al. (1). 77

Figure 14 Kinetic model of binding to pilM (A) and Cd1 (B) during regulatory window assuming the same transcription speed (20 nt/s), association rate (17500 $M^{-1}s^{-1}$), and regulatory CDN concentration (100 µM). Fraction of ON signaling is indicated by blue line. The difference in the size of the regulatory window is caused by pilM having 60% fewer binding competent transcriptional intermediates. RNA states associated with ON signaling are marked in blue-shaded areas. RNA states associated with OFF signaling are marked in gray-shaded areas. Fractions of leaking, false positive, and false negative signaling are marked on the two y-axis representing the the lower and upper limit of regulating CDN concentration. The left limit represents $t_{transcription}<<t_{on}$ and is equivalent to low CDN concentration (100 nM) OFF signaling for pilM and ON signaling for Cd1. The right limit represents $t_{transcription}>t_{on}$ for pilM and $t_{transcription}>>t_{on}$ for Cd1, this is equivalant to high CDN concentration (100 µM) ON signaling for pilM and OFF signaling for Cd1. Dashed lines indicate median response (T50), upper limit of kinetic control character, and lower limit of thermodynamic control character as introduced by Wickiser et al. (55). (C) depicts the qualitative effect of regulatory variables on the size of the regulatory window. The size decreases with transcription speed, increases with sequence length, ligand concentration, association rate, and transcription time, and strongly increases with pausing delays. 83

Figure 15 Markov model simulations of cotranscriptional folding state distribution over time for the pilM (ON) and Cd1 (OFF) riboswitch at 20 nt/s transcription speed with no pausing. State population densities are shown for 100 nM ligand concentration (gray, based on data shown in Figure 13B) and 100 µM ligand concentration (blue, based on data shown in Figure 13C). Areas of the Figure where population densities overlap are shown in a gray-blue color. The thickness of the bar is indicative of a relative population ranging from 5% to 100%. Population densities are also shown in percent at 0.5 s increments above and below the population bars if the changes by more than 5% in the timeframe. Population densities below 1% are marked explicitly. The aptamer domain is formed with nucleotides 76 and 86 when the ligand can first be bound. Upon reaching nucleotides 77 and 87, the aptamer domain can bind ligands and transition up into the holo state. Transcription intermediates of length from 77-101 nucleotides for pilM and 87-147 nucleotides for Cd1 are binding capable. The time frames available for these transitions are indicated as state transitions. PilM can fold the terminator from nucleotide 92 onward, and Cd1 can fold the antiterminator starting with nucleotide 134. Both RNAs transition between the three states until reaching a point of decision with transcription lengths 102 and 148 when the binding is impaired. Figure as shown in (1). 87

Figure 16 Overview of RNA targets. Schematic secondary structures of the RNA targets investigated by 19F-FBS. Stems (P), loops (L), and junctions (J) are annotated, respectively. Tri-, tetra-, and pentaloop sequences are listed explicitly. Figure as published in Binas *et al.* (2)............................... 90

Figure 17 ^{19}F-1D NMR-spectra of the ^{19}F-library fragment mixtures. The ^{19}F-library contains 101 compounds (SI Table 4). Five mixtures of either 20 or 21 ligands were generated to avoid signal overlap. The spectra of the mixtures (A, B, C, D, E) in the screening buffer are displayed. Figure as published in Binas et al. (2).. 91

Figure 18 Determination of Q^{bind}. Four ^{19}F-CPMG experiments are recorded to determine the binding factor Q^{bind} from peak integrals, as discussed in the main text. The relaxation loss at 200 ms relaxation dephasing time relative to 0 ms dephasing for the ^{19}F signal of the ligand is recorded in the presence and absence of a biomolecular target. Figure as published in Binas et al. (2). 95

Figure 19 Interaction table of all fragments and biological targets screened. Hits were classified into no binding (Qbind > 0.67, alternating gray and white), weak (Qbind=0.66-0.33, yellow), or strong binding (Qbind < 0.32, green) in 19F-CPMG experiments. For protein screens, hits for ~5% of the ligands could not unambiguously be assigned (light blue). Figure as published in Binas et al. (2)...... 97

Figure 20 Hit validation and competition experiments: Validation of ^{19}F-CPMG screening hits for the aptamer domains of the three secondary messenger-sensing riboswitches. a) Spectral regions with signals from guanosine (top) and uridine (bottom) residues of the ^{1}H, ^{15}N-correlation experiment of the 76 nt riboswitch with (blue) and without 75 (black) shown under c. b) ^{1}H,^{1}H-TOCSY spectrum with (blue) and without 75 (black). c) ^{1}H-1D-NMR-titration of 75 with the RNA. K_D was determined, according to Williamson (250). d) (Partially) competitive binding of fragments to the 84 nt and 98 nt riboswitch observed in T_2-modulated 1D-^{19}F experiments. Figure as published in Binas et al. (2)...... 99

Figure 21 Correlation matrix of hit clusters, displaying hit correlation between different targets screened by ^{19}F-FBS. Figure as published in Binas et al. (2)... 100

Figure 22 Cheminformatic analysis of hit data for all RNA, DNA, and Protein biomolecules. a) Gaussian distributions for aromatic atoms and SP3 descriptor over categories of biomolecules. SP3 descriptor (sp3 carbon atom count/total carbon atom count) reflects the flatness of the fragment molecules. b) Visualization of classes in a Venn diagram. c) Euclidian distribution of hits to the target biomolecules. (modified (added a percent sign) Figure as published in Binas et al. (2). 101

Figure 23 Fluorescence-based determination of affinity of acridine and coupled acridine-benzamide (P2D11) derivative to a) SAM-Antiterminator, b) A-terminator, and c) 2′dG-Terminator. Figure modified from Binas *et al.* (2). 103

Figure 24 Regulatory understanding gained by the analysis of the regulatory window of CDN riboswitches: Contour plots of transcription speed dependence of the predominate state of pilM and Cd1 at 100 nM (A) and 100 µM CDN (E). Contour lines indicate 5% levels of relative population increasing from violet, blue, green, yellow, to red. At low CDN concentrations, pilM forms terminator and Cd1 forms antiterminator, and at high CDN concentrations, the *holo* state is formed, inverting the switch behavior during the regulatory window as shown in (C). PilM surpasses the limit for pure kinetic control, and Cd1 operates at near saturation with thermodynamic control character. Leaking/false positive signaling occurs only during OFF signaling and is indicated in green. The secondary structures of the predominate states are depicted in (B) and (D). Nucleotides that are transcribed during the regulatory window are marked with dashed lines. Figure modified from Landgraf et al. (1) and used in Figure 13, Figure 14, and Figure 15. 110

Figure 25 Computational analyses, competition experiments, and follow-up chemistry enabled by ^{19}F-NMR FBS: Cheminformatic analysis of hit data for all RNA, DNA, and Protein biomolecules. (A) Visualization of major classes in a Venn diagram. (B) Euclidian distribution of hits by target. (C) (Partially) competitive binding of fragments 9, 99, and 75 to the 84 nt riboswitch aptamer observed in T_2-modulated 1D-^{19}F experiments. Measurements with target and fragment are shown as black (0 ms) and blue (200 ms) lines and after the addition of c-GAMP with green (0 ms) and orange (200 ms) lines. (D) Fluorescence-based titration of acridine and coupled acridine-benzamide (P2D11) derivative to SAM-Antiterminator. Figure modified from Binas et al. (2) and used in Figure 20, Figure 22, and Figure 23 111

Table 1 Published kinetic values of various riboswitches. *values were calculated based on corresponding values provided in the related sources. 41

Table 2 K_D determination by ITC for ligand binding to riboswitch RNA transcripts of different lengths. Values were measured as triplicates. Errors represent the standard deviation of the determined triplicate values. Sequence-aligned constructs are shown in the same line. *: values determined after the publication of Landgraf *et al.* (1). 72

Table 3 Ligand concentration values associated with median ligand response, kinetic control, and thermodynamic control. Values assume no *holo*-complex dissociation, no formation of aptamer competing fold during the transcription time interval, and a large excess of ligand. A transcription speed of 20 nt/s and a k_{on} rate of 17500 $M^{-1}s^{-1}$ were assumed for both aptamers. Threshold values of kinetic and thermodynamic control were applied based on the limits Wickiser *et al*. set (57). 83

SI Figure 1: 10% denaturing polyacrylamide gel-electrophoresis with 7M urea of pilM14-75 and pilM RNAs 75-109 and at ~30°C. Gel was run for 0.5 hours at 200-240V using 1x TBE-buffer as running buffer. Samples contained 0.5µL of a 1/100 dilution of the final RNA stock, formaldehyde, xylene cyanol, and bromophenol blue. Figure as published in Landgraf *et al.* (1). 145

SI Figure 2: Full spectral width 1D spectra of different pilM transcript lengths (A) 1H-1D jump-and-return echo NMR spectra of pilM75 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (B) 1H-1D jump-and-return echo NMR spectra of pilM77 at a concentration of 75 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 2048 scans, and 24.5 ppm spectral width at 600 MHz. (C) 1H-1D jump-and-return echo NMR spectra of pilM79 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (D) 1H-1D jump-and-return echo NMR spectra of pilM80 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (E) 1H-1D jump-and-return echo NMR spectra of pilM81 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (F) 1H-1D jump-and-return echo NMR spectra of pilM82 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (G) 1H-1D jump-and-return echo NMR spectra of pilM83 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (H) 1H-1D jump-and-return echo NMR spectra of pilM84 at a concentration of 100 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 65k time domain points, 256 scans, and 25 ppm spectral width at 600 MHz. (I) 1H-1D jump-and-return echo NMR spectra of pilM88 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. (J) 1H-1D jump-and-return echo NMR spectra of pilM91 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (K) 1H-1D jump-and-return echo NMR spectra of pilM93 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (L) 1H-1D jump-and-return echo NMR spectra of pilM95 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (M) 1H-1D jump-and-return echo NMR spectra of pilM96 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (N) 1H-1D jump-and-return echo NMR spectra of pilM97 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (O) 1H-1D jump-and-return echo NMR spectra of pilM99 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (P) 1H-1D jump-and-return echo NMR spectra of pilM100 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (Q) 1H-1D jump-and-return echo NMR spectra of pilM101 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (R) 1H-1D jump-and-return echo NMR spectra of pilM102 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (S) 1H-1D jump-and-return echo NMR spectra of pilM103 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 512 scans, and 24 ppm spectral width at 950 MHz. (T) 1H-1D jump-and-return echo NMR spectra of pilM104 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (U) 1H-1D jump-and-return echo NMR spectra of pilM105 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. (V) 1H-1D jump-and-return echo NMR spectra of pilM109 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. Figure as published in Landgraf *et al.* (1). 153

SI Figure 3: ITC raw data for the pilM riboswitch: The raw thermogram and the original fit in Sedphat are shown. The pilM-109 measurement was carried out at 50 µM RNA concentration. All others were conducted at 60 µM RNA concentration. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM $MgCl_2$. Figure as published in Landgraf *et al.* (1). 162

SI Figure 4: Kinetic information obtained for strongly binding pilM and Cd1 riboswitch Constructs using kinITC. Figure as published in Landgraf *et al.* (1). 162

SI Figure 5: Graphical description of the three state model used in the Markov modeling. The blue numbers refer to the RNA length ranges for which the indicated rates were applied. Low and high designate which rate was used for low or for high ligand concentration simulations (100 µM to 100 nM), respectively. Figure as published in Landgraf *et al.* (1). 164

SI Figure 6: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the pilM 3′,3′-cGAMP-sensing riboswitch (84 nt). For further hit validation, the CPMG experiment at 400 ms is shown. Figure as published in Binas *et al.*. Hit conformation plots for all other targets can be found in the SI of Binas *et al.* (2). 185

SI Table 1: PCR primers for Cd1 and pilM. Primers are named as the corresponding construct and market, with fwd for the forward primer and rev for the reverse primer. Nucleotides marked with [ ] are 2´ methoxy modified. Table as published in Landgraf *et al.* (1). 143

SI Table 2 Peak List of the resonance assignments for the apo form of pilM Riboswitch found in Figure 11A, B, and C. Table as published in Landgraf *et al.* (1). 154

SI Table 3: List of all biomolecules used in the study listed with their biological host organism (if applicable), PDB accession codes of X-ray structures, and primary publication. Table as published in Binas *et al.* (2). 173

SI Table 4: ^{19}F-library for the ^{19}F-NMR-based fragment screening. Table as published in Binas *et al.* (2). 175

5.2 Methods and supplementary information

DNA plasmids from eurofins® were sourced with a cloned-in sequence containing full-length native riboswitch sequence of the CD-1 and pilM riboswitch preceded by the T7 promotor and a Hammerhead ribozyme and followed by an HDV ribozyme (SI Information 1). The plasmids were transformed into *E.coli* DH5α cells through heat shock, expanded in LB medium, and purified using QIAGEN® plasmid DNA purification kits. For subsequent RNA transcriptions, the plasmids were PCR-amplified to minimize DNA content. The reverse primers were 2′-methoxy modified at the two last residues of their 5′-end to increase the 3′-end homogeneity of the transcribed RNA. The 2′-OMe-modified nucleotides are marked with square brackets in SI Table 1. The forward primers consisted of the noncoding part T7 RNA polymerase promoter followed by the 5′-end of the riboswitch sequence. To assure primer specificity, a melting temperature of 60°C was chosen, and the length of overlapping sequences were adjusted accordingly. The PCR was carried out following the protocol by New England Biolabs® (0.5 µM of each primer, 200 µM dNTPs, 10 ng plasmid) using Phusion polymerase. PCR product length and homogeneity were confirmed through denaturing PAGE.

5.2.1 DNA templates

SI Information 1: DNA template plasmids used for PCRs. The DNA sequences were cloned into a Puc57 vector between EcoRI and SmaI. Information as published in Landgraf *et al.* (1).

Cd1 sequence:

5´-TAATACGACTCACTATAGGGAGACCTATTTAGTTTCTGATGAGAGCGAAAGCTCGAAACAGCTGTGAAGCTGTCAAACTAAATAGGCAAATCTAGAGAAATCTAGTGACGCAAAGCTATAGGGACTAAGGTTTATATACATAAACTATGTCAGCCAGTTGCCAAAAAGAGTCCTAGGTGTATTGTATACCTAAGAAAAGTCTATAATGACTGCCTTTTTGGCAGTCATTTTGTT-3´

pilM sequence:

5´-TAATACGACTCACTATAGGGAGATCGGAGCTGATGAGAGCGAAAGCTCGAAACAGCTGTGAAGCTGTCCTCCGATATCGACAATACTAAACCATCCGCGAGGGTGGGACGGAAAGCCTACAGGGTCTCTCTGAGACAGCCGGGATGCCGAAATATCACAATTCGTGATGCTCGGTCCCGGCATTTCTTTTTGGGTCGGCATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCTACTTCGGTAGGCTTAAGGGAGAAG-3´

T7-Promotor, 5´-Hammerhead ribozyme, riboswitch sequence, 3´-HDV ribozyme

5.2.2 PCR primers

SI Table 1: PCR primers for Cd1 and pilM**.** Primers are named as the corresponding construct and market, with fwd for the forward primer and rev for the reverse primer. Nucleotides marked with [ ] are 2´ methoxy modified. Table as published in Landgraf *et al.* (1).

Primer name	**Primer sequence**
$\text{Cd1}^{1}_{\text{fwd}}$	5´- TAATACGACTCACTATAGGAAACTAAATAGGCAAATCTAGA-3´
$\text{Cd1}^{80}_{\text{rev}}$	5´- [GC]TGACATAGTTTATGTATAT-3´
$\text{Cd1}^{81}_{\text{rev}}$	5´- [GG]CTGACATAGTTTATGTATA-3´
$\text{Cd1}^{82}_{\text{rev}}$	5´-[UG]GCTGACATAGTTTATGTAT-3´
$\text{Cd1}^{83}_{\text{rev}}$	5´-[CU]GGCTGACATAGTTTATGTA-3´
$\text{Cd1}^{86}_{\text{rev}}$	5´-[CA]ACTGGCTGACATAGTTTAT-3´
$\text{Cd1}^{87}_{\text{rev}}$	5´-[GC]AACTGGCTGACATAGTTTA-3´
$\text{Cd1}^{88}_{\text{rev}}$	5´-[GG]CAACTGGCTGACATAGTTT-3´
$\text{Cd1}^{132}_{\text{rev}}$	5´-[UU]ATAGACTTTTCTTAGGTATA-3´
$\text{Cd1}^{143}_{\text{rev}}$	5´-[AA]AGGCAGTCATTATAGACTTTTCT-3´
$\text{Cd1}^{146}_{\text{rev}}$	5´-[CA]AAAAGGCAGTCATTATAGACTT-3´
$\text{Cd1}^{147}_{\text{rev}}$	5´-[CC]AAAAAGGCAGTCATTATAGACT-3´
$\text{Cd1}^{148}_{\text{rev}}$	5´-[GC]CAAAAAGGCAGTCATTATAGAC-3´
$\text{Cd1}^{149}_{\text{rev}}$	5´-[UG]CCAAAAAGGCAGTCATTATAGA-3´
$\text{Cd1}^{150}_{\text{rev}}$	5´-[CU]GCCAAAAAGGCAGTCATTATAG-3´
$\text{Cd1}^{151}_{\text{rev}}$	5´-[AC]TGCCAAAAAGGCAGTCATTATA-3´
$\text{pilM}^{1}_{\text{fwd}}$	5´- TAATACGACTCACTATAGATATCGACAATACTAAACCATCC-3´
$\text{pilM}^{14}_{\text{fwd}}$	5´- TAATACGACTCACTATAGCTAAACCATCCGCGAGGGTGGGACGG-3´
$\text{pilM}^{75}_{\text{rev}}$	5´-[GC]ATCCCGGCTGTCTCAG-3´
$\text{pilM}^{77}_{\text{rev}}$	5´-[CG]GCATCCCGGCTGTCTC-3´
$\text{pilM}^{79}_{\text{rev}}$	5´-[UU]CGGCATCCCGGCTGTC-3´
$\text{pilM}^{80}_{\text{rev}}$	5´-[UU]TCGGCATCCCGGCTGTC-3´
$\text{pilM}^{81}_{\text{rev}}$	5´-[AU]TTCGGCATCCCGGCTGTC-3´
$\text{pilM}^{82}_{\text{rev}}$	5´-[UA]TTTCGGCATCCCGGCTG-3´
$\text{pilM}^{83}_{\text{rev}}$	5´-[AU]ATTTCGGCATCCCGGCTG-3´
$\text{pilM}^{84}_{\text{rev}}$	5´-[GA]TATTTCGGCATCCCG-3´
$\text{pilM}^{88}_{\text{rev}}$	5´-[TT]GTGATATTTCGGCATCCCG-3´
$\text{pilM}^{91}_{\text{rev}}$	5´-[GA]ATTGTGATATTTCGGCATCCC-3´
$\text{pilM}^{93}_{\text{rev}}$	5´-[AC]GAATTGTGATATTTCGGCATCC-3´
$\text{pilM}^{95}_{\text{rev}}$	5´-[TC]ACGAATTGTGATATTTCGGCATC-3´
$\text{pilM}^{96}_{\text{rev}}$	5´-[AT]CACGAATTGTGATATTTCGGCATC-3´
$\text{pilM}^{97}_{\text{rev}}$	5´-[CA]TCACGAATTGTGATATTTCGGC-3´
$\text{pilM}^{99}_{\text{rev}}$	5´-[AG]CATCACGAATTGTGATATTTCG-3´
$\text{pilM}^{100}_{\text{rev}}$	5´-[GA]GCATCACGAATTGTGATATTTC-3´
$\text{pilM}^{101}_{\text{rev}}$	5´-[CG]AGCATCACGAATTGTG-3´
$\text{pilM}^{102}_{\text{rev}}$	5´-[CC]GAGCATCACGAATTGTGATAT-3´
$\text{pilM}^{103}_{\text{rev}}$	5´-[AC]CGAGCATCACGAATTGTGAT-3´
$\text{pilM}^{104}_{\text{rev}}$	5´-[GA]CCGAGCATCACGAATTGTG-3´
$\text{pilM}^{105}_{\text{rev}}$	5´-[GG]ACCGAGCATCACGAATTG-3´
$\text{pilM}^{109}_{\text{rev}}$	5´-[GC]CGGGACCGAGCATCAC-3´

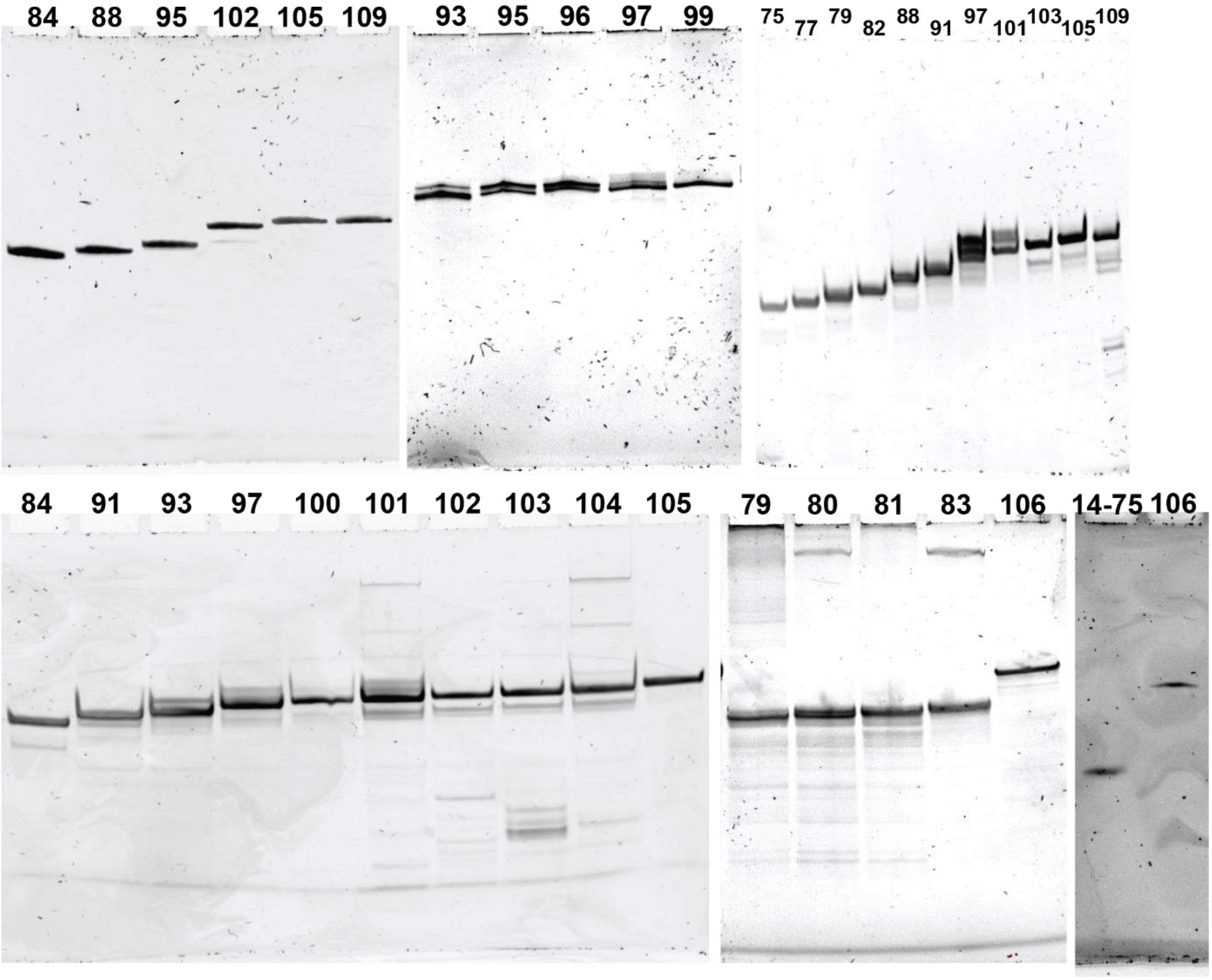
5.2.3 RNA and CDN preparation

In vitro transcriptions were conducted using T7 RNA polymerase. DNA templates were freshly prepared using PCR. The PCR mixtures were added to the transcription mixture without further purification. The transcription mixtures contained transcription buffer (250 mM Tris-HCl pH 8.1), 2 mM spermidine, 20 mM dithiothreitol (DTT), 20% (v/v) DMSO, NTPs were added in sequence adjusted amounts at a total concentration of 10 mM to 15 mM, 5 mM to 27.5 mM $Mg(OAc)_2$, 1 µ/mL yeast inorganic pyrophosphatase (New England Biolabs) and 40 µg/mL T7 RNA polymerase (produced by various members of the Schwalbe group). The NTP to $Mg(OAc)_2$ ratio was optimized before upscaling batch size 100 fold to preparative transcriptions volumes, 10 mM NTP to 15 mM $Mg(OAc)_2$ was frequently found to be an ideal ratio. NTPs were purchased from different providers, Carl Roth GmbH + Co. KG (Germany) and Silantes GmbH (Germany), for isotopically labeled NTPs.

The transcribed RNA was washed in accordance with the protocol developed by Helmling *et al*. (84). The transcription mixtures were washed several times in a 3-10 kDa molecular weight cut-off centrifugal concentrators (Vivaspin 20 ® from Sartorius AG, Germany) using NMR buffer (50 mM Bis-Tris buffer at pH 6.1, 50 mM KCl, and 5 mM $MgCl_2$). Washing was carried out by reducing the volume 5 fold and adding 4 new parts of 1xNMR buffer. This procedure was repeated until the concentrator follow-through had less than 1 OD. The samples were now further concentrated to a volume of about 200 µL volume. The final stock concentration was determined through UV-vis extinction on a Nanodrop (ThermoFisher), and nearest-neighbor corrected extinction coefficients. RNA length and homogeneity were confirmed through denaturing PAGE (SI Figure 1).

C-GAMP and c-di-GMP were purchased from Sigma Aldrich, and 15N labeled c-GAMP was produced through cyclization from labeled nucleotides following the protocol by Kato et al. (258).

5.2.4 PAGE analyses

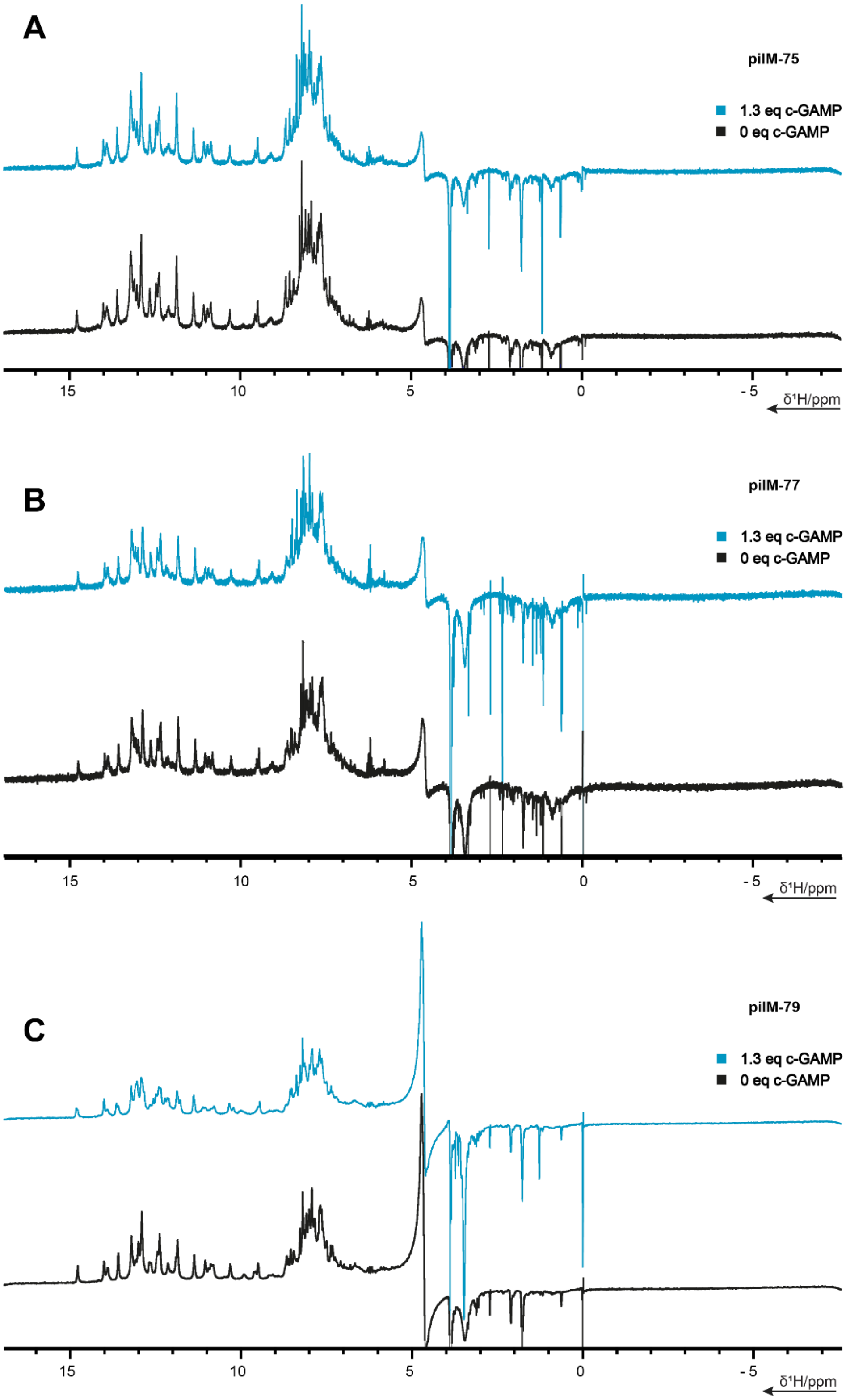


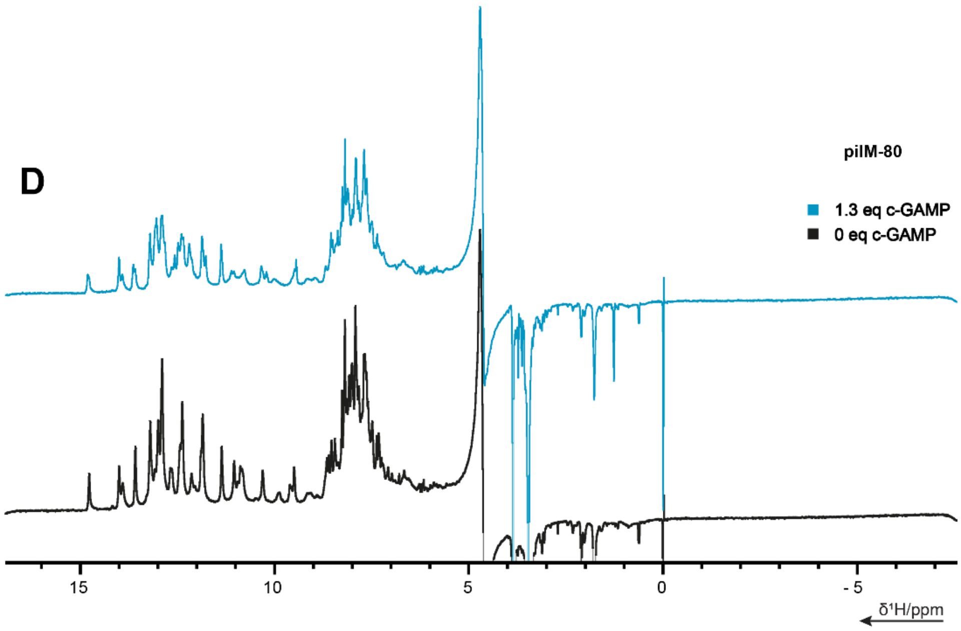
SI Figure 1: 10% denaturing polyacrylamide gel-electrophoresis with 7M urea of pilM14-75 and pilM RNAs 75-109 and at ~30°C. Gel was run for 0.5 hours at 200-240V using 1x TBE-buffer as running buffer. Samples contained 0.5µL of a 1/100 dilution of the final RNA stock, formaldehyde, xylene cyanol, and bromophenol blue. Figure as published in Landgraf *et al.* (1).

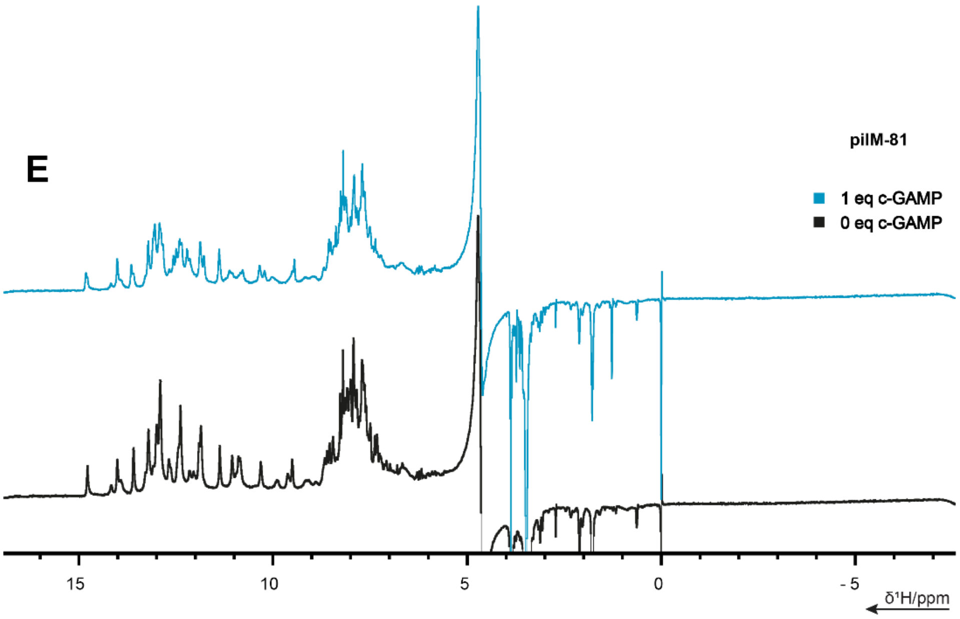
5.2.5 NMR spectroscopy

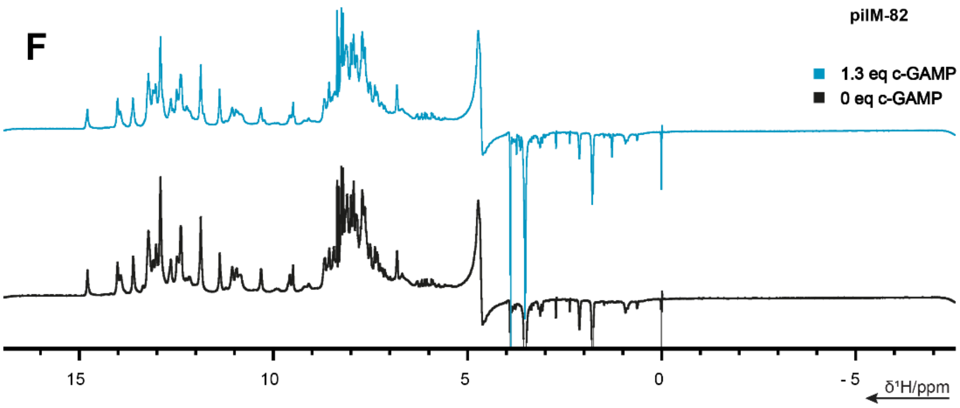
NMR-samples of 180-280 µL volume contained 100-1600 µM RNA, 50 mM Bis-Tris buffer (6.1 pH), 50 mM KCl, 5 mM MgCl2, 5-10% v/v D2O. DSS or Dioxan (for ligand stocks) were added as internal reference in some measurements. All measurements were conducted in Shigemi NMR tubes (Shigemi Inc.) or 3 mm short automation NMR tubes (Bruker). NMR experiments were conducted in cryogenic probe equipped Bruker AV600, AV700, AV800, AV900, and AV950 spectrometers. NMR data were processed with Bruker Topspin 3.5 or later (Bruker Biospin) and Sparky 3.14 (Figure 11, SI Figure 2, and SI Table 2). In 1H 1D and 1H/1H NOESY spectra, water suppression was achieved using jump-and-return echo pulse schemes (259).

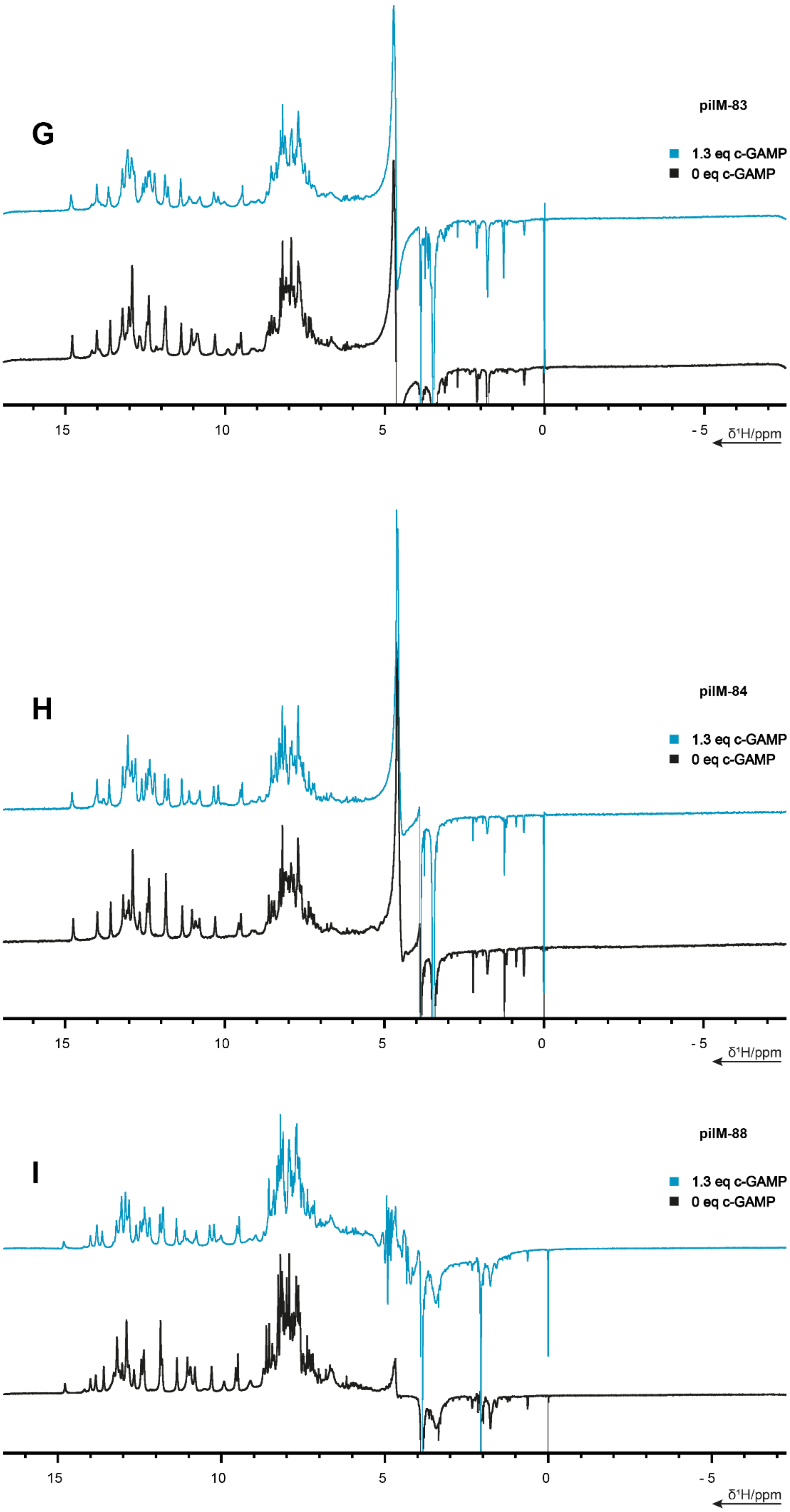
5.2.6 NMR data

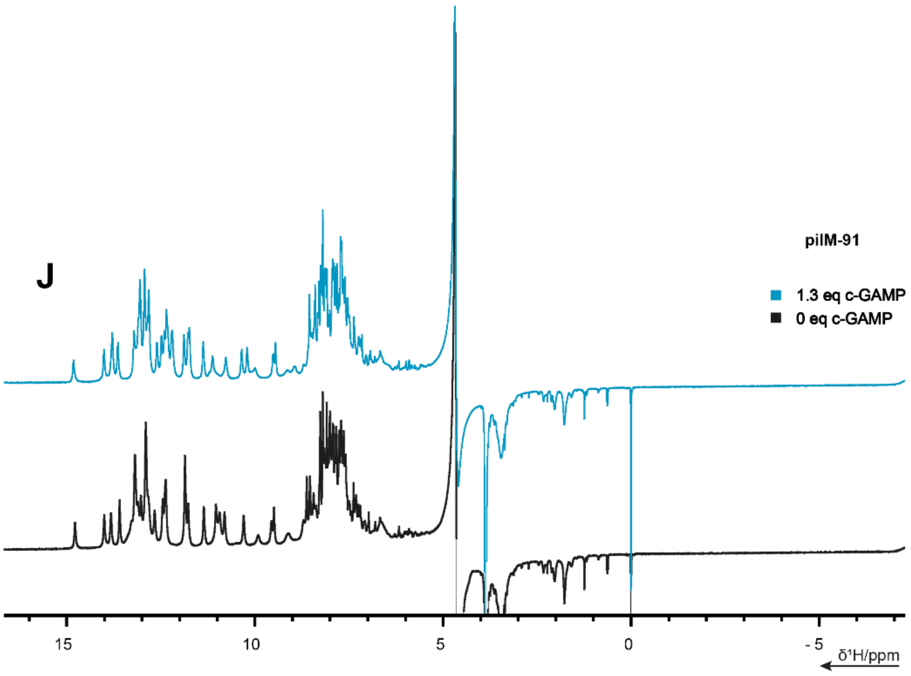

A
pilM-75
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm
B
pilM-77
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm
C
pilM-79
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

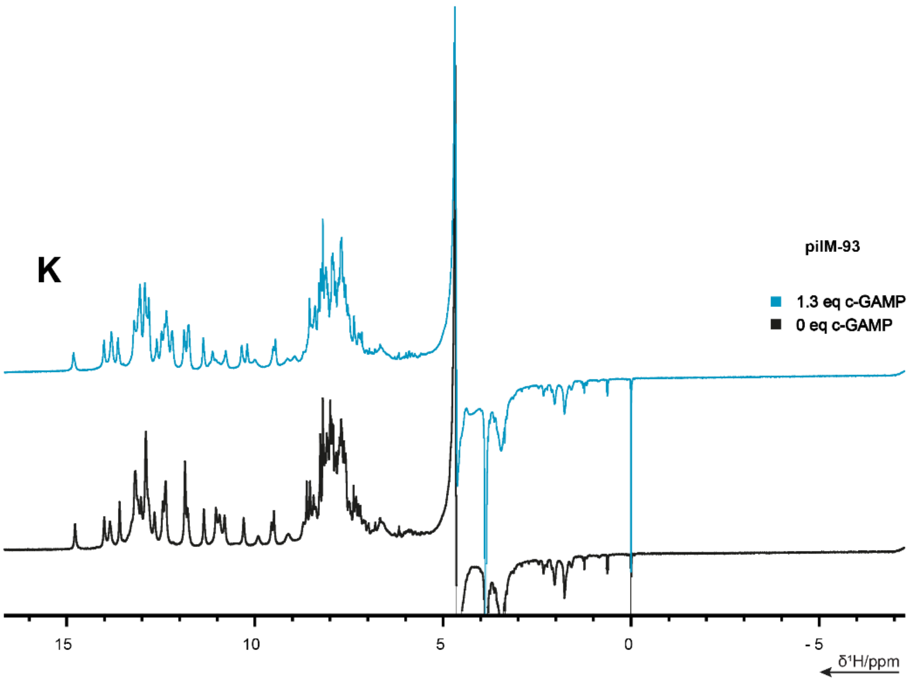

D
pilM-80
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

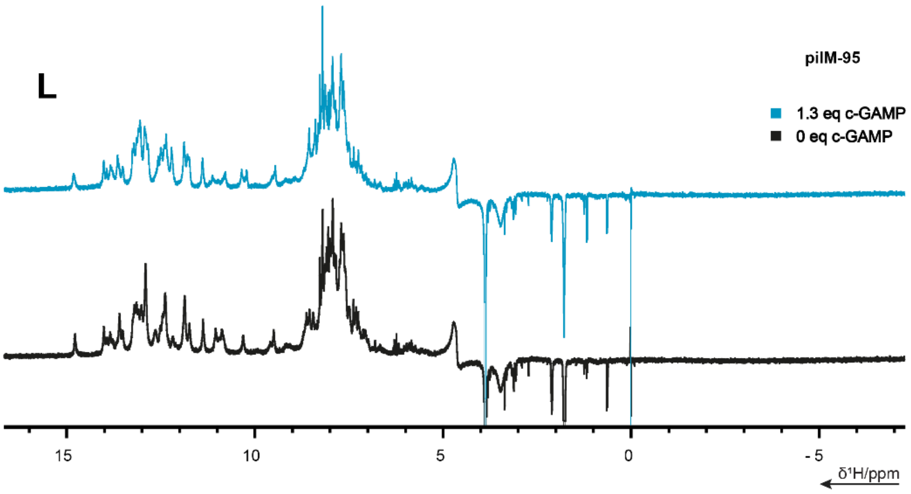

E
pilM-81
■ 1 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

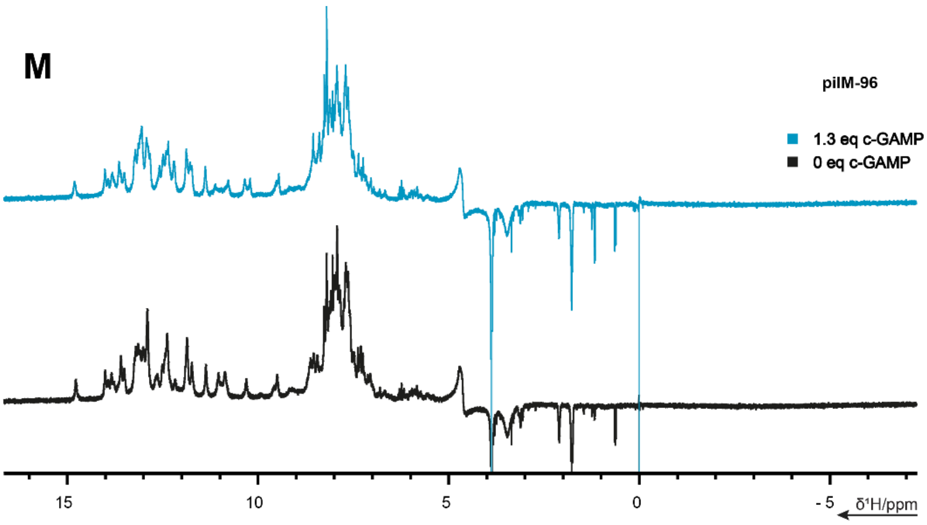

F
pilM-82
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

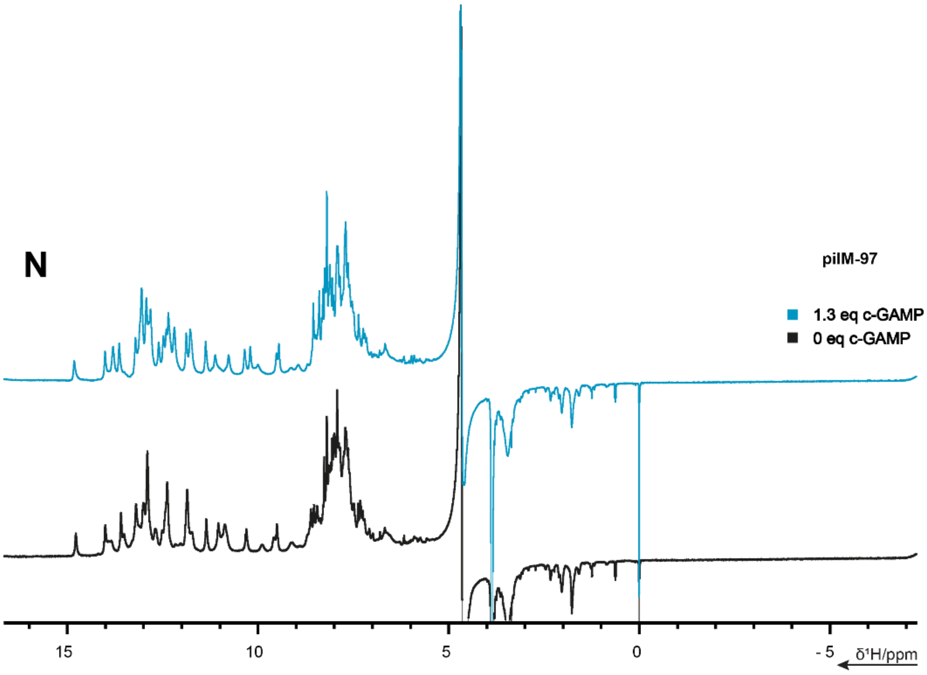

G
pilM-83
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm
H
pilM-84
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm
I
pilM-88
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

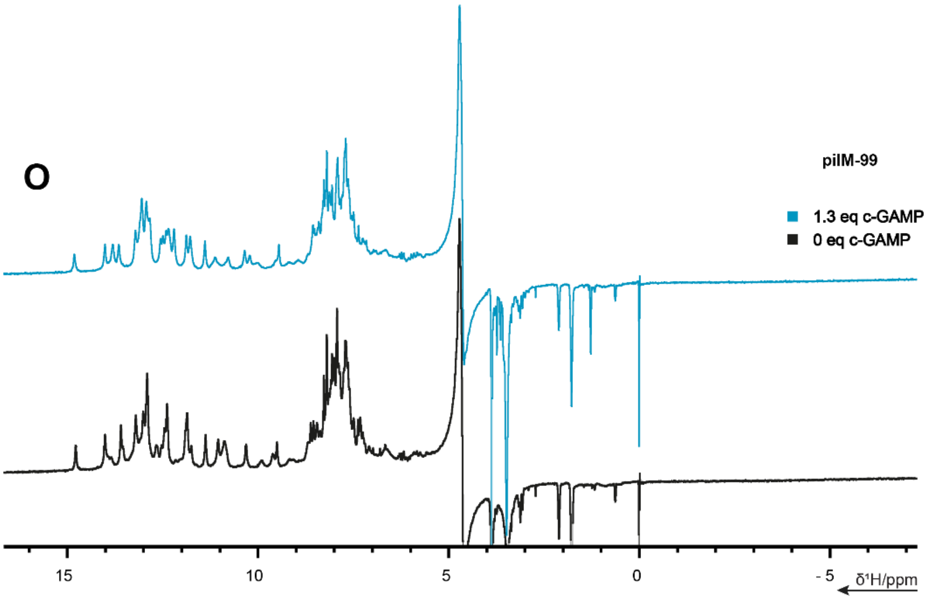

J
pilM-91
🟦 1.3 eq c-GAMP
⬛ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

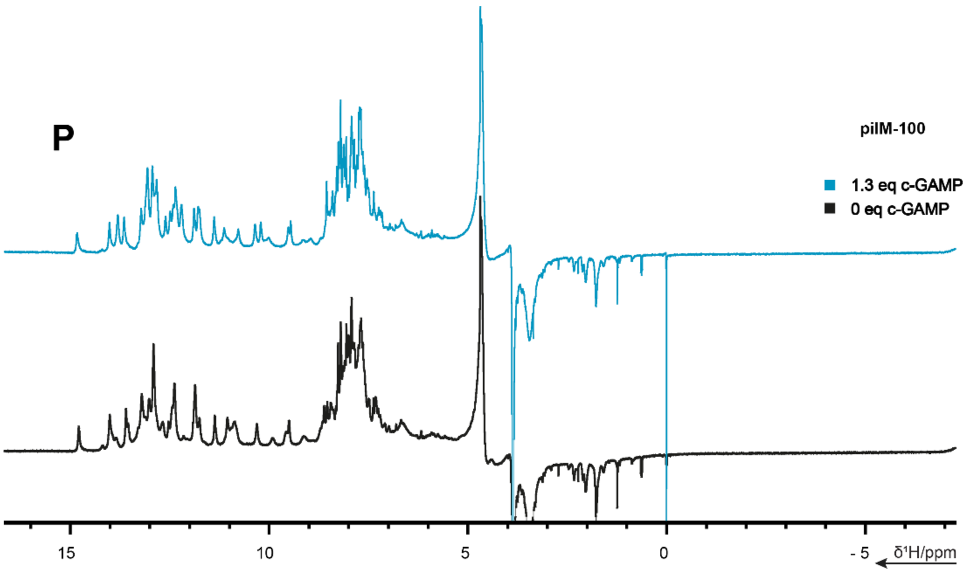

K
pilM-93
🟦 1.3 eq c-GAMP
⬛ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

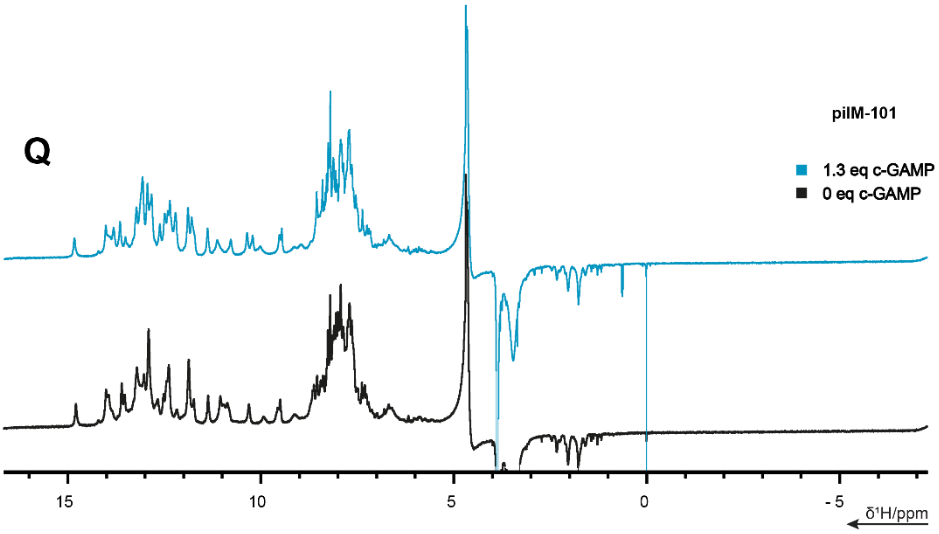

L
pilM-95
🟦 1.3 eq c-GAMP
⬛ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

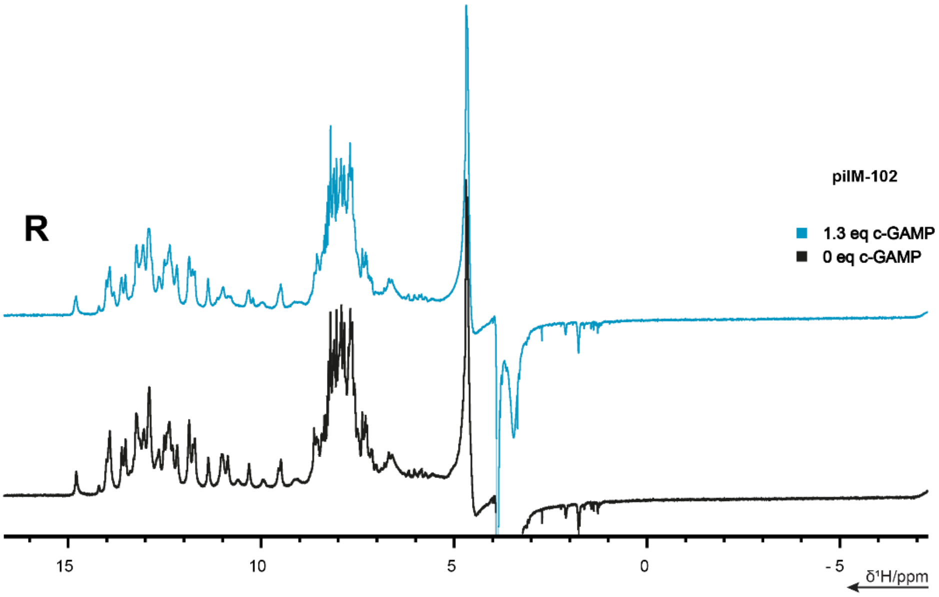

M
pilM-96
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

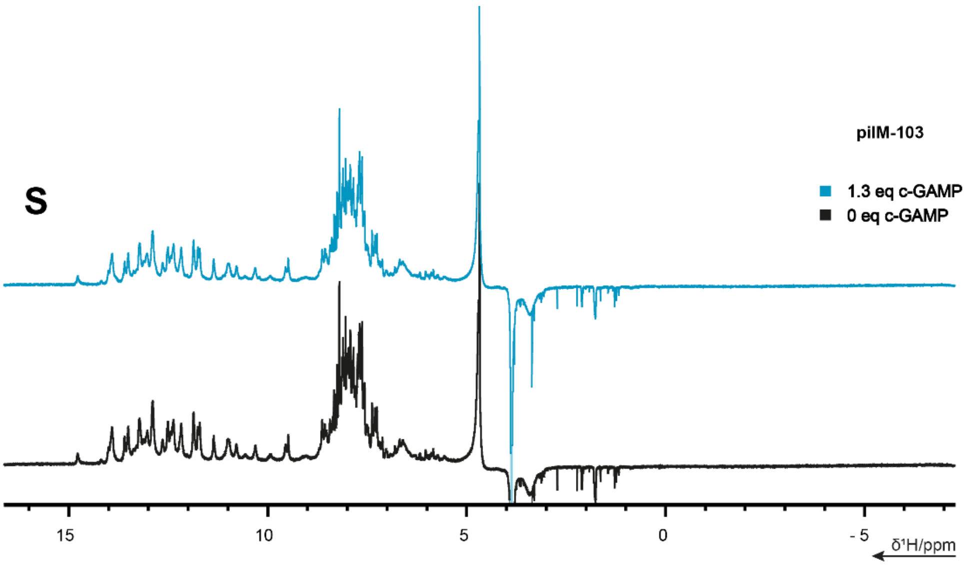

N
pilM-97
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

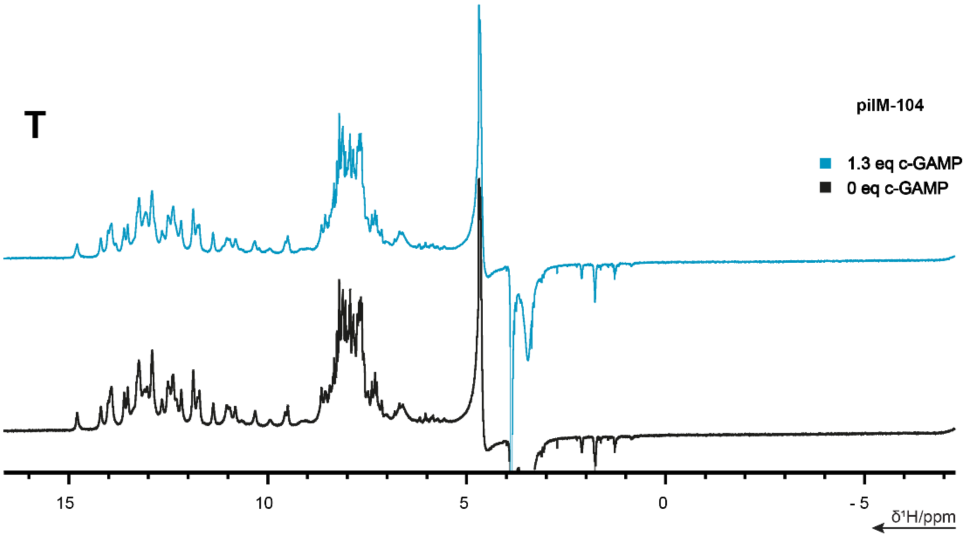

O
pilM-99
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

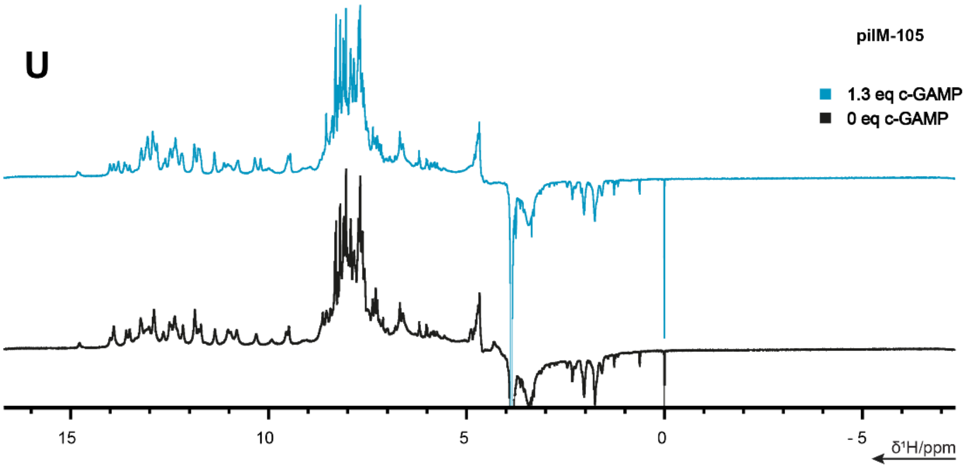

P
pilM-100
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm

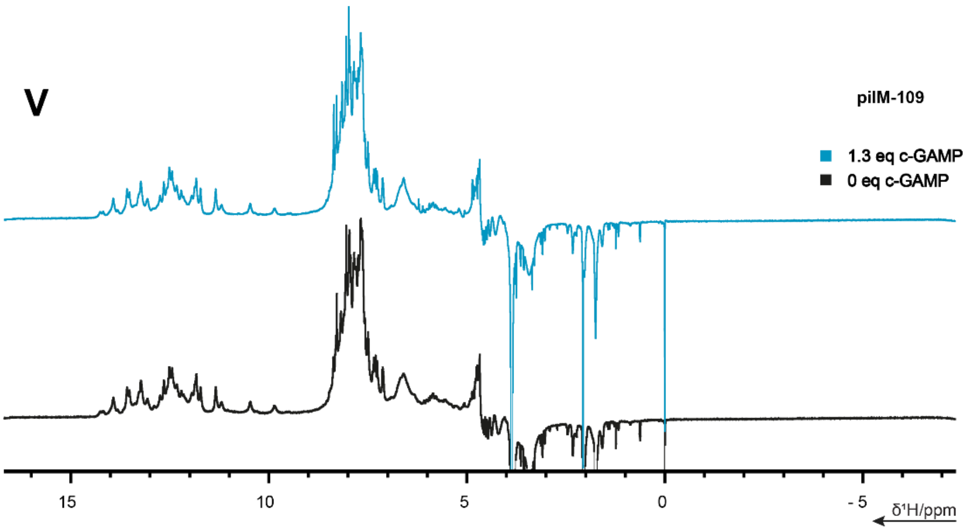

Q
pilM-101
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm


R
pilM-102
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm


S
pilM-103
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm


T
pilM-104
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm


U
pilM-105
■ 1.3 eq c-GAMP
■ 0 eq c-GAMP
15
10
5
0
- 5
δ¹H/ppm



SI Figure 2: Full spectral width 1D spectra of different pilM transcript lengths (A) 1H-1D jump-and-return echo NMR spectra of pilM75 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (B) 1H-1D jump-and-return echo NMR spectra of pilM77 at a concentration of 75 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 2048 scans, and 24.5 ppm spectral width at 600 MHz. (C) 1H-1D jump-and-return echo NMR spectra of pilM79 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (D) 1H-1D jump-and-return echo NMR spectra of pilM80 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (E) 1H-1D jump-and-return echo NMR spectra of pilM81 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (F) 1H-1D jump-and-return echo NMR spectra of pilM82 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (G) 1H-1D jump-and-return echo NMR spectra of pilM83 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (H) 1H-1D jump-and-return echo NMR spectra of pilM84 at a concentration of 100 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 65k time domain points, 256 scans, and 25 ppm spectral width at 600 MHz. (I) 1H-1D jump-and-return echo NMR spectra of pilM88 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. (J) 1H-1D jump-and-return echo NMR spectra of pilM91 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (K) 1H-1D jump-and-return echo NMR spectra of pilM93 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (L) 1H-1D jump-and-return echo NMR spectra of pilM95 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (M) 1H-1D jump-and-return echo NMR spectra of pilM96 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 1024 scans, and 24.5 ppm spectral width at 600 MHz. (N) 1H-1D jump-and-return echo NMR spectra of pilM97 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 600 MHz. (O) 1H-1D jump-and-return echo NMR spectra of pilM99 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 4k time domain points, 512 scans, and 24.5 ppm spectral width at 600 MHz. (P) 1H-1D jump-and-return echo NMR spectra of pilM100 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (Q) 1H-1D jump-and-return echo NMR spectra of pilM101 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (R) 1H-1D jump-and-return echo NMR spectra of pilM102 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (S) 1H-1D jump-and-return echo NMR spectra of pilM103 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 32k time domain points, 512 scans, and 24 ppm spectral width at 950 MHz. (T) 1H-1D jump-and-return echo NMR spectra of pilM104 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 512 scans, and 24 ppm spectral width at 600 MHz. (U) 1H-1D jump-and-return echo NMR spectra of pilM105 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. (V) 1H-1D jump-and-return echo NMR spectra of pilM109 at a concentration of 200 µM at 308K. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM MgCl2. The spectra were recorded with 8k time domain points, 1024 scans, and 24 ppm spectral width at 800 MHz. Figure as published in Landgraf *et al.* (1).

SI Table 2 Peak List of the resonance assignments for the apo form of pilM Riboswitch found in Figure 11A, B, and C. Table as published in Landgraf *et al.* (1).

Assignment 7A	**w1/1H ppm**	**w2/1H ppm**	**Assignment 7B**	**w1/1H ppm**	**w2/1H ppm**	**Assignment 7C**	**w1/1H ppm**	**w2/1H ppm**
U15H3-G38H1	10,807	11,032	**U15H3-G38H1**	10,183	10,775	**U15H3-H3**	10.766	10.771
U15H3-G39H1	10,799	12,927	**U15H3-G39H1**	10,181	12,196	**U15H3-G38H1**	10.77	11.037
U15H3-H3	10,805	10,805	**U15H3-H3**	10,182	10,188	**U22H3-H3**	11.872	11.875
U22H3-G30H1	11,854	12,885	**U22H3-G30H1**	11,761	12,809	**U22H3-G30H1**	11.876	12.901
U22H3-G31H1	11,859	11,355	**U22H3-G31H1**	11,767	11,361	**U22H3-G31H1**	11.873	11.37
U22H3-H3	11,855	11,855	**U22H3-H3**	11,762	11,768	**U22H3-U32H3**	11.872	13.601
U22H3-U32H3	11,854	13,593	**U22H3-U32H3**	11,761	13,616	**G25H1-H1**	12.247	12.246
G25H1-H1	12,149	12,155	**G25H1-H1**	12,204	12,206	**A28H2-H2**	9.485	9.488
G27H1-H1	9,553	9,557	**A28H2-G62H1**	9,449	12,929	**A28H2-G62H1**	9.48	12.903
A28H2-G52H1	9,489	12,902	**A28H2-H2**	9,444	9,446	**A28H2-A63H8**	9.485	8.706
A28H2-H2	9,489	9,490	**A28H2-U55H3**	9,423	14,016	**G29H1-H1**	12.37	12.372
A28H2-U55H3	9,510	14,010	**G29H1-G25H1**	12,358	12,203	**G29H1-G30H1**	12.364	12.902
G29H1-G30H1	12,369	12,885	**G29H1-G25H1**	12,210	12,331	**G30H1-U22H3**	12.901	11.878
G29H1-H1	12,366	12,364	**G29H1-G30H1**	12,353	12,806	**G30H1-G29H1**	12.904	12.37
G30H1-G29H1	12,892	12,362	**G29H1-H1**	12,340	12,348	**G30H1-H1**	12.895	12.901
G30H1-G31H1	12,884	11,357	**G30H1-G29H1**	12,836	12,363	**G30H1-G31H1**	12.907	11.373
G30H1-H1	12,894	12,895	**G30H1-G31H1**	12,825	11,352	**G31H1-U22H3**	11.367	11.877
G30H1-U22H3	12,887	11,860	**G30H1-H1**	12,807	12,812	**G31H1-G30H1**	11.371	12.904
G31H1-G30H1	11,342	12,887	**G30H1-U22H3**	12,808	11,768	**G31H1-H1**	11.366	11.37
G31H1-H1	11,353	11,355	**G31H1-G30H1**	11,363	12,819	**G31H1-U32H3**	11.37	13.603
G31H1-U22H3	11,353	11,859	**G31H1-G33H1**	11,339	11,834	**U32H3-U22H3**	13.6	11.877
G31H1-U32H3	11,350	13,593	**G31H1-H1**	11,354	11,359	**U32H3-G31H1**	13.598	11.367
U32H3-G31H1	13,593	11,354	**G31H1-U22H3**	11,357	11,768	**U32H3-H3**	13.601	13.602
U32H3-H3	13,592	13,592	**G31H1-U32H3**	11,354	13,613	**G33H1-G34H1**	11.874	13.205
U32H3-U22H3	13,589	11,857	**U32H3-G31H1**	13,605	11,359	**G34H1-G33H1**	13.196	11.873
G33H1-G34H1	11,849	13,192	**U32H3-G33H1**	13,606	11,884	**G34H1-H1**	13.208	13.209
G33H1-H1	11,855	11,855	**U32H3-H3**	13,612	13,614	**G35H1-G34H1**	10.953	13.212
G34H1-G33H1	13,191	11,853	**U32H3-U22H3**	13,611	11,768	**G35H1-H1**	10.94	10.947
G34H1-G35H1	13,179	10,944	**G33H1-G31H1**	11,843	11,332	**C37H41-G66H1**	9.031	13.165
G34H1-H1	13,188	13,192	**G33H1-G34H1**	11,890	13,217	**G38H1-U15H3**	11.037	10.771
G35H1-G34H1	10,935	13,193	**G33H1-H1**	11,866	11,881	**G38H1-H1**	11.037	11.043
G35H1-H1	10,933	10,937	**G34H1-G33H1**	13,200	11,886	**U46H3-H3**	10.326	10.327
C37H41-G38H1	9,084	11,028	**G34H1-G35H1**	13,201	11,132	**G50H1-H1**	12.64	12.641
C37H41-G66H1	9,091	13,170	**G34H1-H1**	13,207	13,213	**G50H1-G51H1**	12.632	13.016
C37H41-H41	9,079	9,086	**G35H1-G34H1**	11,144	13,220	**G51H1-G50H1**	13	12.643

C37H41-U15H3	9,092	10,804
G38H1-C37H41	11,031	9,108
G38H1-G39H1	11,021	12,930
G38H1-G66H1	11,003	13,161
G38H1-H1	11,030	11,032
G38H1-U15H3	11,033	10,807
G39H1-G38H1	12,915	11,031
G39H1-U15H3	12,931	10,805
U46H3-G69H1	10,294	9,154
U46H3-G70H1	10,288	13,105
U46H3-H3	10,294	10,295
G50H1-G51H1	12,661	13,022
G50H1-H1	12,664	12,664
G51H1-G50H1	13,022	12,667
G51H1-H1	13,026	13,026
G52H1-A28H2	12,905	9,490
G52H1-H1	12,894	12,895
G52H1-U53H3	12,891	14,775
U53H3-G52H1	14,777	12,900
U53H3-G62H1	14,777	12,900
U53H3-G62H1	14,777	12,900
U53H3-H3	14,766	14,775
U53H3-H3	14,797	14,785
U55H3-A28H2	13,998	9,490
U55H3-G60H1	13,995	12,385
U55H3-G62H1	13,999	12,903
U55H3-H3	13,999	13,999
G60H1-H1	12,366	12,381
G60H1-U55H3	12,381	13,993
G62H1-H1	12,894	12,895
G62H1-U53H3	12,891	14,775
G62H1-U55H3	12,900	14,000
G66H1-C37H41	13,159	9,107
G66H1-G38H1	13,166	11,034
G66H1-H1	13,164	13,159
G69H1-H1	9,119	9,128
G69H1-U46H3	9,143	10,296
G70H1-G71H1	13,102	12,448
G70H1-H1	13,104	13,106
G70H1-U46H3	13,096	10,293
G71H1-G70H1	12,442	13,102
G71H1-H1	12,442	12,446

G35H1-H1	11,131	11,135
G38H1-G39H1	10,779	12,200
G38H1-G66H1	10,772	13,095
G38H1-H1	10,776	10,780
G38H1-U15H3	10,776	10,189
G39H1-G38H1	12,181	10,780
G39H1-H1	12,204	12,206
G39H1-U15H3	12,198	10,184
U46H3-G69H1	10,344	9,116
U46H3-G70H1	10,356	13,076
U46H3-H3	10,334	10,342
G50H1-G51H1	12,628	13,041
G50H1-H1	12,629	12,634
G51H1-G50H1	13,023	12,636
G51H1-H1	13,028	13,034
G52H1-H1	12,929	12,925
G52H1-U53H3	12,827	14,824
U53H3-G52H1	14,814	12,897
U53H3-G62H1	14,799	12,907
U53H3-H3	14,822	14,823
U55H3-A28H2	14,014	9,439
U55H3-G60H1	14,009	12,401
U55H3-G62H1	14,019	12,929
U55H3-H3	14,013	14,016
G60H1-H1	12,391	12,400
G60H1-U55H3	12,401	14,015
G62H1-A28H2	12,926	9,448
G62H1-H1	12,925	12,930
G62H1-U53H3	12,827	14,824
G62H1-U55H3	12,923	14,011
G66H1-G38H1	13,089	10,793
G66H1-H1	13,086	13,084
G69H1-G70H1	9,065	13,033
G69H1-H1	9,098	9,109
G69H1-U46H3	9,094	10,340
G70H1-G69H1	13,060	9,089
G70H1-G71H1	13,066	12,476
G70H1-H1	13,055	13,055
G70H1-U46H3	13,052	10,342
G71H1-G70H1	12,452	13,074
G71H1-H1	12,466	12,473

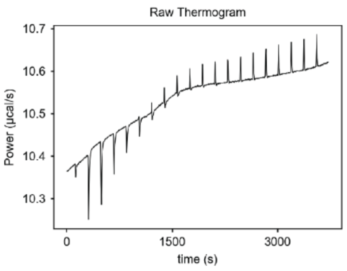
G51H1-H1	13.018	13.018
G51H1-G52H1	13.031	12.892
G52H1-G51H1	12.912	13.022
G52H1-H1	12.902	12.915
G52H1-U53H3	12.898	14.781
U53H3-G52H1	14.777	12.898
U53H3-H3	14.779	14.78
U53H3-G62H1	14.778	12.9
U55H3-A28H2	14.006	9.48
U55H3-H3	14.004	14.005
U55H3-G60H1	13.999	12.391
U55H3-G62H1	14.004	12.905
G60H1-U55H3	12.39	14.007
G60H1-H1	12.392	12.388
G62H1-A28H2	12.899	9.483
G62H1-U53H3	12.896	14.779
G62H1-U55H3	12.892	13.991
G62H1-H1	12.902	12.904
G69H1-U46H3	9.179	10.313
G69H1-H1	9.182	9.187
G70H1-H1	13.055	13.063
G70H1-G71H1	13.063	12.45
G71H1-G70H1	12.45	13.059
G71H1-H1	12.448	12.449

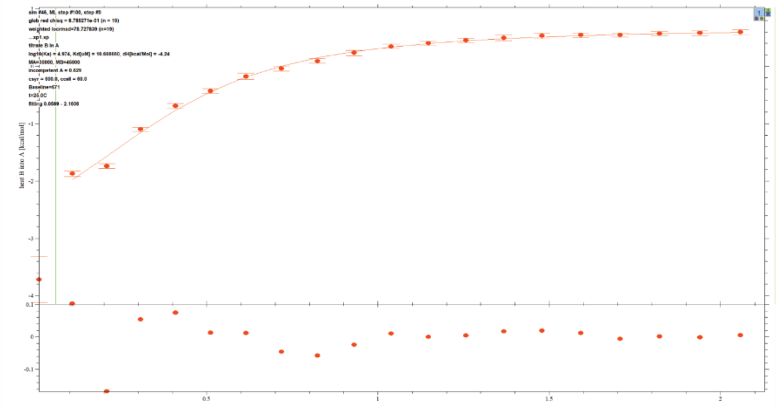
5.2.7 ITC

ITC isotherms were acquired on a MicroCal iTC200 (GE). The measurement volume of 205 µL contained RNAs at a concentration of 60 µM (pilM-109 measurement was carried out at 50 µM RNA concentration). 600 µM c-GAMP solution was added in 4 µL steps after an initial 0.4 µL injection. The obtained isotherms were baseline-corrected and integrated using Nitpic software (260). The resulting thermograms for all pilM measurements were fitted using SEDPHAT (SI Figure 3) (261). k_{on} and k_{off} rates from equilibration time curve fitting were obtained through KinITC software (Affinimeter) (SI Figure 4) (72).

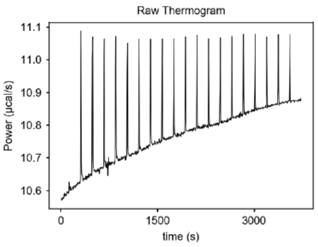
5.2.8 ITC data

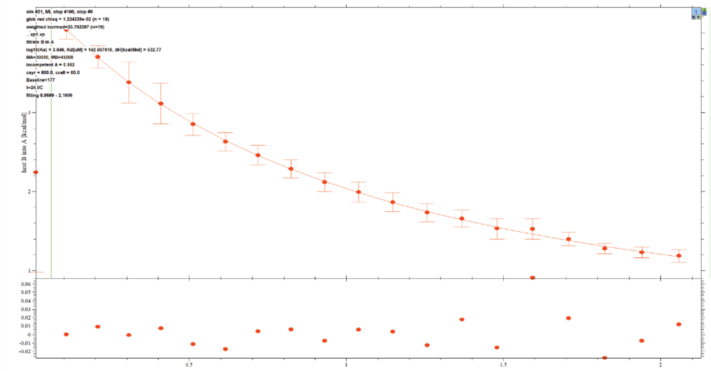
pilM-14-75



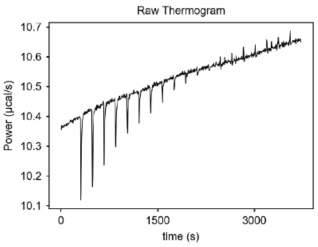


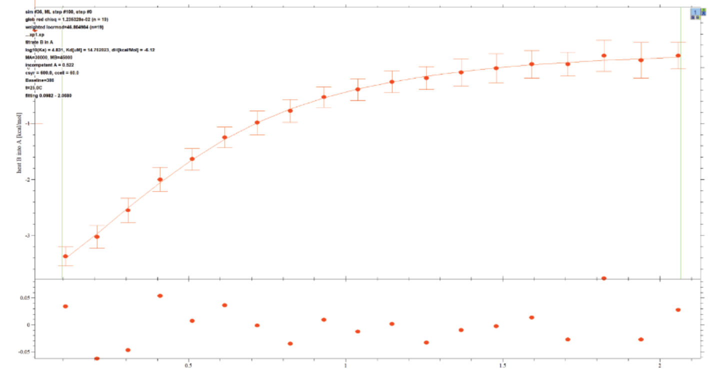
pilM-75



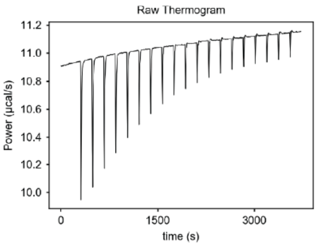


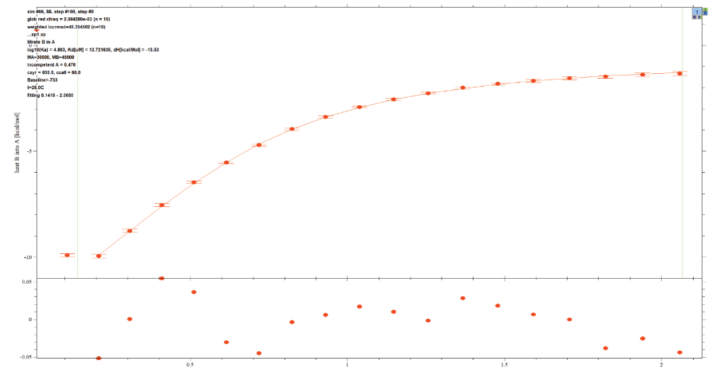
pilM-77



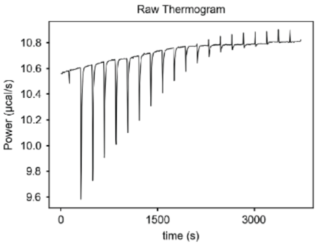


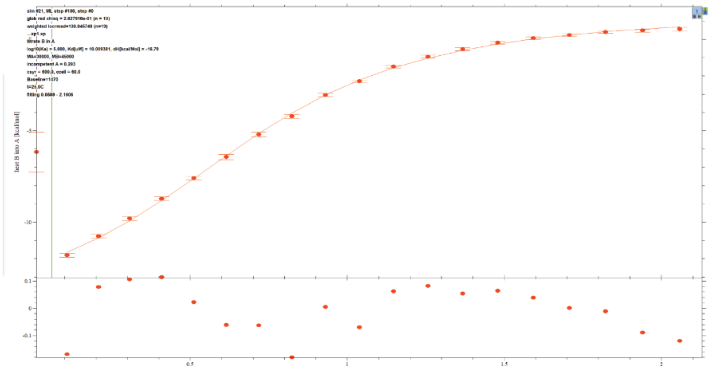
pilM-79



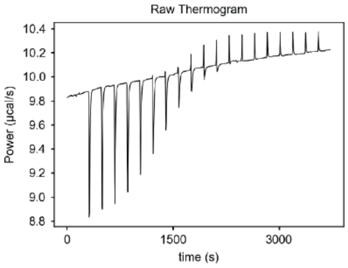


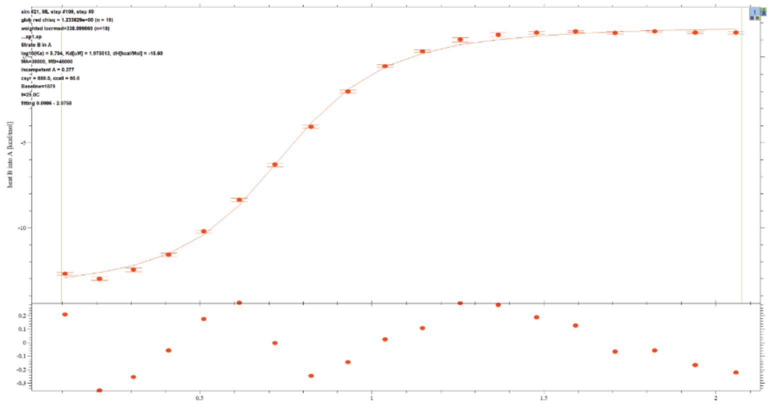
pilM-80



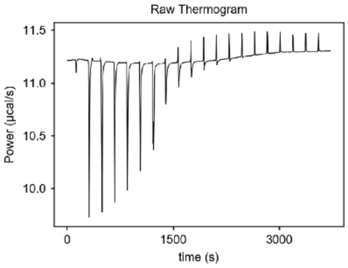


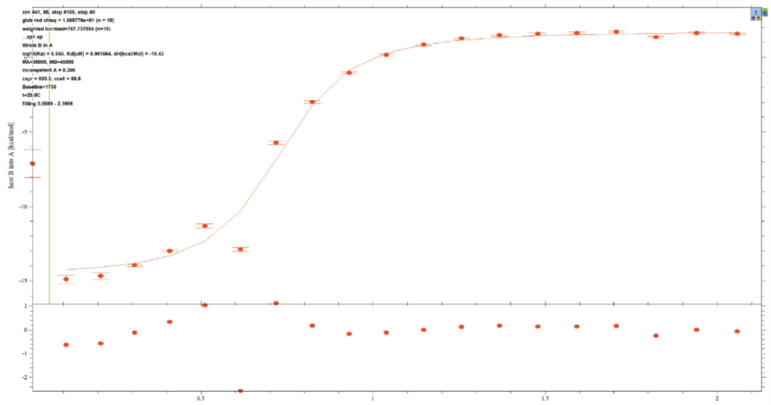
pilM-82



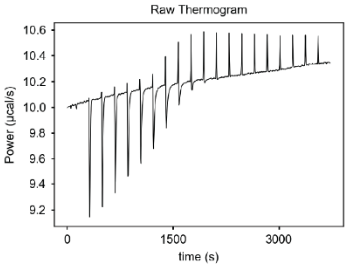


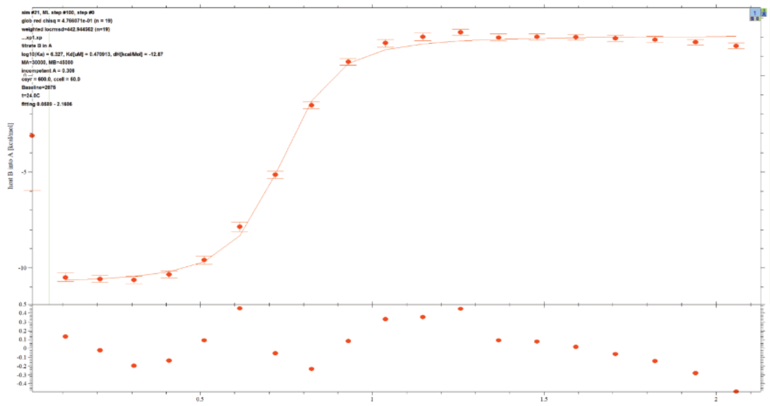
pilM-83



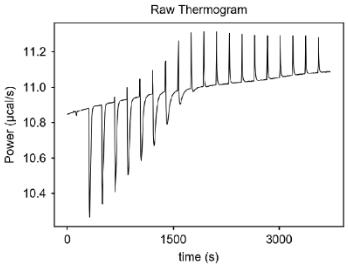


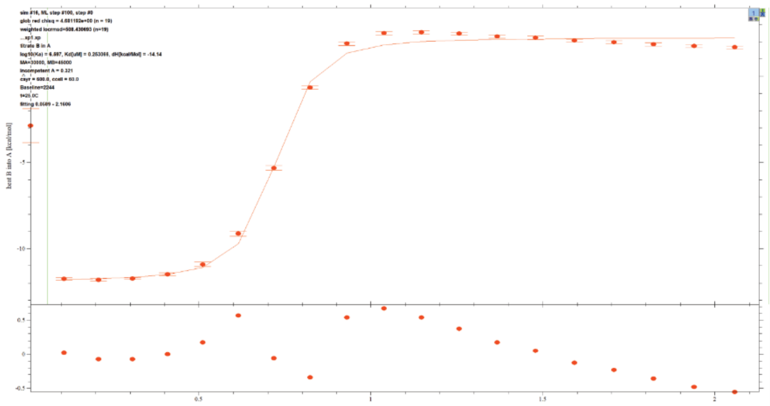
pilM-84



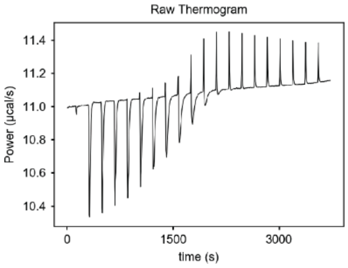


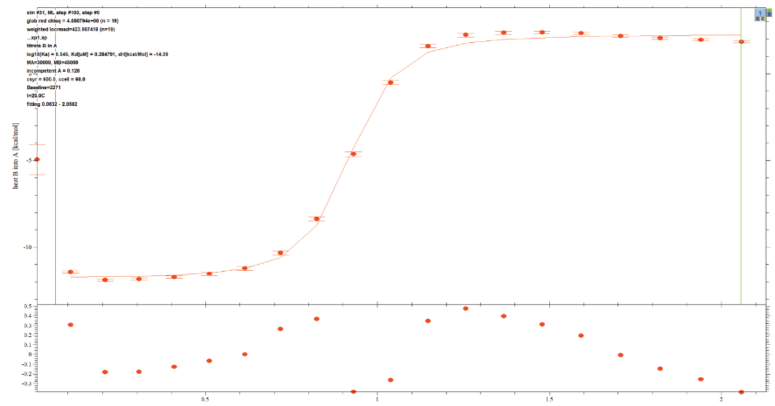
pilM-88



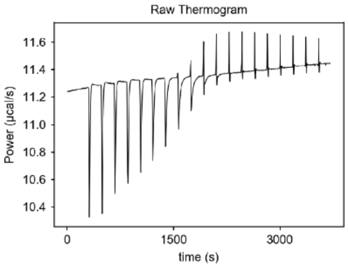


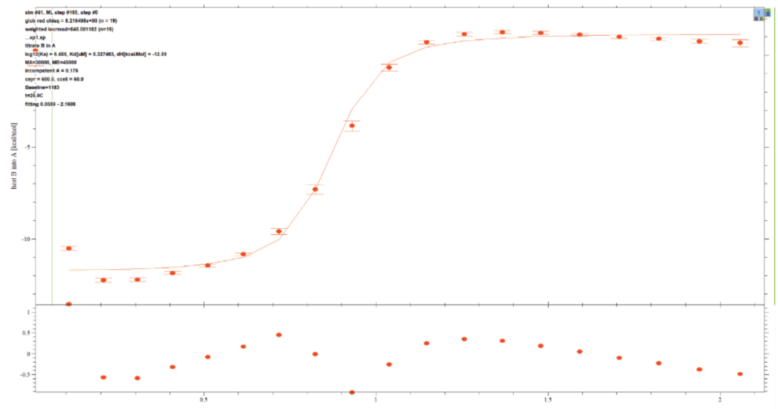
pilM-91



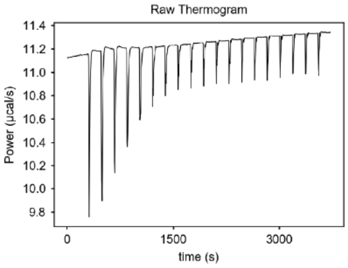


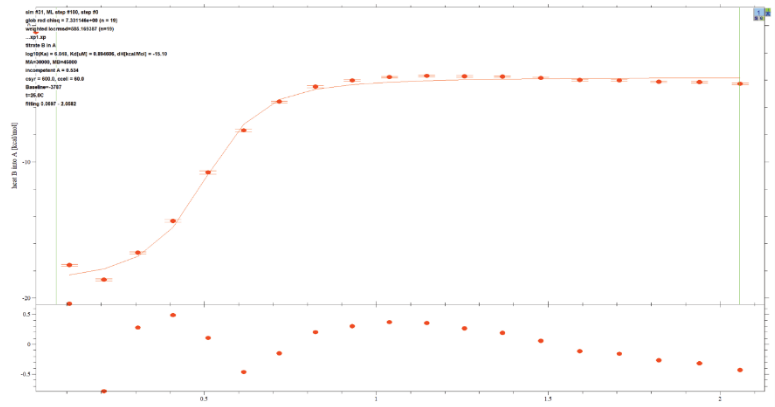
pilM-93



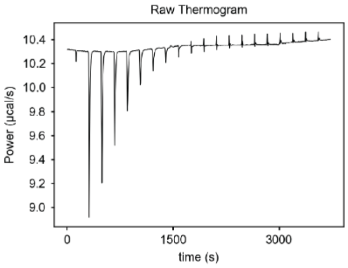


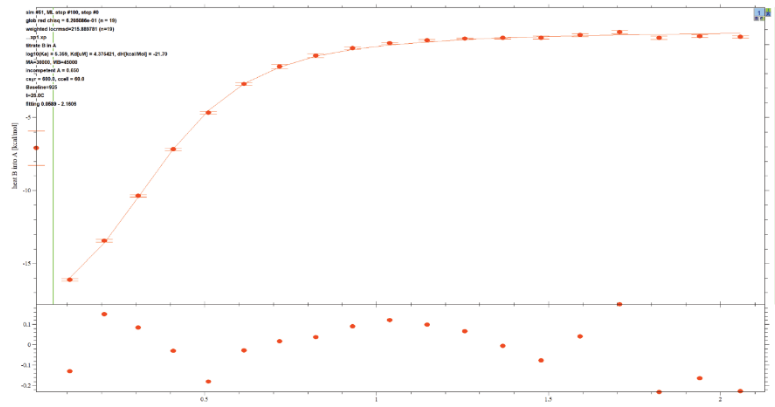
pilM-95



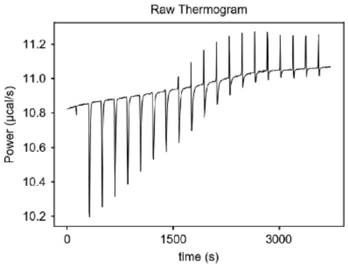


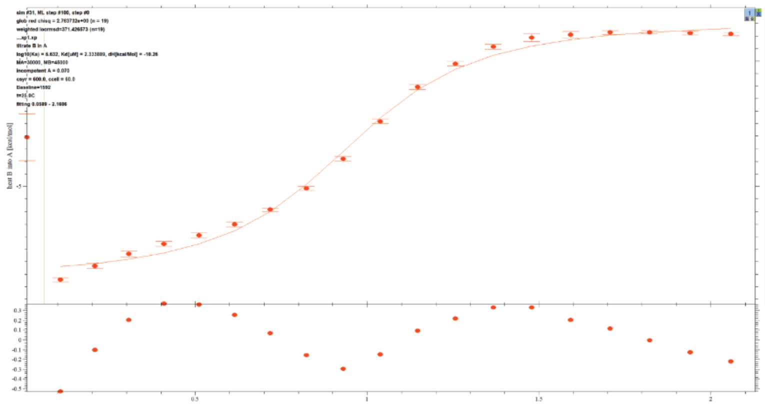
pilM-96



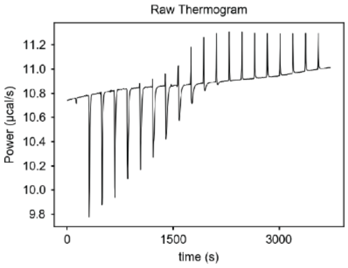


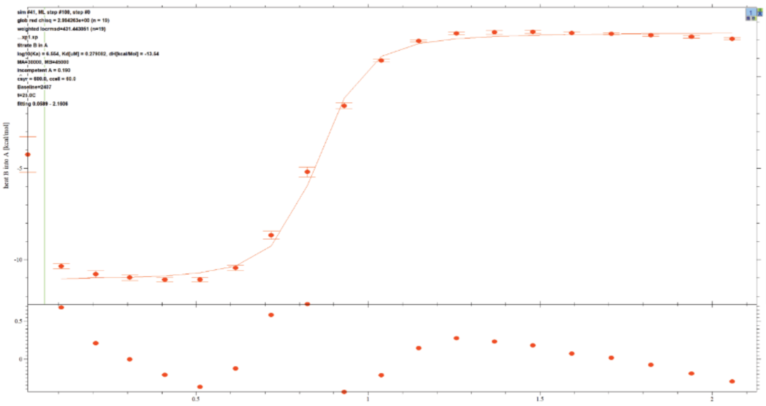
pilM-97



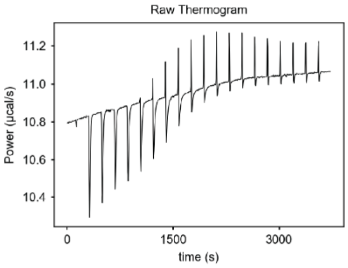


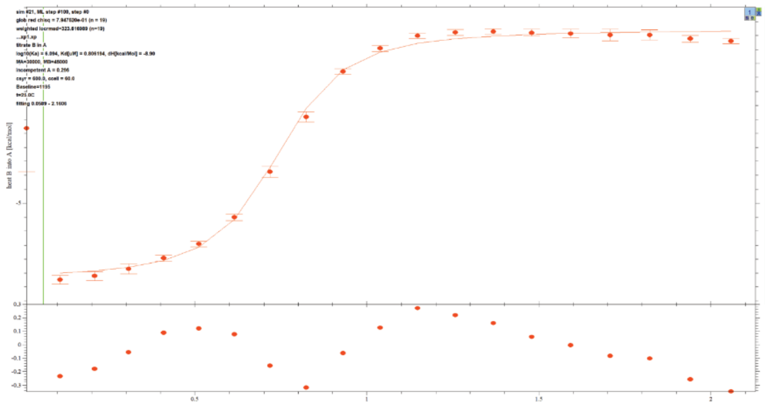
pilM-99



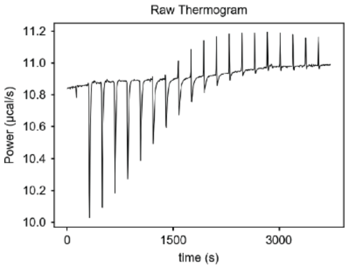


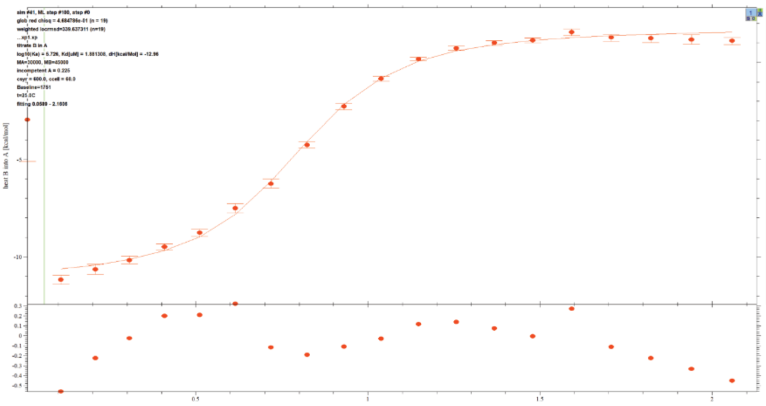
pilM-100



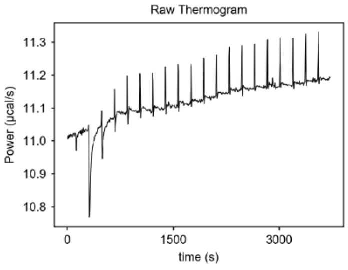


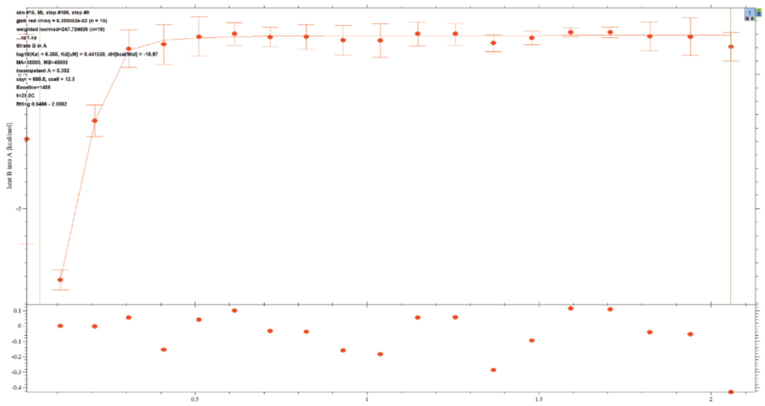
pilM-101



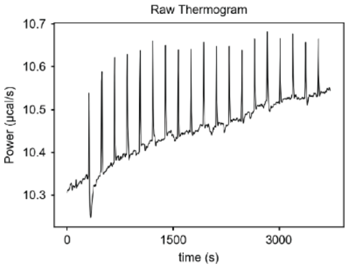


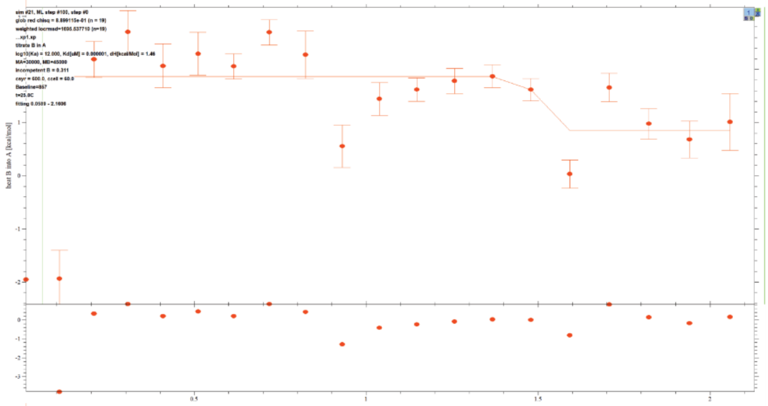
pilM-102



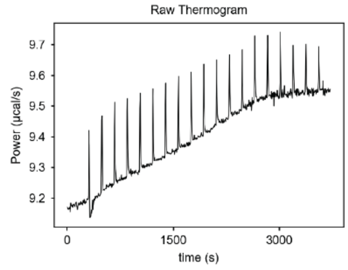


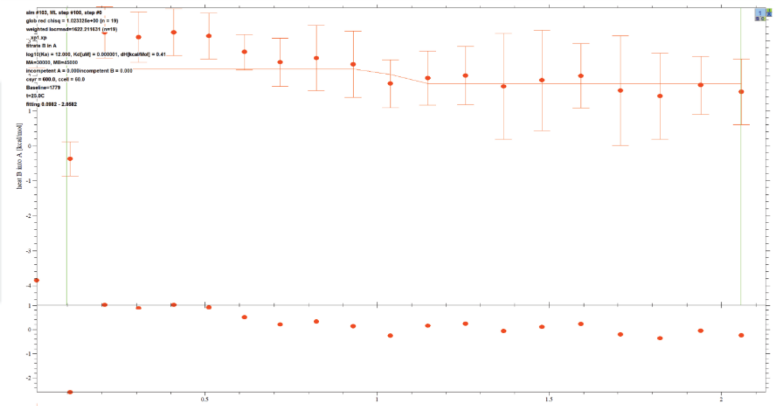
pilM-103



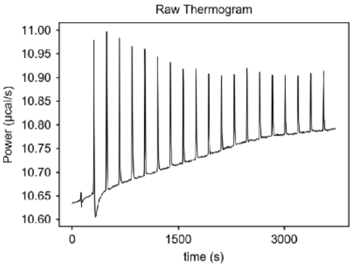


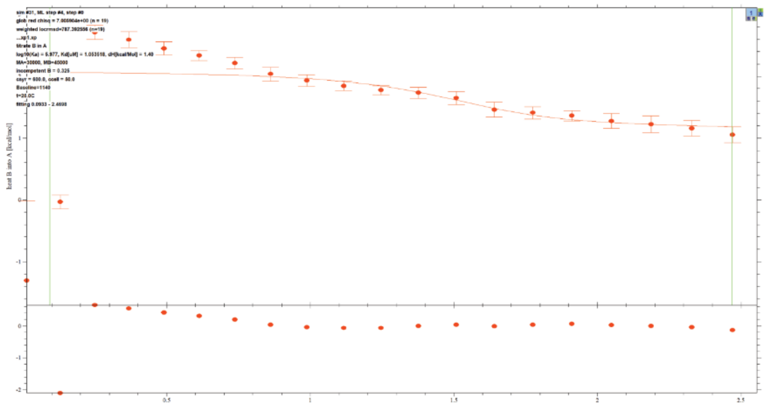
pilM-105





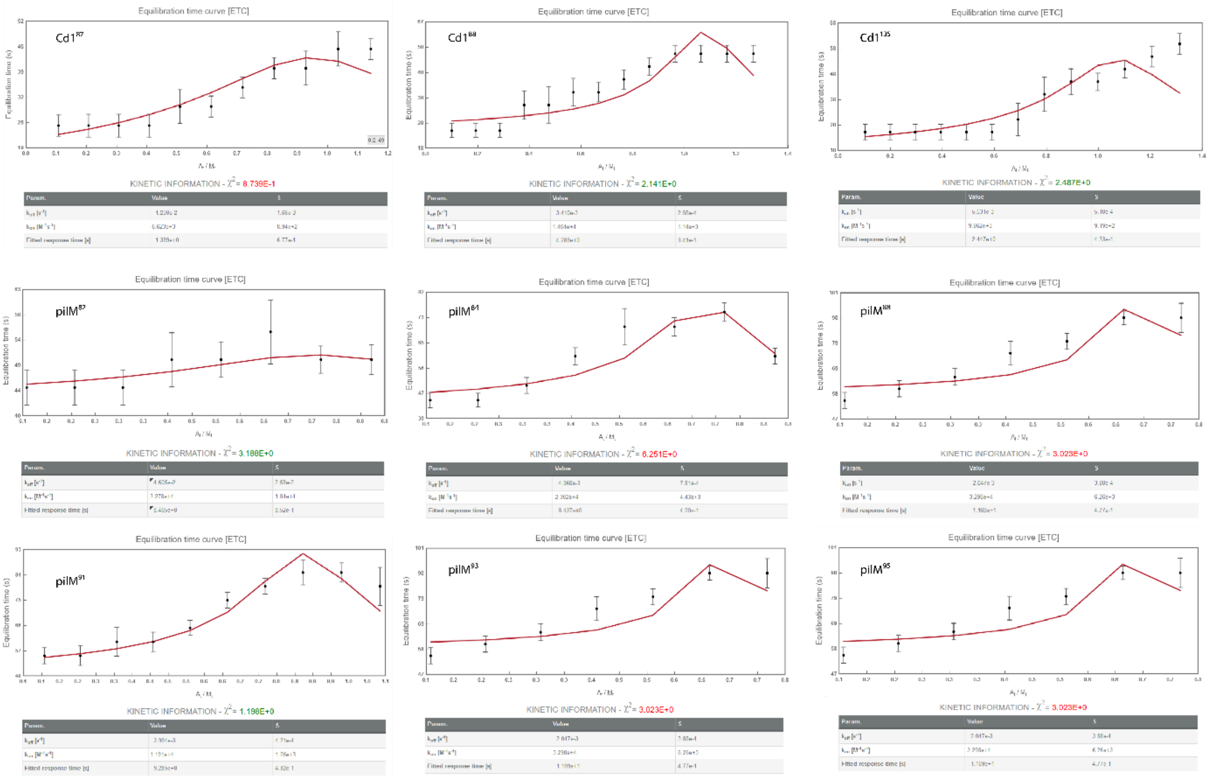
pilM-109





SI Figure 3: ITC raw data for the pilM riboswitch: The raw thermogram and the original fit in Sedphat are shown. The pilM-109 measurement was carried out at 50 µM RNA concentration. All others were conducted at 60 µM RNA concentration. The sample also contained 50 mM Bis-Tris Buffer, 50 mM KCl, and 5 mM $MgCl_2$. Figure as published in Landgraf *et al.* (1).

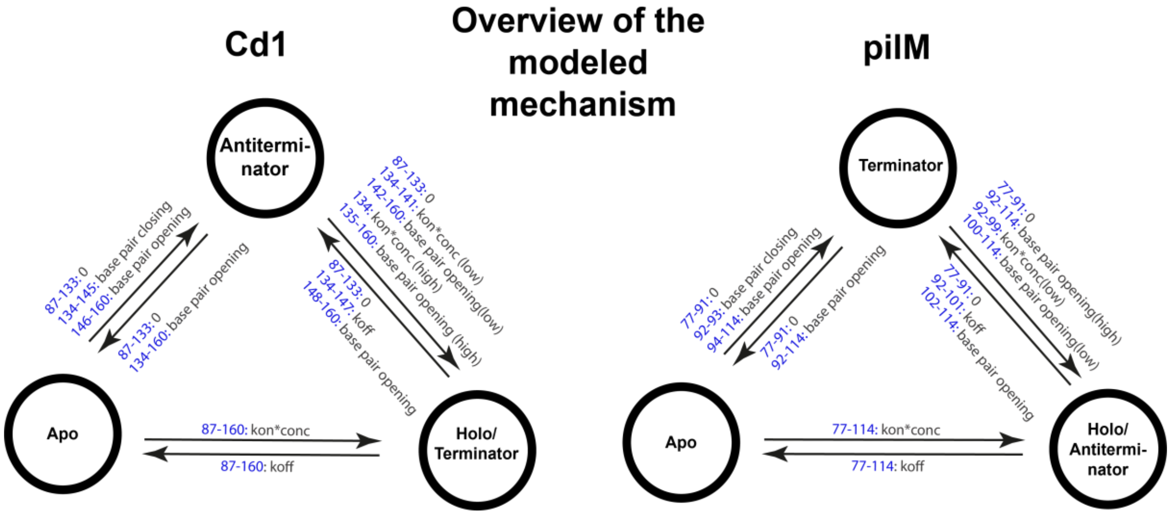
5.2.9 Kinetic data



SI Figure 4: Kinetic information obtained for strongly binding pilM and Cd1 riboswitch Constructs using kinITC. Figure as published in Landgraf *et al.* (1).

5.2.10 Simulation of cotranscriptional folding

Cotranscriptional folding can be simulated using a Markov model. Each possible fold of each transcript length is represented by a state. Rates are assigned to possible transitions between states. These rates correspond either to transcript elongation connecting a state of length n to a state of length n+1 or to a refolding transition connecting two folds of the same length. When a transition between 2 states is not allowed, the corresponding rate is set to 0. Examples of these forbidden transitions are elongations longer than n+1, degradation n-1, retention in the same state (diagonal element of the transition matrix), and transitions into folds that cannot form because their required residues are not transcribed yet. All RNAs of the same length were described by three states. Aptamer fold in the *apo* state, an aptamer-competing-fold which can be a terminator or antiterminator depending on the riboswitch, and a ligand-bound fold in the *holo* state with a ligand bound to the aptamer. The transitions between states are represented through first-order rate equations. The three states are interconnected through a total of six transition rates (262). Transition rates between the *apo* and *holo* states are the k_{on} and k_{off} rates determined in ITC measurements and the subsequent analyses of equilibration time curves. The determined k_{on} rates of high-affinity binding aptamers showed values and enough overlap in their margin of error to be considered constant. This constant rate was extrapolated to lower affinity binding aptamers. The k_{off} rates were determined by dividing the K_D by the constant k_{on} rate. Rates for transcript lengths that were not analyzed were interpolated using the mean of adjacent K_D values. The *apo* aptamer-competing-fold transitions used rates based on the different stability due to the number of formed base pairs. These rates were first introduced by Fürtig *et al*. (185). The *holo* and aptamer competing fold transitions require an association or discociation event and a refolding of the structure. Since bound ligand and folded base pairs contribute stabilizing hydrogen bonds to the *holo*-structure, a concerted mechanism is likely to require more energy to accord. The mechanism can therefore be assumed to be a two-step process where the slow rate acts as a limiting step. Transitions were modeled as a one-step process using the slower rate to aid computation speed. In practice, this resulted in the use of base pair difference for rates unless the k_{off} rate were lower due to a high-affinity aptamer or the product of k_{on} and ligand concentration, was lower because of low ligand concentration. The rate values of state transitions are illustrated in SI Figure 5. In addition to riboswitch behavior over time, changes in ligand concentration, transcription speed, base pair closing, and pausing duration were modeled and graphed as contour plots. To model pausing of the RNA polymerase, Pausing was set to occur 10 nucleotides before the pause site at nucleotide 145 of the Cd1 riboswitch. This setting allows only residues to engage into folding that have left the polymerase at the time of encountering the pause site and would not be occluded from folding while being in the exit tunnel (263).


Cd1
Overview of the modeled mechanism
pilM
Antitermi-nator
87-133: 0
134-145: base pair closing
146-160: base pair opening
87-133: 0
134-160: base pair opening
87-133: 0
134-141: kon*conc (low)
142-160: base pair opening(low)
134: kon*conc (high)
135-160: base pair opening (high)
87-133: 0
134-147: koff
148-160: base pair opening
Apo
87-160: kon*conc
87-160: koff
Holo/ Terminator
Terminator
77-91: 0
92-93: base pair closing
94-114: base pair opening
77-91: 0
92-114: base pair opening
77-91: 0
92-114: base pair opening(high)
92-99: kon*conc(low)
100-114: base pair opening(low)
77-91: 0
92-101: koff
102-114: base pair opening
Apo
77-114: kon*conc
77-114: koff
Holo/ Antitermi-nator

SI Figure 5: Graphical description of the three state model used in the Markov modeling. The blue numbers refer to the RNA length ranges for which the indicated rates were applied. Low and high designate which rate was used for low or for high ligand concentration simulations (100 µM to 100 nM), respectively. Figure as published in Landgraf *et al.* (1).

5.2.11 Source code of Markov model

Example of a contour plot showing concentration dependence over time is used in Figure 13A PilM.

```
l=38;
states=3;
n=l*states+1;
KM=Array[k,{n,n},{1,1}];
KM//MatrixForm;

For[j=1,j<n+1,j++,For[i=1,i<n+1,i++,k[i,j]=0]]
KM//MatrixForm;

(*synth=transscriptionspeed*)
For[i=1,i<n+1,i++,k[i,i+1]=synth]
k[l+1,l+2]=0;
k[2*l+1,2*l+2]=0;

(*pausing at 119 length for few seconds*)
lengthexitthunnel=10;
(*!!!!!matrix elements below need to be comented out if the is no pausing or pausing duration is set to 1/synth
which means no pausing because it yields normal synth speed.!!!!!!*)
k[44-lengthexitthunnel,1+44-lengthexitthunnel]=1.0/(pausingduration);
k[l+44-lengthexitthunnel,l+1+44-lengthexitthunnel]=1.0/(pausingduration);
k[2*l+44-lengthexitthunnel,2*l+1+44-lengthexitthunnel]=1.0/(pausingduration);
KM//MatrixForm;


m2=2; (*77*)
m3=3;(*78*)
m4=4;(*79*)
m5=5;(*80*)
m6=6;(*81*)
m7=7;(*82*)
m8=8;(*83*)
```

```
m9=9;(*84*)
m10=10;(*85*)
m11=11;(*86*)
m12=12;(*87*)
m13=13;(*88*)
m14=14;(*89*)
m15=15;(*90*)
m16=16;(*91*)
m17=17;(*92*)
m18=18;(*93*)
m19=19;(*94*)
m20=20;(*95*)
m21=21;(*96*)
m22=22;(*97*)
m23=23;(*98*)
m24=24;(*99*)
m25=25;(*100*)
m26=26;(*101*)
m27=27;(*102*)
m28=28;(*103*)
m29=29;(*104*)
m30=30;(*105*)
m31=31;(*106*)
m32=32;(*107*)
m33=33;(*108*)
m34=34;(*109*)
m35=35;(*110*)
m36=36;(*111*)
m37=37;(*112*)
m38=38;(*113*)
m39=39 ;(*114*)

kd2=19000*10^-9;
kd3=(kd2+kd4)/2;
kd4=17000*10^-9;
kd5=8000*10^-9;
kd6=(kd5+kd7)/2;
kd7=2100*10^-9;
kd8=1670*10^-9;
kd9=430*10^-9;
kd10=(kd9+kd13)/2;
kd11=(kd9+kd13)/2;
kd12=(kd9+kd13)/2;
kd13=250*10^-9;
kd14=(kd13+kd16)/2;
kd15=(kd13+kd16)/2;
kd16=310*10^-9;
kd17=(kd16+kd18)/2;
kd18=380*10^-9;
kd19=(kd18+kd20)/2;
kd20=1500*10^-9;
kd21=3000*10^-9;
kd22=1600*10^-9;
kd23=(kd22+kd24)/2;
kd24=250*10^-9;
kd25=500*10^-9;
kd26=1300*10^-9;
kd27=17000*10^-9;
kdapomittel=(kd9+kd13+kd16)/3; (* best estimate what the Kd of compleat aptamer is*)

(*apo->holo* with const kon*)
For [m=2,m<l+2,m++,k[m,2*l+m]=kon*conc]
k[m-1,2*l+m-1];(*prints a k to check if line aboth works*)

(*holo->apo* with const kon to calculate koff from kd*)
k[m2+l*2,m2]=kd2*kon;
k[m3+l*2,m3]=kd3*kon;
```

```
k[m4+l*2,m4]=kd4*kon;
k[m5+l*2,m5]=kd5*kon;
k[m6+l*2,m6]=kd6*kon;
k[m7+l*2,m7]=kd7*kon;
k[m8+l*2,m8]=kd8*kon;
k[m9+l*2,m9]=kd9*kon;
k[m10+l*2,m10]=kd10*kon;
k[m11+l*2,m11]=kd11*kon;
k[m12+l*2,m12]=kd12*kon;
k[m13+l*2,m13]=kd13*kon;
k[m14+l*2,m14]=kd14*kon;
k[m15+l*2,m15]=kd15*kon;
k[m16+l*2,m16]=kd16*kon;
k[m17+l*2,m17]=kdapomittel*kon;
k[m18+l*2,m18]=kdapomittel*kon;
k[m19+l*2,m19]=kdapomittel*kon;
k[m20+l*2,m20]=kdapomittel*kon;
k[m21+l*2,m21]=kdapomittel*kon;
k[m22+l*2,m22]=kdapomittel*kon;
k[m23+l*2,m23]=kdapomittel*kon;
k[m24+l*2,m24]=kdapomittel*kon;
k[m25+l*2,m25]=kdapomittel*kon;
k[m26+l*2,m26]=kdapomittel*kon;
k[m27+l*2,m27]=kdapomittel*kon;
k[m28+l*2,m28]=kdapomittel*kon;
k[m29+l*2,m29]=kdapomittel*kon;
k[m30+l*2,m30]=kdapomittel*kon;
k[m31+l*2,m31]=kdapomittel*kon;
k[m32+l*2,m32]=kdapomittel*kon;
k[m33+l*2,m33]=kdapomittel*kon;
k[m34+l*2,m34]=kdapomittel*kon;
k[m35+l*2,m35]=kdapomittel*kon;
k[m36+l*2,m36]=kdapomittel*kon;
k[m37+l*2,m37]=kdapomittel*kon;
k[m38+l*2,m38]=kdapomittel*kon;
k[m39+l*2,m39]=kdapomittel*kon;

k[m2,m2+l]= 0;(*apo->term*)
k[m3,m3+l]= 0;
k[m4,m4+l]= 0;
k[m5,m5+l]= 0;
k[m6,m6+l]= 0;
k[m7,m7+l]= 0;
k[m8,m8+l]= 0;
k[m5,m5+l]= 0;
k[m6,m6+l]= 0;
k[m7,m7+l]= 0;
k[m8,m8+l]= 0;
k[m9,m9+l]= 0;
k[m10,m10+l]= 0;
k[m11,m11+l]= 0;
k[m12,m12+l]= 0;
k[m13,m13+l]= 0;
k[m14,m14+l]= 0;
k[m15,m15+l]= 0;
k[m16,m16+l]= 0;
k[m17,m17+l]= baseclosing;(*+1*)
k[m18,m18+l]= baseclosing;(*+2*)
k[m19,m19+l]= 2.590597845;(*-1*)
k[m20,m20+l]= 1.056087738;(*-2*)
k[m21,m21+l]= 0.43052661;(*-3*)
k[m22,m22+l]= 0.175509246;(*-4*)
k[m23,m23+l]= 0.175509246;(*-4*)
k[m24,m24+l]= 0.071548412;(*-5*)
k[m25,m25+l]= 0.029167553;(*-6*)
k[m26,m26+l]= 0.011890497;(*-7*)
k[m27,m27+l]= 0.011890497;(*-7*)
```

```
k[m28,m28+l]= 0.011890497;(*-7*)
k[m29,m29+l]= 0.000805564;(*-10*)
k[m30,m30+l]= 0.000805564;(*-10 ab hier eigentlich mehr als 10*)
k[m31,m31+l]= 0.000805564;(*-10*)
k[m32,m32+l]= 0.000805564;(*-10*)
k[m33,m33+l]= 0.000805564;(*-10*)
k[m34,m34+l]= 0.000805564;(*-10*)
k[m35,m25+l]= 0.000805564;(*-10*)
k[m36,m36+l]= 0.000805564;(*-10*)
k[m37,m37+l]= 0.000805564;(*-10*)
k[m38,m38+l]= 0.000805564;(*-10*)
k[m39,m39+l]= 0.000805564;(*-10*)


k[m2+l,m2]=0; (*term->apo*)
k[m3+l,m3]= 0;
k[m4+l,m4]= 0;
k[m5+l,m5]= 0;
k[m6+l,m6]= 0;
k[m7+l,m7]= 0;
k[m8+l,m8]= 0;
k[m5+l,m5]= 0;
k[m6+l,m6]= 0;
k[m7+l,m7]= 0;
k[m8+l,m8]= 0;
k[m9+l,m9]= 0;
k[m10+l,m10]= 0;
k[m11+l,m11]= 0;
k[m12+l,m12]= 0;
k[m13+l,m13]= 0;
k[m14+l,m14]= 0;
k[m15+l,m15]= 0;
k[m16+l,m16]= 0;
k[m17+l,m17]= 2.590597845;(*-1*)
k[m18+l,m18]= 1.056087738;(*-2*)
k[m19+l,m19]= 0.43052661;(*-3*)
k[m20+l,m20]= 0.175509246;(*-4*)
k[m21+l,m21]= 0.071548412;(*-5*)
k[m22+l,m22]= 0.029167553;(*-6*)
k[m23+l,m23]= 0.029167553;(*-6*)
k[m24+l,m24]= 0.011890497;(*-7*)
k[m25+l,m25]= 0.004847301;(*-8*)
k[m26+l,m26]= 0.001976059;(*-9*)
k[m27+l,m27]= 0.000805564;(*-10*)
k[m28+l,m28]= 0.000805564;(*-10 ab hier eigentlich mehr als 10*)
k[m29+l,m29]= 0.000805564;(*-10*)
k[m30+l,m30]= 0.000805564;(*-10*)
k[m31+l,m31]= 0.000805564;(*-10*)
k[m32+l,m32]= 0.000805564;(*-10*)
k[m33+l,m33]= 0.000805564;(*-10*)
k[m34+l,m34]= 0.000805564;(*-10*)
k[m35+l,m35]= 0.000805564;(*-10*)
k[m36+l,m36]= 0.000805564;(*-10*)
k[m37+l,m37]= 0.000805564;(*-10*)
k[m38+l,m38]= 0.000805564;(*-10*)
k[m39+l,m39]= 0.000805564;(*-10*)


k[m2+l,m2+l*2]= 0; (*term->holo*)
k[m3+l,m3+l*2]= 0;
k[m4+l,m4+l*2]= 0;
k[m5+l,m5+l*2]= 0;
k[m6+l,m6+l*2]= 0;
k[m7+l,m7+l*2]= 0;
k[m8+l,m8+l*2]= 0;
k[m5+l,m5+l*2]= 0;
k[m6+l,m6+l*2]= 0;
```

```
k[m7+l,m7+l*2]= 0;
k[m8+l,m8+l*2]= 0;
k[m9+l,m9+l*2]= 0;
k[m10+l,m10+l*2]= 0;
k[m11+l,m11+l*2]= 0;
k[m12+l,m12+l*2]= 0;
k[m13+l,m13+l*2]= 0;
k[m14+l,m14+l*2]= 0;
k[m15+l,m15+l*2]= 0;
k[m16+l,m16+l*2]= 0;
k[m17+l,m17+l*2]= kon*conc;(*-1*)
k[m18+l,m18+l*2]= kon*conc;(*-2*)
k[m19+l,m19+l*2]= kon*conc;(*-3*)
k[m20+l,m20+l*2]= kon*conc;(*-4*)
k[m21+l,m21+l*2]= kon*conc;(*-5*)
k[m22+l,m22+l*2]= kon*conc;(*-6*)
k[m23+l,m23+l*2]= kon*conc;(*-6*)
k[m24+l,m24+l*2]= kon*conc;(*-7*) (*wahrscheinlich wäre es für das model richtiger maximal die 7 bp des p1 zu brechen*)
k[m25+l,m25+l*2]= 0.004847301;(*-8*)
k[m26+l,m26+l*2]= 0.001976059;(*-9*)
k[m27+l,m27+l*2]= 0.000805564;(*-10*)
k[m28+l,m28+l*2]= 0.000805564;(*-10 ab hier eigentlich mehr als 10*)
k[m29+l,m29+l*2]= 0.000805564;(*-10*)
k[m30+l,m30+l*2]= 0.000805564;(*-10*)
k[m31+l,m31+l*2]= 0.000805564;(*-10*)
k[m32+l,m32+l*2]= 0.000805564;(*-10*)
k[m33+l,m33+l*2]= 0.000805564;(*-10*)
k[m34+l,m34+l*2]= 0.000805564;(*-10*)
k[m35+l,m35+l*2]= 0.000805564;(*-10*)
k[m36+l,m36+l*2]= 0.000805564;(*-10*)
k[m37+l,m37+l*2]= 0.000805564;(*-10*)
k[m38+l,m38+l*2]= 0.000805564;(*-10*)
k[m39+l,m39+l*2]= 0.000805564;(*-10*)

kon*conc

k[m2+l*2,m2+l]= 0;(*holo->term*)
k[m3+l*2,m3+l]= 0;
k[m4+l*2,m4+l]= 0;
k[m5+l*2,m5+l]= 0;
k[m6+l*2,m6+l]= 0;
k[m7+l*2,m7+l]= 0;
k[m8+l*2,m8+l]= 0;
k[m5+l*2,m5+l]= 0;
k[m6+l*2,m6+l]= 0;
k[m7+l*2,m7+l]= 0;
k[m8+l*2,m8+l]= 0;
k[m9+l*2,m9+l]= 0;
k[m10+l*2,m10+l]= 0;
k[m11+l*2,m11+l]= 0;
k[m12+l*2,m12+l]= 0;
k[m13+l*2,m13+l]= 0;
k[m14+l*2,m14+l]= 0;
k[m15+l*2,m15+l]= 0;
k[m16+l*2,m16+l]= 0;
k[m17+l*2,m17+l]= kd17*kon; (*+1*)
k[m18+l*2,m18+l]= kd18*kon;(*+2*)
k[m19+l*2,m19+l]= kd19*kon;
k[m20+l*2,m20+l]= kd20*kon;
k[m21+l*2,m21+l]= kd21*kon;
k[m22+l*2,m22+l]= kd22*kon;
k[m23+l*2,m23+l]= kd23*kon;
k[m24+l*2,m24+l]= kd24*kon;
k[m25+l*2,m25+l]= kd25*kon;
k[m26+l*2,m26+l]= kd26*kon;
k[m27+l*2,m27+l]= 0.011890497;(*-7*)
```

```
k[m28+l*2,m28+l]= 0.011890497;(*-7*)
k[m29+l*2,m29+l]= 0.000805564;(*-10*)
k[m30+l*2,m30+l]= 0.000805564;(*-10*ab hier eigentlich mehr als 10*)
k[m31+l*2,m31+l]= 0.000805564;(*-10*)
k[m32+l*2,m32+l]= 0.000805564;(*-10*)
k[m33+l*2,m33+l]= 0.000805564;(*-10*)
k[m34+l*2,m34+l]= 0.000805564;(*-10*)
k[m35+l*2,m35+l]= 0.000805564;(*-10*)
k[m36+l*2,m36+l]= 0.000805564;(*-10*)
k[m37+l*2,m37+l]= 0.000805564;(*-10*)
k[m38+l*2,m38+l]= 0.000805564;(*-10*)
k[m39+l*2,m39+l]= 0.000805564;(*-10*)

KM//MatrixForm ;(*hier kann man sich die matrix for transformation anschauen*)



K=Transpose[KM];
For[i=1,i<=n,i++,K[[i,i]]=-Sum[KM[[i,j]],{j,1,n}]];K//MatrixForm ;(*hier kann man sich die transponierte matrix
anschauen*)
Clear[p];
p0=Table[0,n];
Part[p0,1]=1;
p0;
(*Length[p0]*)
p[t_,p0_]:=MatrixExp[t K].p0
(*p[t,p0]*)

t1=SessionTime[];
TimeObject[]
(*list of all variables in for matrix*)
synth=20;
conc= 100*10^-6;
kon=17000;
baseclosing=400;  (*4000 macht schon 10 min rechen zeit bitte keine 4000000 eintragen*)
plotend=3;
pausingduration=1/synth;
(*
(*apo plot*)
p200=Plot[((*{p[t, p0][[1]]}+*){p[t, p0][[m2]]}+{p[t, p0][[m3]]}+{p[t, p0][[m4]]}+{p[t, p0][[m5]]}+{p[t, p0][[m6]]}+{p[t,
p0][[m7]]}+{p[t, p0][[m8]]}+{p[t, p0][[m9]]}+{p[t, p0][[m10]]}+{p[t, p0][[m11]]}+{p[t, p0][[m12]]}+{p[t, p0][[m13]]}+{p[t,
p0][[m14]]}+{p[t, p0][[m15]]}+{p[t, p0][[m16]]}+{p[t, p0][[m17]]}+{p[t, p0][[m18]]}+{p[t, p0][[m19]]}+{p[t,
p0][[m20]]}+{p[t, p0][[m21]]}+{p[t, p0][[m22]]}+{p[t, p0][[m23]]}+{p[t, p0][[m24]]}+{p[t, p0][[m25]]}+{p[t,
p0][[m26]]}+{p[t, p0][[m27]]}+{p[t, p0][[m28]]}+{p[t, p0][[m29]]}+{p[t, p0][[m30]]}+{p[t, p0][[m31]]}+{p[t,
p0][[m32]]}+{p[t, p0][[m33]]}+{p[t, p0][[m34]]}+{p[t, p0][[m35]]}+{p[t, p0][[m36]]}+{p[t, p0][[m37]]}+{p[t,
p0][[m38]]}+{p[t, p0][[m39]]})/1, {t, 0, plotend}, PlotRange \[Rule]
{0,1.1},PlotStyle\[Rule]{Yellow,Thick},AspectRatio\[Rule]1]
(*terminator plot*)
p201=Plot[({p[t, p0][[m2+l]]}+{p[t, p0][[m3+l]]}+{p[t, p0][[m4+l]]}+{p[t, p0][[m5+l]]}+{p[t, p0][[m6+l]]}+{p[t,
p0][[m7+l]]}+{p[t, p0][[m8+l]]}+{p[t, p0][[m9+l]]}+{p[t, p0][[m10+l]]}+{p[t, p0][[m11+l]]}+{p[t, p0][[m12+l]]}+{p[t,
p0][[m13+l]]}+{p[t, p0][[m14+l]]}+{p[t, p0][[m15+l]]}+{p[t, p0][[m16+l]]}+{p[t, p0][[m17+l]]}+{p[t, p0][[m18+l]]}+{p[t,
p0][[m19+l]]}+{p[t, p0][[m20+l]]}+{p[t, p0][[m21+l]]}+{p[t, p0][[m22+l]]}+{p[t, p0][[m23+l]]}+{p[t, p0][[m24+l]]}+{p[t,
p0][[m25+l]]}+{p[t, p0][[m26+l]]}+{p[t, p0][[m27+l]]}+{p[t, p0][[m28+l]]}+{p[t, p0][[m29+l]]}+{p[t, p0][[m30+l]]}+{p[t,
p0][[m31+l]]}+{p[t, p0][[m32+l]]}+{p[t, p0][[m33+l]]}+{p[t, p0][[m34+l]]}+{p[t, p0][[m35+l]]}+{p[t, p0][[m36+l]]}+{p[t,
p0][[m37+l]]}+{p[t, p0][[m38+l]]}+{p[t, p0][[m39+l]]})/1.0, {t, 0, plotend}, PlotRange \[Rule]
{0,1.1},PlotStyle\[Rule]{Red,Thick},AspectRatio\[Rule]1]
(*holo plot*)
p202=Plot[({p[t, p0][[m2+2*l]]}+{p[t, p0][[m3+2*l]]}+{p[t, p0][[m4+2*l]]}+{p[t, p0][[m5+2*l]]}+{p[t, p0][[m6+2*l]]}+{p[t,
p0][[m7+2*l]]}+{p[t, p0][[m8+2*l]]}+{p[t, p0][[m9+2*l]]}+{p[t, p0][[m10+2*l]]}+{p[t, p0][[m11+2*l]]}+{p[t,
p0][[m12+2*l]]}+{p[t, p0][[m13+2*l]]}+{p[t, p0][[m14+2*l]]}+{p[t, p0][[m15+2*l]]}+{p[t, p0][[m16+2*l]]}+{p[t,
p0][[m17+2*l]]}+{p[t, p0][[m18+2*l]]}+{p[t, p0][[m19+2*l]]}+{p[t, p0][[m20+2*l]]}+{p[t, p0][[m21+2*l]]}+{p[t,
p0][[m22+2*l]]}+{p[t, p0][[m23+2*l]]}+{p[t, p0][[m24+2*l]]}+{p[t, p0][[m25+2*l]]}+{p[t, p0][[m26+2*l]]}+{p[t,
p0][[m27+2*l]]}+{p[t, p0][[m28+2*l]]}+{p[t, p0][[m29+2*l]]}+{p[t, p0][[m30+2*l]]}+{p[t, p0][[m31+2*l]]}+{p[t,
p0][[m32+2*l]]}+{p[t, p0][[m33+2*l]]}+{p[t, p0][[m34+2*l]]}+{p[t, p0][[m35+2*l]]}+{p[t, p0][[m36+2*l]]}+{p[t,
p0][[m37+2*l]]}+{p[t, p0][[m38+2*l]]}+{p[t, p0][[m39+2*l]]})/1, {t, 0, plotend}, PlotRange \[Rule]
{0,1.1},PlotStyle\[Rule]{Green,Thick},AspectRatio\[Rule]1]
*)
```

```
(*sum of apo, term and holo finalstate*)
p203=Plot[({p[t, p0][[m39]]}+{p[t, p0][[m39+l]]}+{p[t, p0][[m39+2*l]]})/1, {t, 0, plotend},PlotRange ->
{0,1.1},PlotStyle->{black,Thick},AspectRatio->1]

EmitSound[Sound[{SoundNote["C"],SoundNote["G"],SoundNote["C5"]}]]
t2=SessionTime[];
TimeObject[]
t2-t1

(*variables of 2nd dimention of contourplot*)
(*!!!!!!! you have to rename the second dimention in plot p204,p205,p206 or it will be wrong with no error
message saying it is wrong!!!!!*)
y1=100*10^-6;
y2=100*10^-9;

t3=SessionTime[];
p204=ContourPlot[   ({p[t, p0][[1]]}+{p[t, p0][[m2]]}+{p[t, p0][[m3]]}+{p[t, p0][[m4]]}+{p[t, p0][[m5]]}+{p[t,
p0][[m6]]}+{p[t, p0][[m7]]}+{p[t, p0][[m8]]}+{p[t, p0][[m9]]}+{p[t, p0][[m10]]}+{p[t, p0][[m11]]}+{p[t, p0][[m12]]}+{p[t,
p0][[m13]]}+{p[t, p0][[m14]]}+{p[t, p0][[m15]]}+{p[t, p0][[m16]]}+{p[t, p0][[m17]]}+{p[t, p0][[m18]]}+{p[t,
p0][[m19]]}+{p[t, p0][[m20]]}+{p[t, p0][[m21]]}+{p[t, p0][[m22]]}+{p[t, p0][[m23]]}+{p[t, p0][[m24]]}+{p[t,
p0][[m25]]}+{p[t, p0][[m26]]}+{p[t, p0][[m27]]}+{p[t, p0][[m28]]}+{p[t, p0][[m29]]}+{p[t, p0][[m30]]}+{p[t,
p0][[m31]]}+{p[t, p0][[m32]]}+{p[t, p0][[m33]]}+{p[t, p0][[m34]]}+{p[t, p0][[m35]]}+{p[t, p0][[m36]]}+{p[t,
p0][[m37]]}+{p[t, p0][[m38]]}+{p[t, p0][[m39]]})/1, {t, 0, plotend},{conc,y1,y2},ContourStyle->None,ContourShading-
>Automatic,ColorFunction->"Rainbow",PlotLegends->BarLegend[{"Rainbow",{0,1}}],PlotRange->{0,1},Contours-
>20,ColorFunctionScaling->False]

p205=ContourPlot[({p[t, p0][[m2+l]]}+{p[t, p0][[m3+l]]}+{p[t, p0][[m4+l]]}+{p[t, p0][[m5+l]]}+{p[t, p0][[m6+l]]}+{p[t,
p0][[m7+l]]}+{p[t, p0][[m8+l]]}+{p[t, p0][[m9+l]]}+{p[t, p0][[m10+l]]}+{p[t, p0][[m11+l]]}+{p[t, p0][[m12+l]]}+{p[t,
p0][[m13+l]]}+{p[t, p0][[m14+l]]}+{p[t, p0][[m15+l]]}+{p[t, p0][[m16+l]]}+{p[t, p0][[m17+l]]}+{p[t, p0][[m18+l]]}+{p[t,
p0][[m19+l]]}+{p[t, p0][[m20+l]]}+{p[t, p0][[m21+l]]}+{p[t, p0][[m22+l]]}+{p[t, p0][[m23+l]]}+{p[t, p0][[m24+l]]}+{p[t,
p0][[m25+l]]}+{p[t, p0][[m26+l]]}+{p[t, p0][[m27+l]]}+{p[t, p0][[m28+l]]}+{p[t, p0][[m29+l]]}+{p[t, p0][[m30+l]]}+{p[t,
p0][[m31+l]]}+{p[t, p0][[m32+l]]}+{p[t, p0][[m33+l]]}+{p[t, p0][[m34+l]]}+{p[t, p0][[m35+l]]}+{p[t, p0][[m36+l]]}+{p[t,
p0][[m37+l]]}+{p[t, p0][[m38+l]]}+{p[t, p0][[m39+l]]})/1.0, {t, 0, plotend}, {conc,y1,y2},ContourStyle-
>None,ContourShading->Automatic,ColorFunction->"Rainbow",PlotLegends-
>BarLegend[{"Rainbow",{0,1}}],PlotRange->{0,1},Contours->20,ColorFunctionScaling->False]

p206=ContourPlot[({p[t, p0][[m2+2*l]]}+{p[t, p0][[m3+2*l]]}+{p[t, p0][[m4+2*l]]}+{p[t, p0][[m5+2*l]]}+{p[t,
p0][[m6+2*l]]}+{p[t, p0][[m7+2*l]]}+{p[t, p0][[m8+2*l]]}+{p[t, p0][[m9+2*l]]}+{p[t, p0][[m10+2*l]]}+{p[t,
p0][[m11+2*l]]}+{p[t, p0][[m12+2*l]]}+{p[t, p0][[m13+2*l]]}+{p[t, p0][[m14+2*l]]}+{p[t, p0][[m15+2*l]]}+{p[t,
p0][[m16+2*l]]}+{p[t, p0][[m17+2*l]]}+{p[t, p0][[m18+2*l]]}+{p[t, p0][[m19+2*l]]}+{p[t, p0][[m20+2*l]]}+{p[t,
p0][[m21+2*l]]}+{p[t, p0][[m22+2*l]]}+{p[t, p0][[m23+2*l]]}+{p[t, p0][[m24+2*l]]}+{p[t, p0][[m25+2*l]]}+{p[t,
p0][[m26+2*l]]}+{p[t, p0][[m27+2*l]]}+{p[t, p0][[m28+2*l]]}+{p[t, p0][[m29+2*l]]}+{p[t, p0][[m30+2*l]]}+{p[t,
p0][[m31+2*l]]}+{p[t, p0][[m32+2*l]]}+{p[t, p0][[m33+2*l]]}+{p[t, p0][[m34+2*l]]}+{p[t, p0][[m35+2*l]]}+{p[t,
p0][[m36+2*l]]}+{p[t, p0][[m37+2*l]]}+{p[t, p0][[m38+2*l]]}+{p[t, p0][[m39+2*l]]})/1, {t, 0, plotend},
{conc,y1,y2},ContourStyle->None,ContourShading->Automatic,ColorFunction->"Rainbow",PlotLegends-
>BarLegend[{"Rainbow",{0,1}}],PlotRange->{0,1},Contours->20,ColorFunctionScaling->False]

EmitSound[Sound[{SoundNote["C"],SoundNote["G"],SoundNote["C5"]}]]
t4=SessionTime[];
t4-t3

(*non time dependend contourplots*)
z1=100*10^-6;
z2=100*10^-9;
t=0.2;
```

5.2.12 Fragment-based screening of the iNEXT library

The pilM RNA was prepared as outlined in chaper 5.2.3 All additional RNAs were synthesized in-house using T7 RNAP for in vitro transcription (264). DNA templates were generated using PCR run-off or linearizing plasmids carrying the desired sequence, and both methods included the necessary T7 promotor. In vitro transcription was carried out at 10 to 20 ml scales based on the predicted yield, with transcription conditions tuned for yield and sample purity. Subsequently, the samples were buffer exchanged into an NMR buffer. If further purification was necessary due to observed RNA byproducts, preparative PAGE or HPLC preceded the buffer exchange step (84). Analytical PAGE and UV-Vis spectroscopy were used to determine sample concentration and purity throughout the preparation. The same transcription conditions, using 15N-isotopically labeled rNTPs, were used to generate 15N-uniformly labeled RNA for several of the follow-up tests.

Library design aspects of fragment pooling

Fragment library quality control was conducted by performing ^{1}D-, ^{1}H- and ^{19}F-spectra of each chemical individually. All fragments in the library used for screening were examined for discrepancies. This manual spectra analysis allowed substances with incorrect or extra signals to be ruled out.

The fragments were distributed across five pools. Each fragment pool contained 20 or 21 fragments in 90% d6-DMSO with 10% D2O. The concentration of each fragment in the pooled fragment stocks was 2.5 mM. The RNA concentration in the samples was around 50 µM aiming for a 1:1 ratio relative to each fragment. The pools were composed with minimal signal overlap utilizing the chemical shift data obtained during the initial quality control step. The screening samples were assembled into NMR tubes using the SamplePro Tube robot. The total sample volume, including 5% D2O serving as the locking solvent, was 170 µL. The screening buffer was 25 mM KPi at pH 6.2, 50 mM KCl, and 5 mM MgCl2 in an H2O, D2O, and d6-DMSO mix of 94%, 5%, and 1%, respectively. This was used for RNA and reference measurements. DNAs were measured at 25 mM KPi, pH 7.0, and 70 mM KCl, in the same H2O, D2O, and d6-DMSO mix. MptpA and PtkA buffer conditions were 25 mM and 50 mM HEPES/NaOH at pH 7.0 and pH 7.5 with 150 mM and 300 mM NaCl, both containing 10 mM DTT and 10 mM MgCl2. EphA2 buffer was 20 mM Tris at pH 8, 200 mM NaCl, 5 mM MgCl2, 3 mM TCEP, and T7 and rS1 buffer conditions were 25 mM KPi at pH 7.2, 150 mM KCl, and 5 mM DTT. The NMR screening data of 19F-1D and ^{19}F-CPMG-T2 measurements were acquired at mixing times of 0 ms, 200 ms, and 400 ms. The screening requires subsequent comparison of target and target-free measurements. Since the measurements of different targets contained the same or similar buffers, for protein targets, a single set of target-free pool measurements could be used as a shared reference, greatly reducing measurement effort and cost. The minor differences in the buffer of protein targets resulted in a 5% loss of possible targets because these could no longer be unambiguously assigned. This could be resolved through additional pool measurement in identical protein buffers.

Hit validation for a subset of targets

Three RNA targets under study were selected for further testing. To demonstrate the binding of the single strong binding fragment, samples of exclusively these fragments with the corresponding RNA were made using the same conditions and buffers. Some of these samples were subsequently used for competition experiments with native ligands. For competition experiments, an equimolar quantity amount of ligand compared to RNA was added.

In addition, samples of ^{15}N-labeled RNA with an RNA concentration of 1.25 mM and fragment concentrations of 50–100 mM were generated to enable RNA observed changes introduced by the fragment RNA interaction. Reference samples, in this case, contained no fragments.

NMR setup and methodology

A Bruker AVIIIHD-600 NMR spectrometer with a 5 mm 1H/19F [13C,15N]-TCI prodigy cryo-probe and a high throughput sample changer with active cooling for sample storage were used to acquire the data. Spectra were acquired: 19F-1D, water-suppressed proton 1D, and 19F-1D with CPMG spinlock (0, 200, and 400 ms). 19F spectra were recorded at 298 K, without 1H-decoupling, and subsequently processed with a line broadening function of 10 Hz. Adiabatic WURST pulses with a bandwidth of 120 ppm were applied for 2 ms to the CPMG spinlock. The Topspin wavemaker software determines the pulse during the measurement. The inter-pulse delay of the CPMG spin lock used in the screening was 9 ms. Screening data were acquired, processed, and collected using Topspin 4.0 (Bruker Biospin) with its integrated fragment-based screening software tool. Signal integration ranges were established on the target free reference samples and transferred to target measurements. Each hit was manually reviewed for deviations in peak positions. 5% of fragment signals with these deviations, observed for protein targets, were omitted. All remaining data was processed to yield Qbind coefficients (3.2.1 19F-CPMG screening of RNA and other macromolecules). Fragments were categorized based on Qbind into nonbinders (Qbind>0.66), weak binders (Qbind<0.66 and Qbind>0.33), and strong binders (Qbind<0.33).

Follow-up research was carried out for strong hits of three of the RNA targets, all known riboswitches. The same screening settings were used to perform 19F-1D-CPMG spectra with single fragment samples. Follow-up measurements included: 1H,1H-TOCSY with excitation sculpting and 15N-SFHMQC, proton 1Ds with either excitation sculpting(265) or jump-and-return echo(266) scheme for water suppression.

5.2.13 Biological context and available structural data of screened targets

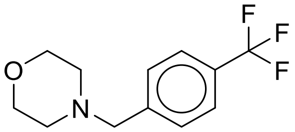
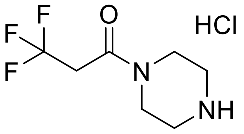
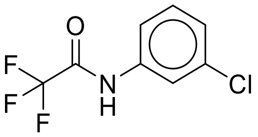
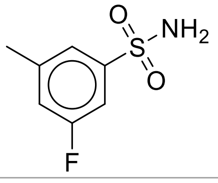
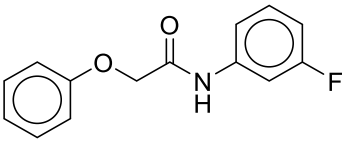
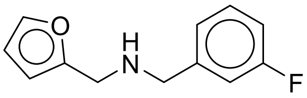
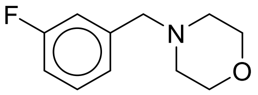
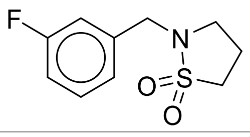
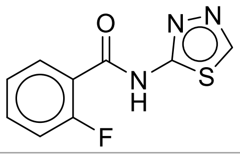
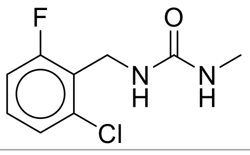
SI Table 3: List of all biomolecules used in the study listed with their biological host organism (if applicable), PDB accession codes of X-ray structures, and primary publication. Table as published in Binas *et al.* (2)**.**

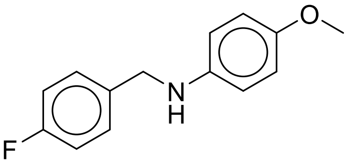
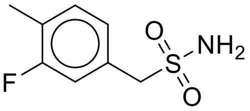
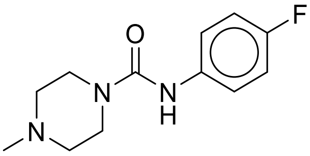
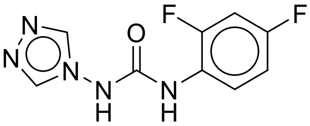
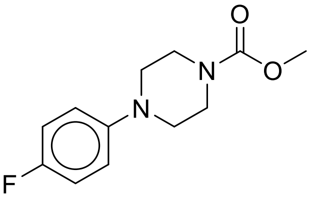
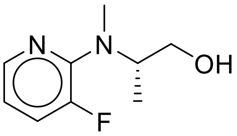
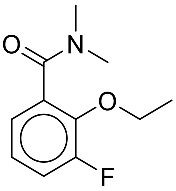
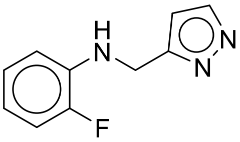
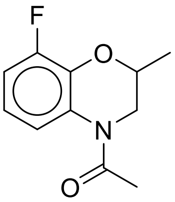
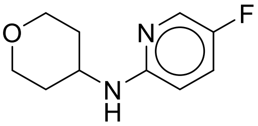
	Organism	X-Ray	NMR
Riboswitches and Aptamers			
Guanidine (Gdn-II)-sensing riboswitch (49 nt)	*Escherichia Coli*	5NDI(267)	
ZMP-sensing riboswitch (76 nt)	*Thermosinus carboxydivorans*	4ZNP(268)	
thiM TPP-sensing riboswitch (80 nt)	*E. Coli*	2GDI(269)	
pilM 3′, 3′-cGAMP-sensing riboswitch (84 nt)	*Geobacter metallireducens*	4YAZ(42)	
TenA TTP- sensing riboswitch (94 nt)	*Staphylococcus aureus*		
cyclic di-GMP-1 riboswitch (98 nt)	*Clostridium difficile*	3MXH*(81) 3IRW*(80) 3IWN*(270)	
Adenine-sensing riboswitch (127 nt)	*Vibrio vulnificus*	1y26$^{\#}$(271) 5E54$^{\#}$(73) 4TZX$^{\#}$(272)	(24) (264)
Riboswitch Elements			
2′-deoxyguanosine-sensing-riboswitch terminator (39 nt)	*Mesoplasma florum*		(53)
SAM-sensing riboswitch anti-terminator (38 nt)	*Bacillus subtilis*		
Adenine-sensing riboswitch terminator (51 nt)	*B. subtilis*		(273)
Adenine-sensing riboswitch expression platform (60 nt)	*V. vulnificus*		(264) (274)
Other RNAs			
RNA with GAAG tetraloop (14 nt)	*artificial*		2F87*(275)
RNA CUUG tetraloop (14 nt)	*artificial*		1RNG*(276)
Hammerhead ribozyme (54 nt)		1MME(277)	
Hepatitis delta virus ribozyme (70 nt)	*Hepatitis delta virus*	1DRZ(278)	
tRNA$^{\mathrm{fMet}}$ (77 nt)	*E. coli*	(279)	
DNA			
cMyc G-Quadruplex (22 nt)	*Homo sapiens*		1XAV(280) 2L7V(281)
cKit G-Quadruplex (24 nt)	*H. sapiens*	3QXR*(282)	2O3M(283)

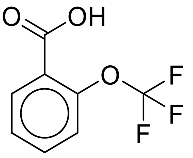
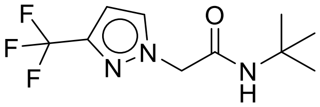
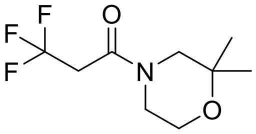
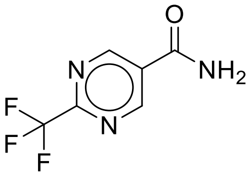
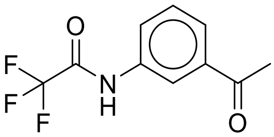
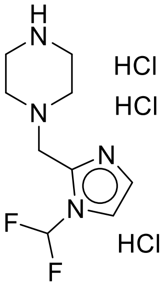
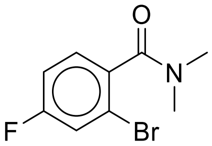
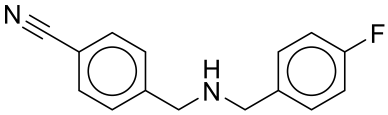
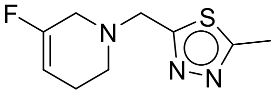
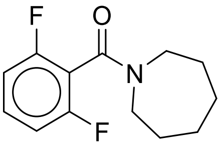
		2WO2	
DNA Duplex (24 nt)	*artificial*	1BNA(284)	
Tel26 G-Quadruplex (26 nt)	*H. sapiens*		2HY9(285) 5Z80(286)
wtTel26 G-Quadruplex (26 nt)	*H. sapiens*		5MVB(287) 2JPZ(288)
Proteins			
Mycobacterium tuberculosis Protein Tyrosine Phosphatase A (MptpA, 18 kDa)	*Mycobacterium tuberculosis*	1U2P(289)	2LUO(290)
Protein tyrosine Kinase A (PtkA, 30 kDa)	*M. tuberculosis*		6F2X(291)
Receptor tyrosine kinase EphA2 (34 kDa)	*H. sapiens*	5I9U(292)	
Ribosomal protein S1 (61 kDa)	*V. vulnificus*		2MFI*#(293) 2KHI*#(294)
T7 RNA polymerase (100 kDa)	*Escherichia phage T7*	1MSW(295)	

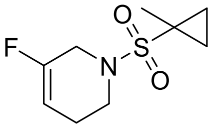
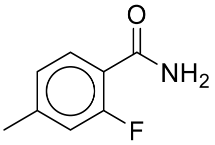
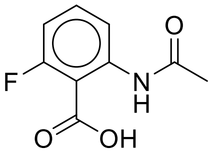
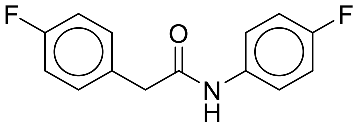
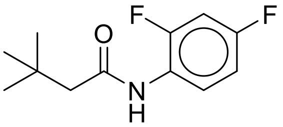
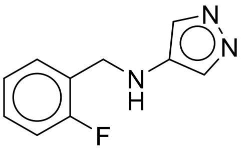
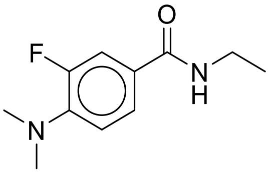
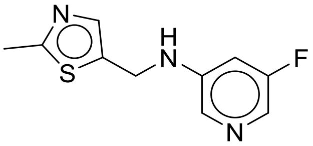
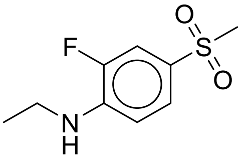
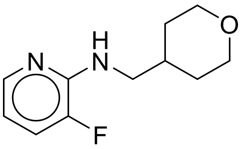
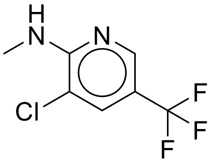
5.2.14 Fragment library

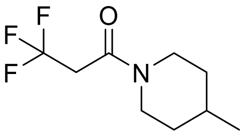
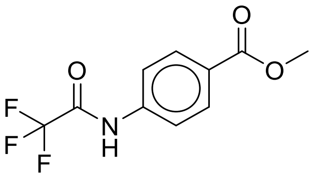
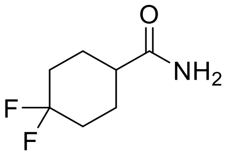
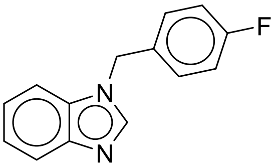
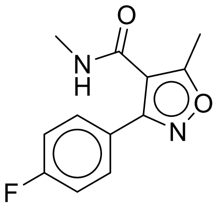
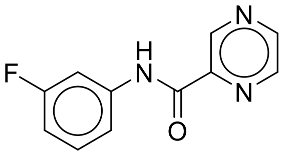
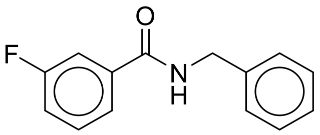
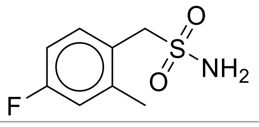
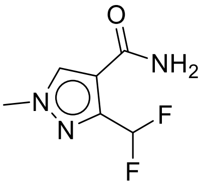
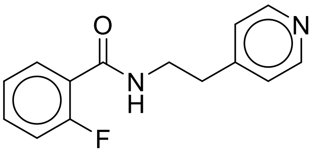
SI Table 4: ^{19}F-library for the ^{19}F-NMR-based fragment screening. Table as published in Binas *et al.* (2)**.**

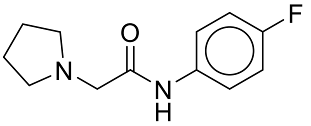
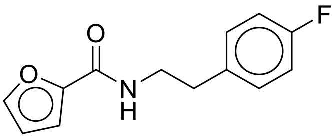
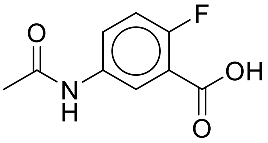
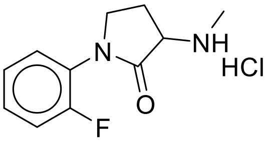
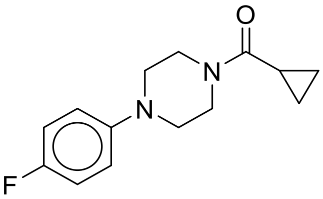
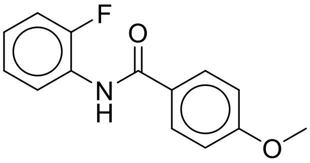
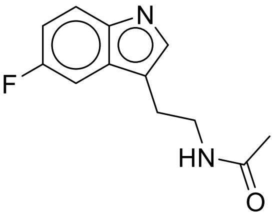
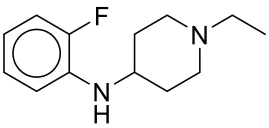
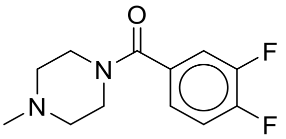
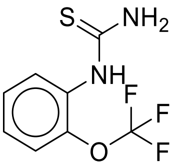
	SMILES code	**Formula**	**Chemical structure**
1.	FC(F)(F)c1ccc(cc1)CN2CCOCC2	C12H14F3NO	
2.	Cl.O=C(CC(F)(F)F)N1CCNCC1	C7H12ClF3N2O	
3.	Clc1cc(NC(=O)C(F)(F)F)ccc1	C8H5ClF3NO	
4.	Fc1cc(C)cc(c1)S(=O)(N)=O	C7H8FNO2S	
5.	Fc1cccc(c1)NC(=O)COc2ccccc2	C14H12FNO2	
6.	Fc2cccc(CNCc1ccco1)c2	C12H12FNO	
7.	Fc2cc(CN1CCOCC1)ccc2	C11H14FNO	
8.	O=S2(=O)CCCN2Cc1cccc(F)c1	C10H12FNO2S	
9.	O=C(Nc1nncs1)c2ccccc2F	C9H6FN3OS	
10.	Clc1cccc(F)c1CNC(=O)NC	C9H10ClFN2O	

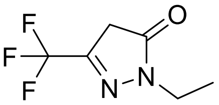
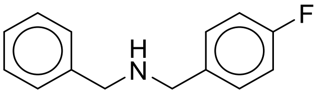
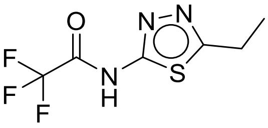
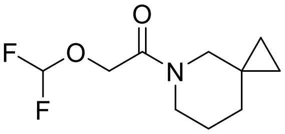
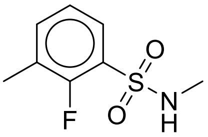
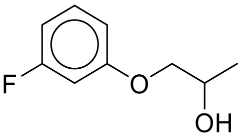
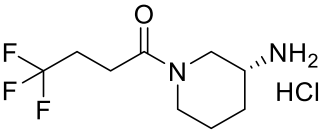
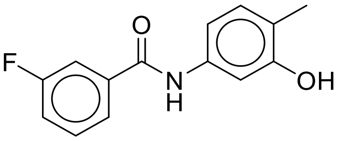
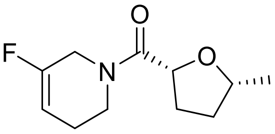
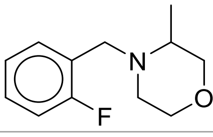
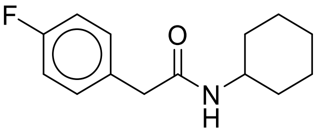
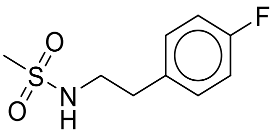
11.	COc2ccc(NCc1ccc(F)cc1)cc2	C14H14FNO	
12.	Cc1ccc(CS(N)(=O)=O)cc1F	C8H10FNO2S	
13.	O=C(Nc1ccc(F)cc1)N2CCN(C)CC2	C12H16FN3O	
14.	O=C(Nc1ccc(F)cc1F)Nn2cnnc2	C9H7F2N5O	
15.	COC(=O)N1CCN(CC1)c2ccc(F)cc2	C12H15FN2O2	
16.	Fc1cccnc1N(C)[C@@H](C)CO	C9H13FN2O	
17.	CCOc1c(cccc1F)C(=O)N(C)C	C11H14FNO2	
18.	Fc2ccccc2NCc1ccnn1	C10H10FN3	
19.	CC(=O)N2CC(C)Oc1c(F)cccc12	C11H12FNO2	
20.	Fc1ccc(nc1)NC2CCOCC2	C10H13FN2O	

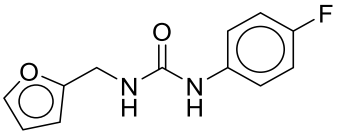
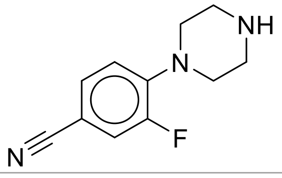
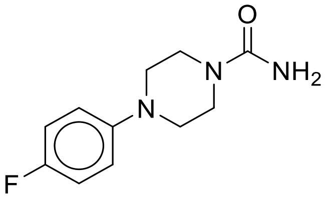
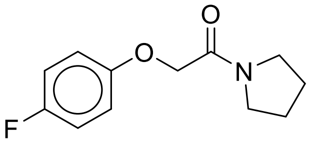
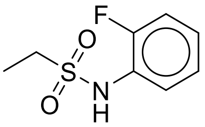
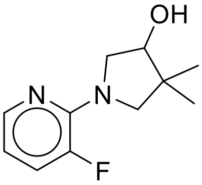
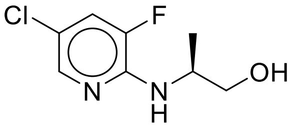
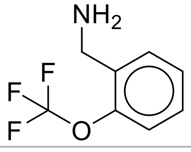
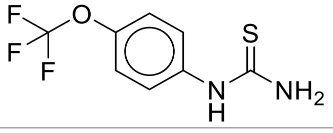
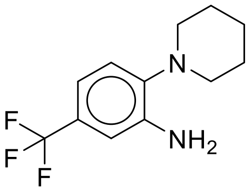
21.	FC(F)(F)Oc1ccccc1C(=O)O	C8H5F3O3	
22.	O=C(Cn1ccc(n1)C(F)(F)F)NC(C)(C)C	C10H14F3N3O	
23.	CC1(C)CN(CCO1)C(=O)CC(F)(F)F	C9H14F3NO2	
24.	FC(F)(F)c1ncc(cn1)C(N)=O	C6H4F3N3O	
25.	O=C(Nc1cc(ccc1)C(C)=O)C(F)(F)F	C10H8F3NO2	
26.	Cl.Cl.Cl.FC(F)n2ccnc2CN1CCNCC1	C9H17Cl3F2N4	
27.	O=C(c1ccc(F)cc1Br)N(C)C	C9H9BrFNO	
28.	Fc2ccc(CNCc1ccc(C#N)cc1)cc2	C15H13FN2	
29.	FC=1CN(CCC=1)Cc2nnc(C)s2	C9H12FN3S	
30.	O=C(c1c(F)cccc1F)N2CCCCCC2	C13H15F2NO	

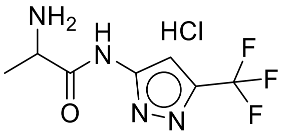
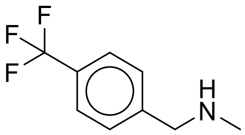
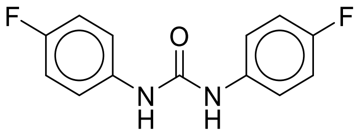
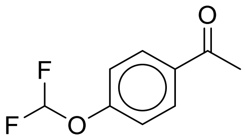
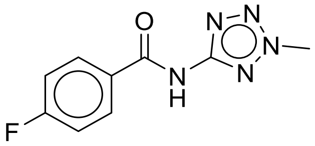
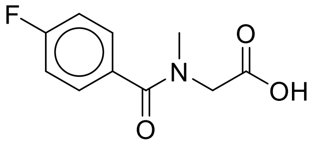
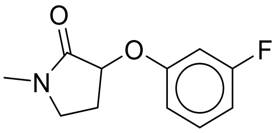
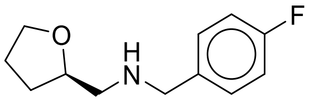
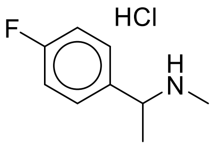
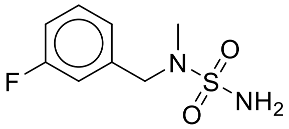
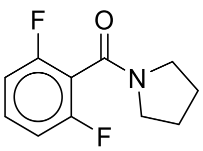
31.	FC=1CN(CCC=1)S(=O)(=O)C2(C)CC2	C9H14FNO2S	
32.	O=C(N)c1ccc(C)cc1F	C8H8FNO	
33.	O=C(C)Nc1cccc(F)c1C(=O)O	C9H8FNO3	
34.	O=C(Nc1ccc(F)cc1)Cc2ccc(F)cc2	C14H11F2NO	
35.	O=C(Nc1ccc(F)cc1F)CC(C)(C)C	C12H15F2NO	
36.	Fc2ccccc2CNc1cnnc1	C10H10FN3	
37.	Fc1cc(ccc1N(C)C)C(=O)NCC	C11H15FN2O	
38.	Cc2ncc(CNc1cc(F)cnc1)s2	C10H10FN3S	
39.	Fc1cc(ccc1NCC)S(=O)(C)=O	C9H12FNO2S	
40.	Fc2cccnc2NCC1CCOCC1	C11H15FN2O	
41.	Clc1cc(cnc1NC)C(F)(F)F	C7H6ClF3N2	

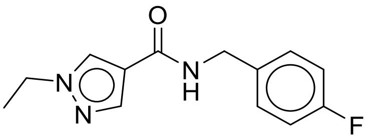
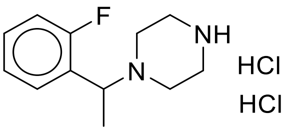
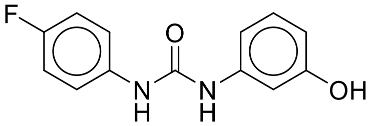
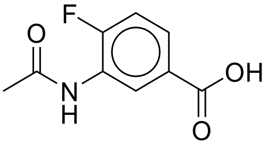
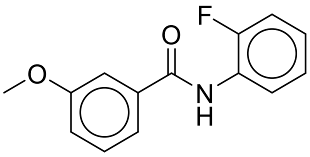
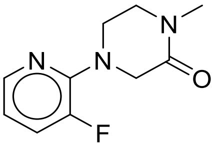
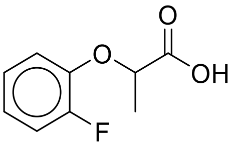
42.	CC1CCN(CC1)C(=O)CC(F)(F)F	C9H14F3NO	
43.	O=C(Nc1ccc(cc1)C(=O)OC)C(F)(F)F	C10H8F3NO3	
44.	FC1(F)CCC(CC1)C(N)=O	C7H11F2NO	
45.	Fc1ccc(cc1)Cn3cnc2ccccc23	C14H11FN2	
46.	CNC(=O)c2c(C)onc2c1ccc(F)cc1	C12H11FN2O2	
47.	O=C(Nc1cccc(F)c1)c2cnccn2	C11H8FN3O	
48.	O=C(NCc1ccccc1)c2cc(F)ccc2	C14H12FNO	
49.	Fc1cc(C)c(CS(N)(=O)=O)cc1	C8H10FNO2S	
50.	FC(F)c1nn(C)cc1C(N)=O	C6H7F2N3O	
51.	Fc1ccccc1C(=O)NCCc2ccncc2	C14H13FN2O	

52.	O=C(Nc1ccc(F)cc1)CN2CCCC2	C12H15FN2O	
53.	O=C(NCCc1ccc(F)cc1)c2ccco2	C13H12FNO2	
54.	O=C(C)Nc1cc(c(F)cc1)C(=O)O	C9H8FNO3	
55.	Cl.Fc1ccccc1N2CCC(NC)C2=O	C11H14ClFN2O	
56.	O=C(C1CC1)N2CCN(CC2)c3ccc(F)cc3	C14H17FN2O	
57.	COc1ccc(cc1)C(=O)Nc2ccccc2F	C14H12FNO2	
58.	CC(=O)NCCc2cnc1ccc(F)cc12	C12H13FN2O	
59.	CCN2CCC(Nc1ccccc1F)CC2	C13H19FN2	
60.	O=C(N1CCN(C)CC1)c2ccc(F)c(F)c2	C12H14F2N2O	
61.	FC(F)(F)Oc1ccccc1NC(N)=S	C8H7F3N2OS	

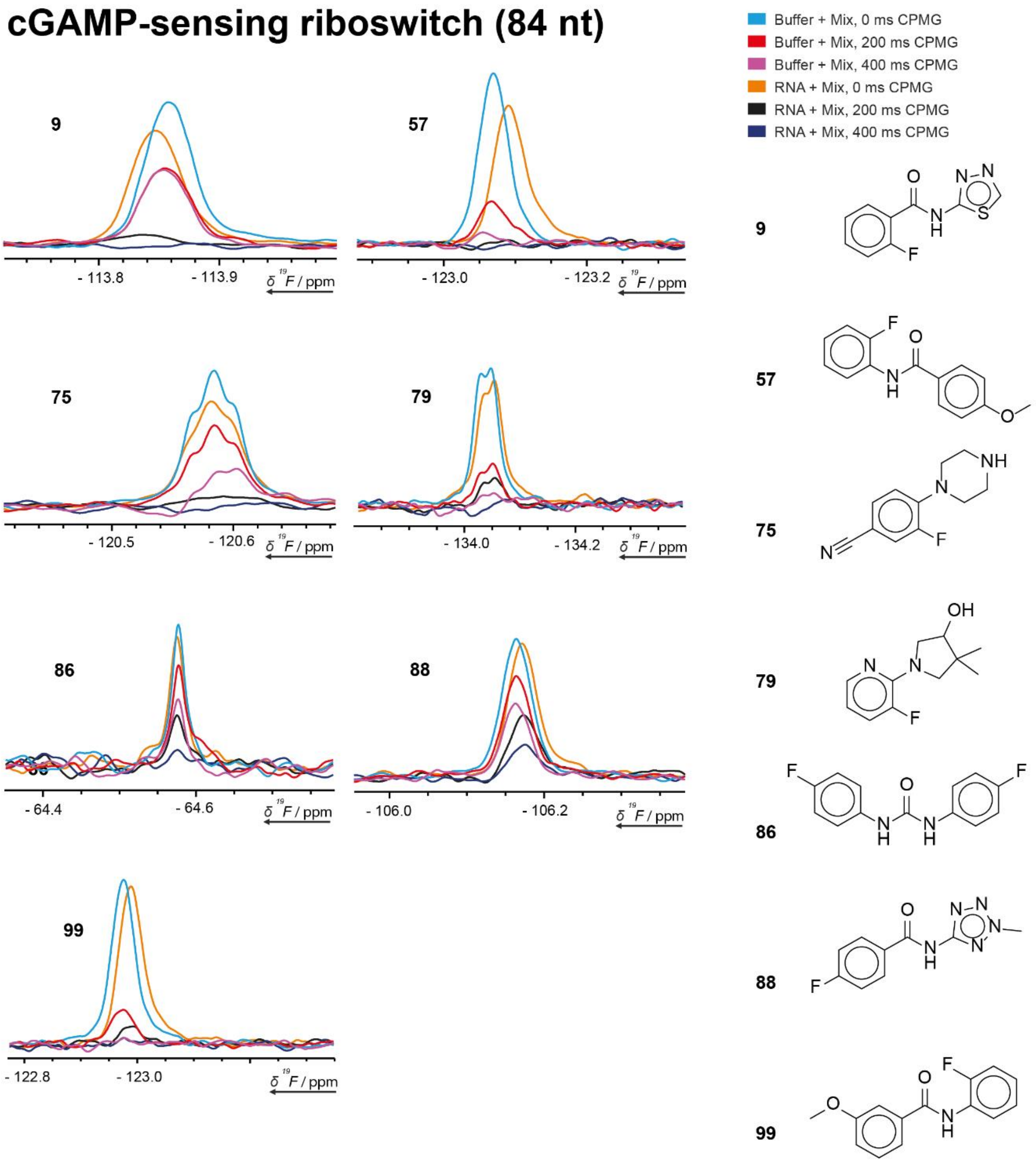
62.	FC(F)(F)C=1CC(=O)N(CC)N=1	C6H7F3N2O	
63.	Fc2ccc(CNCc1ccccc1)cc2	C14H14FN	
64.	O=C(Nc1nnc(CC)s1)C(F)(F)F	C6H6F3N3OS	
65.	FC(F)OCC(=O)N2CCCC1(CC1)C2	C10H15F2NO2	
66.	Fc1c(C)cccc1S(=O)(=O)NC	C8H10FNO2S	
67.	CC(O)COc1cccc(F)c1	C9H11FO2	
68.	Cl.N[C@@H]1CCCN(C1)C(=O)CCC(F)(F)F	C9H16ClF3N2O	
69.	Cc2ccc(NC(=O)c1cccc(F)c1)cc2O	C14H12FNO2	
70.	O=C(N1CC(F)=CCC1)[C@H]2CC[C@@H](C)O2	C11H16FNO2	
71.	Fc2ccccc2CN1CCOCC1C	C12H16FNO	
72.	O=C(NC1CCCCC1)Cc2ccc(F)cc2	C14H18FNO	
73.	Fc1ccc(CCNS(C)(=O)=O)cc1	C9H12FNO2S	

74.	O=C(NCc1ccco1)Nc2ccc(F)cc2	C12H11FN2O2	
75.	N#Cc1cc(F)c(cc1)N2CCNCC2	C11H12FN3	
76.	O=C(N)N1CCN(CC1)c2ccc(F)cc2	C11H14FN3O	
77.	O=C(COc1ccc(F)cc1)N2CCCC2	C12H14FNO2	
78.	O=S(=O)(Nc1ccccc1F)CC	C8H10FNO2S	
79.	Fc1cccnc1N2CC(C)(C)C(O)C2	C11H15FN2O	
80.	C[C@@H](CO)Nc1ncc(Cl)cc1F	C8H10ClFN2O	
81.	FC(F)(F)Oc1ccccc1CN	C8H8F3NO	
82.	FC(F)(F)Oc1ccc(cc1)NC(N)=S	C8H7F3N2OS	
83.	FC(F)(F)c1cc(N)c(cc1)N2CCCCC2	C12H15F3N2	

84.	Cl.FC(F)(F)c1cc(NC(=O)C(C)N)nn1	C7H10ClF3N4O	
85.	CNCc1ccc(cc1)C(F)(F)F	C9H10F3N	
86.	O=C(Nc1ccc(F)cc1)Nc2ccc(F)cc2	C13H10F2N2O	
87.	FC(F)Oc1ccc(cc1)C(C)=O	C9H8F2O2	
88.	O=C(Nc1nn(C)nn1)c2ccc(F)cc2	C9H8FN5O	
89.	O=C(c1ccc(F)cc1)N(C)CC(=O)O	C10H10FNO3	
90.	CN2CCC(Oc1cc(F)ccc1)C2=O	C11H12FNO2	
91.	Fc2ccc(CNC[C@H]1CCCO1)cc2	C12H16FNO	
92.	Cl.Fc1ccc(cc1)C(C)NC	C9H13ClFN	
93.	Fc1cc(CN(C)S(N)(=O)=O)ccc1	C8H11FN2O2S	
94.	O=C(c1c(F)cccc1F)N2CCCC2	C11H11F2NO	

95.	O=C(NCc1ccc(F)cc1)c2cn(CC)nc2	C13H14FN3O	
96.	Cl.Cl.CC(c1ccccc1F)N2CCNCC2	C12H19Cl2FN2	
97.	Oc2cccc(NC(=O)Nc1ccc(F)cc1)c2	C13H11FN2O2	
98.	O=C(C)Nc1cc(ccc1F)C(=O)O	C9H8FNO3	
99.	COc1cc(ccc1)C(=O)Nc2ccccc2F	C14H12FNO2	
100.	CN1CCN(CC1=O)c2ncccc2F	C10H12FN3O	
101.	CC(Oc1ccccc1F)C(=O)O	C9H9FO3	

5.2.15 Hit confirmation



SI Figure 6: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the pilM 3′,3′-cGAMP-sensing riboswitch (84 nt). For further hit validation, the CPMG experiment at 400 ms is shown. Figure as published in Binas *et al.*. Hit conformation plots for all other targets can be found in the SI of Binas *et al.* (2).

Acknowledgments

I thank Prof. Harald Schwalbe for allowing me to work in his research group. He allowed me to pursue my research with great freedom and to get it right. He provided an accommodating environment for all members of his group, allowing us to do science as equals among peers and learn every day. In doing so, he acted as a mentor in scientific excellence, compassion, and leadership. Danke AV.

I want to thank Tom Landgraf for his contributions to our shared publications and his support when writer's block was at its worst. I must thank Boris Fürtig for his aid that allowed me to build on his work in Markov modeling riboswitches.

I want to thank Oliver Binas, Tom Landgraf, and Vanessa de Jesus for their initiative and leadership in our ^{19}F-NMR screening project and Jason Martins, Daniel Hymon, Hannes Berg, Jasleen Kaur Bains, Thomas Biedenbänder, Boris Fürtig, Santosh Lakshmi Gande, Anna Niesteruk, Andreas Oxenfarth, Nusrat Shahin Qureshi, Tatjana Schamber, Robbin Schnieders, Alix Tröster, Anna Wacker, Julia Wirmer-Bartoschek, Maria Alexandra Wirtz Martin, Elke Stirnal, Kamal Azzaoui, Marcel Jules José Blommers, Christian Richter, and Sridhar Sreeramulu for their contributions.

I want to thank Bozana Knezic and Oliver Binas for their work on our book chapter on ribozymes.

I need to thank Christian Richter, Sridhar Sreeramulu, Martin Hengesbach, Elke Stirnal, Jan Ferner, Vanessa de Jesus and Boris Fürtig for their guidance, excellent expertise, and willingness to share it with someone who asks too many the difficult questions. I thank Robin Krishnathas and Elke Stirnal for sharing my responsibility in keeping part of our infrastructure running.

I am indebted to Jan Hering, Boris Fürtig, and Sven Warhaut for their mentorship and for leaving bright steps to flow. Kerstin Dathe and Daniela Dworak have my deep gratitude for their mastery of managing chaos with a human touch. I want to thank all members of the Schwalbe group for their friendship and kindness. And I cannot thank my family and spouse enough for their support and patience.

I witnessed Anna Paulus, Ernst Egert, and Robert Heide giving their whole life to our Chemistry. I will not forget them and their sacrifice.

Funds from DFG supported the work for the priority program 1879: ”Nucleotide Second Messenger Signaling in Bacteria”, and for the collaborative research center 902: “Molecular principles of RNA-based regulation”, by EU-funded project iNEXT-discovery. Work at BMRZ is supported by the state of Hesse. Boehringer Ingelheim Fonds supported Vanessa de Jesus. Fonds der Chemischen Industrie supported Robbin Schnieders.