

NMR-based structural and functional studies
of
RNAs and their complexes

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 15 – Biowissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Anna Katharina Weickhmann
aus Bad Homburg

Frankfurt 2021

(D 30)

vom Fachbereich 15 – Biowissenschaften der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Sven Klimpel

Gutachter: Prof. Dr. Jens Wöhnert

Zweitgutachterin: Prof. Michaela Müller-McNicoll, PhD

Datum der Disputation: 19.07.2022

This thesis is based on the following publication:

Wolter AC, **Weickmann AK**, Nasiri AH, Hantke K, Ohlenschläger O, Wunderlich CH, Kreutz C, Duchardt-Ferner E, Wöhnert J (2017) A Stably Protonated Adenine Nucleotide with a Highly Shifted pK_a Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. *Angew Chem Int Ed Engl.* 2017 Jan 2;56(1):401-404.

Keller H*, **Weickmann AK***, Bock T, Wöhnert J (2018) Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitch. *RNA.* 2018 Jul 13. (* equal contributions)

Weickmann AK, Keller H, Duchardt-Ferner E, Strebitzer E, Juen MA, Kremser J, Wurm JP, Kreutz C, Wöhnert J (2018) NMR resonance assignment for the SAM/SAH binding riboswitch RNA bound to S-Adenosylhomocysteine. *Biomol NMR Assign.* 2018 Jul 26.

Weickmann AK, Keller H, Wurm JP, Strebitzer E, Juen MA, Kremser J, Weinberg Z, Kreutz C, Duchardt-Ferner E, Wöhnert J. The structure of the SAM/SAH-binding riboswitch. *Nucleic Acids Res.* 2018 Dec 27.

CONTENT

I	Zusammenfassung.....	1
II	Introduction	7
1.	RNA – Beyond the Central Dogma of Biology	7
1.1.	RNA Fundamental Building Blocks.....	8
1.2.	RNA Structure Elements – The Building Blocks of Complex Structures	9
1.3.	Protonated Nucleobases Increase Number of Potential Interactions and Stability 12	
2.	Aptamers – The Proof that RNA Has High Versatility	15
2.1.	SELEX – Selecting Active RNA <i>in vitro</i>	15
2.2.	GTP Binding Aptamers – Many Different Ways to Recognize the Same Ligand	17
3.	Riboswitches – From <i>in vitro</i> to <i>in vivo</i> (or Theory Found in Nature)	22
3.1.	Challenges in Riboswitch Research.....	27
3.2.	SAM and SAH Binding Riboswitches	28
3.3.	Cyclic Dinucleotide Messenger c-di-GMP and cGAMP Binding Riboswitches	37
4.	ITC on RNA – Beyond Affinity Measurements.....	42
5.	RNA Structure Determination using NMR – For Well-Behaved and Challenging RNA .	44
5.1.	From Resonance Assignment to Structure – The Traditional Approach	45
5.2.	When Your RNA Doesn’t Behave as You Want it to	47
6.	Aim of this Work	47
III	Publications.....	49
1.	Wolter et al – Structure of the GTP Class II aptamer.....	49
2.	Keller et al - Adenine protonation for ligand promiscuity	54
3.	Weickmann et al – NMR resonance assignment of SAM/SAH riboswitch	68
4.	Weickmann et al – Structure of the SAM/SAH binding Riboswitch	75

IV	Discussion	88
1.	Structural Features for Stability and Ligand Binding	88
1.1.	Protonation as Structural Features.....	88
1.2.	Promiscuity.....	89
2.	SAM/SAH Binding Riboswitch.....	91
2.1.	Ligand Recognition Compared to Other SAM Binding Riboswitches	91
2.2.	Comparison NMR and Crystal Structure	92
3.	Conclusion	94
V	References	95
VI	Appendix.....	105
1.	Supplementary Information Wolter et al	105
2.	Supplementary Information Keller et al	129
3.	Supplementary Information Weickmann et al (structure)	140

I ZUSAMMENFASSUNG

Das zentrale Dogma der Biologie ist seit 1970 etabliert. Es beinhaltet die drei zentralen Makromoleküle einer Zelle: DNA als Trägerin der genetischen Information, RNA (messenger RNA, mRNA) als Botenmolekül und Protein als Prozessor aller Reaktionen in einer Zelle. Das Stigma des langweiligen Botenmoleküls hat RNA spätestens mit der Entdeckung ihrer katalytischen Eigenschaften verloren. So ist sie in der Lage sich selbst katalytisch zu spalten, wenn sie eine dafür günstige Konformation einnimmt. Es wurde außerdem gezeigt, dass RNA Liganden jeder Art, wie ganze Zellen, Metabolite oder gar Ionen, spezifisch und selektiv binden kann. Dies wurde zuerst in *in vitro* Experimenten gezeigt. Dafür wurden die Liganden auf einer Festphase immobilisiert und eine zufällige RNA Bibliothek auf ihre Bindungseigenschaften geprüft. Durch den iterativen Prozess, SELEX, kann schließlich im Idealfall mindestens eine RNA isoliert werden, die den Liganden affin und spezifisch bindet. Diese RNA nennt man Aptamer.

Die Idee des linearen Vorgangs, dass DNA in RNA und schließlich in Proteine umgeschrieben wird, trifft außerdem nicht nur im Labor, sondern auch in zellulären Prozessen nicht ausschließlich zu. In Bakterien wird häufig nicht die komplette mRNA in Proteine übersetzt. Dies ist für biologische Systeme, die normalerweise sehr effizient arbeiten, um unnötige Energieverschwendung zu vermeiden, ungewöhnlich. Es konnte gezeigt werden, dass sich gelegentlich vor den Genen regulatorische Einheiten befinden, die mit regulatorischen Proteinen interagieren und die Genaktivierung steuern. Bis heute gibt es Gensequenzen, denen keine regulatorischen Proteine zugeschrieben werden können und für die außerdem keine regulatorische Funktion bekannt ist. Wie bereits bei Aptameren gezeigt, ist RNA durchaus in der Lage, Liganden direkt zu binden. Diese Eigenschaft wurde bei bakterieller mRNA entdeckt, später auch in höheren Organismen. Diese Einheiten werden Riboswitches genannt und funktionieren wie Schalter, die lediglich in einer Richtung schalten. Sie befinden sich häufig in der 5' nicht-translatierten Region (UTR), selten auch in der 3' UTR, und setzen sich aus einer Aptamer-Domäne und einer Regulationsdomäne zusammen. Die Aptamer-Domäne bindet den Liganden, der oft in direkter funktioneller Verbindung zum darauffolgenden Gen steht. Die durch die Bindung induzierte Konformationsänderung führt dazu, dass auch die Regulationsdomäne ihre Konformation ändert. Man unterscheidet zwischen transkriptionellen und translationalen Riboswitches. Im ersten Fall enthält die

Regulationsdomäne die (Anti-)Terminatorstammhelix, welche sich als Reaktion auf die Bindung des Liganden an die Aptamer-Domäne ausbilden kann. Hierbei werden zwei Fälle unterschieden: Wird die Terminatorstammhelix ausgebildet, spricht man von einem OFF-Schalter (englisch *off*, aus), da die Transkription verhindert wird. Bei der Ausbildung der Anti-Terminatorstammhelix hingegen wird von einem ON-Schalter (englisch *on*, an) gesprochen, da die Transkription ermöglicht wird. Bei translationalen Riboswitches enthält die Regulationsdomäne die ribosomale Bindestelle (RBS) mit der Shine-Dalgarno-Sequenz. Die Bindung des Liganden in der Aptamer-Domäne kann bewirken, dass die RBS freigelegt wird und das Ribosom ungehindert binden kann. Die Translation kann stattfinden und man spricht auch hier von einem ON-Schalter. Im Gegensatz kann es aber auch dazu kommen, dass die RBS bzw. die Shine-Dalgarno-Sequenz in eine stabile 3D-Struktur integriert werden und damit die Bindung des Ribosoms an die mRNA verhindert wird. Diese Art von Riboswitch wird ebenso als OFF-Schalter bezeichnet. Oft lassen sich die Aptamer- und Regulationsdomäne nicht klar voneinander trennen, da sie sich überschneiden.

Eine große Gruppe von Aptameren sind die Guanosintriphosphat (GTP) Aptamere. Diese zeigt sehr eindrücklich, wie RNA trotz der limitierten Anzahl an Bausteinen unterschiedliche Strategien nutzt, um denselben Liganden zu erkennen. Es wurden elf verschiedene Klassen dieser Aptamere identifiziert, die sich in Sequenzkonservierung, Größe und Ligandenerkennungsstrategie maßgeblich unterscheiden. Bisher ist die 3D-Struktur einer dieser Klassen aufgeklärt, des GTP Klasse I Aptamer. Die Ligandenerkennungsstrategie eines weiteren wurde untersucht und es wurde gezeigt, dass das GTP Klasse V Aptamer den Liganden in einer G-Quadruplex-Struktur erkennt. Die komplette Struktur des GTP Klasse II Aptamers wird in der ersten Publikation gezeigt. Die vorhergesagte Stamm-Schlaufen-Sekundärstruktur bildet nach Ligandenbindung eine kompakte Struktur aus. Hierbei bildet sich, wie in der Sekundärstruktur vorhergesagt, eine Stammhelix aus. GTP wird in der apikalen Schlaufe durch ein Basentriplet gebunden. Interessanterweise zeichnet die Struktur ein stabil protoniertes Adenin unterhalb der GTP-Bindestelle aus. Dieses wurde durch eine Kombination aus weiterführenden NMR- und ITC-Experimenten untersucht und charakterisiert. Es zeigte sich, dass die protonierte Base einen pK_s -Wert hat, der weit von der Neutralität verschoben ist. Die Protonierung ist auch noch bei sehr basischen Puffern stabil. Die Protonierung von Nukleobasen ist bisher selten beschrieben und wird daher in strukturellen Betrachtungen auch selten berücksichtigt. Erschwerend hinzu kommt, dass Kristallstrukturen die Position von

Wasserstoffatomen nicht auflösen und diese schließlich lediglich in die Struktur modelliert werden. Wasserstoffbrückenbindungen werden auf Grundlage von räumlicher Nähe angenommen. Diese Vorgehensweise führt dazu, dass die Protonierung von Nucleobasen lange nicht berücksichtigt wurde. Die Struktur des GTP Klasse II Aptamers zeigt eindrücklich ein Beispiel für strukturelle Protonierung, bei der RNA in der Lage ist die Protonierung weitab der Neutralität zu stabilisieren und damit eine stabile Architektur auszubilden.

Protonierungen von Nucleobasen wurden im natürlichen Kontext z.B. im Zusammenhang mit dem Luteovirus P1-P2 *frameshifting* (englisch *frame*, Rahmen und *shifting*, verschieben), bei der Kontrolle der Konformationsänderungen im Spliceosom oder bei der katalytischen Selbstspaltung des HDV-Ribozym beschrieben. Diese Art der funktionellen Protonierung wird auch von den zyklischen di-Nukleotiden (CDN) bindenden Riboswitches genutzt, um zwei CDN mit ähnlicher Affinität zu binden. CDN gehören zu den sekundären Signalmolekülen. Dabei handelt es sich um zwei Nukleotide, die über einen der Zuckeralkohole (2' oder 3') des einen Nukleotides zur Phosphatgruppe des anderen Nukleotides und umgekehrt miteinander verbunden sind. Kombinatorisch sind eine Reihe von verschiedenen zyklischen di-Nukleotiden denkbar, bekannt sind bisher c-di-GMP (bestehend aus zwei zyklisch verknüpften Guanosenen, G_{α} und G_{β}), c-di-AMP (bestehend aus zwei zyklisch verknüpften Adenosinen) und cGAMP (bestehend aus je einem zyklisch verknüpften Guanosin und Adenosin). Bakterielle CDN sind 3'-3' verknüpft und es wurden bisher alle drei Kombinationen beschrieben. Im Gegensatz dazu sind eukaryotische CDN 3'-2' verknüpft und es wurde bisher nur ein cGAMP Isomer beschrieben. In ihrer biologischen Funktion werden die bakteriellen Isomere im Kontext mit Biofilmbildung, Beweglichkeit und Virulenz beschrieben, während die eukaryotischen als Agonisten der angeborenen Immunantwort beschrieben wurden. Die genauen regulatorischen Zusammenhänge sind noch nicht vollständig aufgeklärt. c-di-GMP Riboswitches wurden als regulatorische Einheit beschrieben und deren Kristallstruktur aufgeklärt. Der Riboswitch bildet im Ligand gebundenen Zustand eine H-förmige Struktur aus, in deren Zentrum der Ligand eingebettet wird. Hierbei wird das eine G_{α} durch ein Guanosin des Riboswitches und das andere G_{β} durch ein Watson-Crick Basenpaar erkannt. Mutationsexperimente führten dazu, dass bei einer G-zu-A Mutation an der G_{α} -Bindestelle die Selektivität des Riboswitches verändert wurde. Die Mutante bindet sowohl c-di-GMP als auch cGAMP mit ähnlichen Bindungsaffinitäten. Riboswitches, die cGAMP binden wurden auch in bakteriellen Genomen gefunden. Hierbei ist die Promiskuität unterschiedlich stark

ausgeprägt. Die Untersuchung des Bindungsmodus und der damit verbundenen Promiskuität ist in der zweiten Publikation beschrieben. Hier wurde gezeigt, dass die Riboswitches beide Liganden nur binden können, wenn zur Bindung von c-di-GMP das Ligand bindende A protoniert vorliegt. Auch diese Protonierung konnte mit weiterführenden NMR- und ITC-Experimenten charakterisiert werden. Die Untersuchungen einer solch großen RNA sind mit NMR-Spektroskopie herausfordernd. Hierbei wurde ausgenutzt, dass die Kristallstruktur bereits bekannt war, welche allerdings auch den oben genannten Gründen die Protonierung nicht zeigte. Auch diese Protonierung zeigt einen pK_s -Wert, der weit von der Neutralität verschoben ist und außerdem bei unterschiedlichen pH stabil ist.

In den beiden untersuchten Beispielen wurden zwei verschiedene Arten von Protonierung gezeigt: eine strukturelle und eine funktionelle. Das GTP Klasse II Aptamer benutzt die Protonierung als strukturelle Basis für stabile Ausbildung der Ligandenbindungsstelle. Hierbei werden durch die Protonierung des Adenins mehr nutzbare Wasserstoffbrücken ausgebildet und damit die Tertiärstruktur stabilisiert. Im Unterschied dazu nutzen die promiskuitiven CDN Riboswitches die Protonierung, um verschiedene Liganden binden zu können und es kommt damit zu einer Verschiebung der Funktionalität. Der regulatorische Nutzen ist allerdings noch unbekannt.

Auch bei den SAM Riboswitches wurde ein promiskuitiver Vertreter beschrieben. SAM Riboswitches gehören zu den am längsten bekannten Klassen der Riboswitches. Bis heute sind hier die meisten unterschiedlichen Klassen bekannt. SAM wird häufig als Donor für funktionelle Gruppen benutzt, besonders häufig aber nicht ausschließlich als Methylgruppendonor für die Methylierung einer Reihe unterschiedlicher Substrate (z.B. DNA, Proteine, Metabolite etc.). Bei dieser Reaktion entsteht SAH als Nebenprodukt, welches sich nur in der fehlenden (transferierten) Methylgruppe und einer fehlenden Ladung am Sulfonium von SAM unterscheidet. Zusätzlich ist SAH zelltoxisch, da es affin an Methyltransferasen bindet und damit diese essenzielle Reaktion inhibiert. Eine enge Kontrolle der SAH-Konzentration ist daher kritisch. SAM bindende Riboswitches haben zu SAM eine bis zu 1000-fach höhere Bindungsaffinität im Vergleich zu SAH. Die Beschreibung eines translationalen OFF-Riboswitches, der SAM und SAH mit ähnlicher Affinität bindet, ist daher überraschend. Zumal seine Genassoziation fast ausschließlich zu Methyltransferasen ist, deren Regulation durch SAH wenig sinnvoll erscheint. Um ein besseres Verständnis für die Funktion des SAM/SAH Riboswitches zu erhalten, wurde seine 3D-Struktur mittels NMR-Spektroskopie

aufgeklärt, wie in der vierten Publikation beschrieben. Dafür mussten zunächst alle NMR-Resonanzen der Sequenz und des Liganden zugeordnet werden, wie in der dritten Publikation beschrieben. Dabei wurde als Ligand SAH gewählt, da dieser chemisch stabiler und damit für die teils tagelangen NMR-Messungen besser geeignet ist. Zusätzlich wurden Mutanten bzw. verwandte Liganden mittels ITC-Experimenten auf ihre Bindungseigenschaften untersucht, um die Bedeutung der Linkerlänge, einzelner Basenpaare und funktioneller Gruppen des Liganden zu untersuchen. Bei anderen bekannten SAM Riboswitches umschließt die RNA den Liganden fast komplett. Dabei wird zum einen das Sulfoniumion spezifisch durch die Carboxylgruppen verschiedener Uracil-Nukleotide koordiniert und erkannt. Außerdem bildet sich eine Bindetasche aus, die genug Platz für die stabile Bindung der Methylgruppe hat. Beim SAH Riboswitch wird die Selektivität für SAH dadurch erreicht, dass die Bindetasche sterisch keinen Platz für die Methylgruppe von SAM bereitstellt. Die Struktur des SAM/SAH Riboswitches zeigt deutlich, dass die SAH-Adeningruppe spezifisch durch ein reverses Hoogsteen-Basenpaar zu einem Uracil erkannt wird. Es bilden sich zusätzlich einige Wasserstoffbrückenbindungen aus, die den Zuckerrest des Liganden stabilisieren. Der Aminocarboxypropyl-Rest des Liganden wird allerdings nicht mehr spezifisch erkannt und die Struktur zeigt, dass die RNA den Liganden nicht komplett umschließt. Die Genassoziation mit überwiegend Methyltransferasen lässt darauf schließen, dass der SAM/SAH Riboswitch einen Minimalerkennungsmodus für SAM zeigt und die Bindung von SAH mit ähnlicher Bindungsaffinität ein Zufall ist. Dieser Modus ergibt im biologischen Kontext nur Sinn, wenn die intrazelluläre SAH-Konzentration durch andere regulatorische Mechanismen niedrig gehalten wird. Eine bioinformatische Suche ergab darauf, dass die Organismen, die einen SAM/SAH Riboswitch kodieren außerdem meistens auch einen SAH-Degradationsweg kodieren. Kurz nach der Publikation der NMR-Struktur wurde außerdem eine Kristallstruktur des SAM/SAH Riboswitches u.a. gebunden an SAM veröffentlicht. Ein Vergleich beider Strukturen zeigt, dass der Riboswitch tatsächlich SAM und SAH mit fast identischer 3D-Struktur bindet. Eine Selektion zwischen den beiden Liganden ist daher in der Tat nicht möglich. Die Kristallstruktur unterscheidet sich von der NMR-Struktur nur in der Lage eines Adenins. Dieses Adenin bildet in der NMR-Struktur ein A-Basentriplet aus, welches sich unterhalb der Bindetasche befindet. In Gegensatz wird in der Kristallstruktur zwar ein äquivalentes Basentriplet ausgebildet, wobei gesagtes Adenin jedoch durch den Adeninrest

eines zweiten SAH ersetzt wird. Der Unterschied ist höchstwahrscheinlich auf Kristallpackungseffekte zurückzuführen und daher auch wahrscheinlich nicht physiologisch. Zusammenfassend wurden in dieser Arbeit drei verschiedene Ligand bindende RNA-Strukturen untersucht, die alle sehr unterschiedliche Strategien zur Bindung der Liganden nutzen. Obwohl Portionierungen bei Aptameren und Riboswitches selten beschrieben wurden, haben sie eine maßgebliche Funktion in den beiden oben beschriebenen Strukturen. Obwohl bisher im Hinblick auf alle bekannten RNA Strukturen eher selten beschrieben, gibt es doch, neben den genannten zwei, einige Beispiele für strukturelle oder funktionelle Protonierungen. Auch in Hinblick auf zukünftige bzw. Verbesserung bestehender RNA-Strukturvorhersage-Programme, welche für Proteine schon lange genutzt werden, müssen protonierte Nucleobasen ernsthaft in Betracht gezogen werden. Außerdem konnte gezeigt werden, dass zwei der untersuchten Riboswitches zwei Liganden mit ähnlicher Affinität binden. Die genutzte Strategie ist hierbei unterschiedlich. Während bei den promiskuitiven CDN Riboswitches der regulatorische Nutzen noch unbekannt ist, konnte für den SAM/SAH Riboswitch gezeigt werden, dass SAH nur zufällig aufgrund der wahrscheinlich sehr niedrigen intrazellulären SAH-Konzentration gebunden wird und dieser daher wahrscheinlich später in der evolutionären Entwicklung entstanden ist.

Riboswitches halten es weiterhin spannend.

II INTRODUCTION

1. RNA – BEYOND THE CENTRAL DOGMA OF BIOLOGY

The three central macromolecules in a cell – DNA, RNA and protein – were put in context by Francis Crick in 1958 (updated 1970) when he proposed the “Central Dogma of Biology” (Crick 1970). He states that DNA, the carrier of all genetic information, is copied into RNA. The produced RNA (messenger RNA, mRNA) is then translated into proteins which perform all biochemical reactions in a cell. In this dogma, RNA is described to be solely a mediator molecule in protein synthesis. Since RNA was perceived as a mediator molecule, it remained in the shadow of protein research. Research interests focused on DNA as the carrier of genetic information and proteins as the actors in all chemical reactions in all living cells.

Furthermore, mRNA was assumed to be mainly single stranded and therefore of limited interest for (structural) biology, but tRNAs (transfer RNA) were known to be large and structured since the late 1950s. tRNAs are key players in proteins synthesis. Their properties fit into the picture of RNA being a mediator molecule with the sole function of decoding the information on the mRNA for the transfer of amino acids to a growing peptide chain. The first sequence and an proposal for a secondary structure of yeast alanine tRNA was published in 1965 by Holley (Holley et al. 1965), who was awarded the Nobel prize in Medicine in 1968 for it. With that, the interest in RNA and their 3D structures grew. Hence, between 1971 and 1976 further structures of tRNAs were solved (yeast tRNA^{Phe} at 3 Å (Kim et al. 1971), several E. coli tRNA species at 7 Å (Brown et al. 1972), yeast tRNA^{Phe} at 4 Å (Kim et al. 1973), yeast tRNA^{Phe} at 3 Å (Robertus et al. 1974)).

In 1962 Alexander Rich already proposed the idea of an RNA world (Rich et al. 1962), claiming that the earliest forms of life used RNA as its only macromolecule instead of DNA and proteins. Only five years later, Carl Woese proved that RNA possesses catalytic properties (Woese 1967). RNA has the intrinsic property to deprotonate its 2'-OH group and when in an optimal geometry undergo a nucleophilic attack resulting in a self-cleavage of the RNA backbone. This showed that RNA holds the potential for a catalytic activity. It took until the early 1980s when it was shown that RNAs are indeed catalytically active *in vivo*, like proteins. RNase P (Guerrier-Takada et al. 1983) and self-splicing intron I (Garriga et al. 1986; Garriga and Lambowitz 1986) were described – two large, structured RNAs that are catalytically active. Altman and Cech

were rewarded with the Nobel Prize in Chemistry in 1989 for their discovery. Many more catalytically active RNAs, called ribozymes, and other classes of functional RNAs have been found since (Doherty and Doudna 2000) enlightening the potential of RNA and destroying the superstition of a (boring) mediator molecule.

Additionally, the discovery that the ribosome is the largest RNA enzyme (Dahlberg 1989), again enhanced the importance of RNA in the process of protein synthesis. Even though RNA is limited to its four building blocks, it was found to form a magnificent number of intriguing structural interaction and act as an important player in the orchestra of biochemical reactions and interactions in *in vivo*. With the growing number of functional RNAs and their structures solved, more and more features of RNA were found. Not only the tRNA structures elucidated the high potential of RNA to fold into ordered complexity, but also the growing numbers of RNA classes interacting in the jungle of gene expression, editing and modification enhances the biochemical potential and importance of RNA.

1.1. RNA FUNDAMENTAL BUILDING BLOCKS

RNAs are polynucleotides which compose of the four nucleotides that are connected in a 5'-3' direction (Figure 1). In each nucleotide the ribose ring containing five carbon atoms is connected to the nucleobase via a N-glycosidic bond from N1/N7 of the nucleobase to the ribose C1' atom. RNA contains four nucleobases: adenine (ADE, A), guanine (GUA, G),

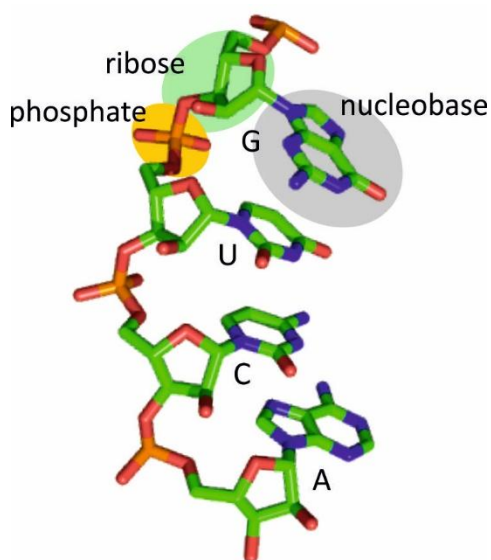


Figure 1: RNA strand with all four nucleotides. The nucleotide's ribose is marked in green, while the nucleobase is marked in grey. The phosphate is marked in orange (from structure 6HAG).

cytosine (CYT, C) and uracil (URA, U) - the former two being purines and the latter being pyrimidines. The phosphate group is connected to the sugar via an ester bridge to the C5' atom of its own nucleotide's ribose and the hydroxyl group at the C3' atom of the following ribose.

In contrast to RNA, DNA uses thymidine instead of uracil. Additionally, DNA lacks the sugar 2'OH group and is thus called a deoxyribose. These small differences result in great structural and functional differences between the two macromolecules. While DNA is chemically stable due to the deoxyribose, RNA can catalyze self-splicing reactions due to the nucleophilic attack of the potentially deprotonated

2'OH group. Additionally, the sugar conformation is influenced by the presence or absence of the 2'OH group. While the deoxyribose is mainly found in a 2'-endo conformation, the ribose sterically favors a 3'-endo conformation. Finally, the ribose 2'OH group is a putative hydrogen bond donor and acceptor, enabling RNA to form more hydrogen bonds than DNA.

1.2. RNA STRUCTURE ELEMENTS – THE BUILDING BLOCKS OF COMPLEX STRUCTURES

RNA is highly diverse in the usage of the donor and acceptor properties of the nucleobase or ribose to form immensely complex structure elements. Like DNA, RNA forms Watson-Crick double helices between two antiparallel strands by forming the canonical (Watson-Crick) A-U and G-C base pairs. A-form helices are sterically favored by RNAs due to the 3'-endo ribose conformation, while DNA mainly forms B-form helices. The A-form helix has 11 base pairs per turn, a rise of 2.8 Å and a tilt of 22°. The major groove of the A-form helix is deep and narrow, while the minor groove is shallow and wide. At high salt concentrations (>20%) an A'-helix is formed which only has 12 base pairs per turn resulting in a stretched helix comparable to Z-DNA.

The simplest form of a non-canonically structured RNA is the so-called hairpin (Figure 2A), where the 5'- and 3'-end of the RNA strand base pair to form a stem helix which is closed by

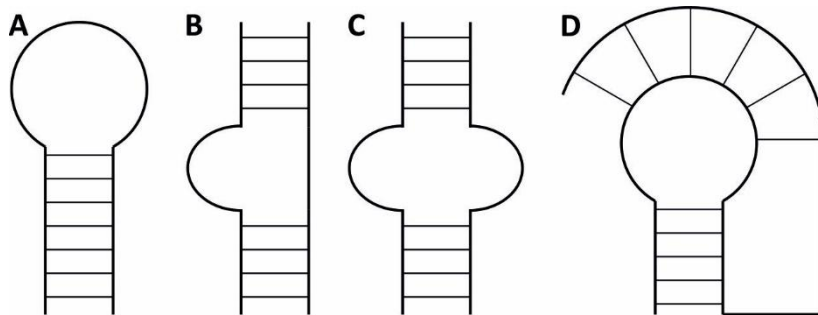


Figure 2: Secondary structure elements. A shows an hairpin, B a bulge and C an internal loop. D shows a schematic pseudoknot secondary structure.

a stretch of unpaired nucleotides that form a loop. Very common is a loop of four nucleotides (see tetraloops). Those regular elements are often

interrupted by bulges or internal loops of unpaired

RNA nucleotides, the former describing a stem structure that is interrupted by unpaired nucleotides in only one strand (Figure 2B), while the latter contains unpaired nucleotides in both strands (Figure 2C). In a more complex case, those bulges and internal loops could hold additional stem structures and hereby forming a junction where three or more stems are connected by several unpaired nucleotides.

Due to the high number of possibilities to interact in complex structure elements, RNA has the potential to form very intriguing structures with many possible applications in nature or

biotechnology. The combination and expansion of the described patterns give RNAs the possibility to form even more complex structure elements.

Tetraloops are a special kind of hairpin structures and compose of a stem helix capped by a loop of four unpaired nucleotides. The sequences and structures of the tetraloop loop region are highly conserved and often serve to initiate RNA-folding or as tertiary anchoring points within or between RNAs (tetraloop-receptor). Interestingly, about half of the hairpin loops in 16S-rRNA are tetraloops (Woese et al. 1990). In 16S-rRNA, GNRA- and UNGC-tetraloops are most often found. The tetraloops are named by the consensus sequence of the four nucleotides in the loop (N = any nucleotide, R = purines). Interestingly, the GNRA tetraloop in contrast to the UNGC tetraloop was found to mediate tertiary interactions, so-called tetraloop-receptor interactions. The UNCG tetraloop family which comprises the most stable tetraloop structures, on the other hand is thought to nucleate folding. Additionally, the CUUG builds the third abundant class of tetraloops and is the most flexible of the three (Jucker and Pardi 1995).

U-turns were first observed in the tRNA-anticodon loop (Quigley and Rich 1976) and are a very common motif for reversal of chain direction by inducing a 180° turn. While the loop length is variable, the U-turns are characterized by a U H3N3 forming a long-range hydrogen bond to the phosphate of residue +3. Additionally, the 2'OH of the same U forms a second long-range interaction to N7 of purine +2. A UNR consensus sequence can be derived.

Kissing loops are two hairpin structures that interact with its loops to form stable long-range interactions between RNA strands. These interactions are mediated by canonical base pairing of at least two nucleotides of the interacting hairpin loop regions. To date, kissing loops have been found in many RNAs, e.g. tRNA, mRNA, and rRNA (Chang and Tinoco 1994; Scarabino et al. 1999; Tok et al. 2000).

Coaxial Stacking describes structures where two or more discontinuous stems are packed on each other to form a helix structure. The individual helices can be interrupted by non-canonical base pairs, unpaired nucleotides or whole loops and junctions.

Pseudoknots are hairpins whose loop interact with a 5'- or 3'-end of the RNA strand to form three new loops and a second helix (Figure 2D). The additional helix can coaxially stack on the first when one loop is eliminated. The loops connecting the two helices can either be flexible and unpaired or form tertiary non-canonical base pairs to one of the helices for further stabilization. Those structures seem to be the simplest way to form stable three-dimensional

structures and are associated with the assembly of ribonucleoprotein complexes, translation regulation (IRES (Otto and Puglisi 2004; Pfingsten et al. 2006, 2007), also see riboswitches), programmed ribosomal frameshifting (e.g. HIV or Beet western yellow virus mRNAs (Farabaugh 1996; Nixon and Giedroc 2000), stop codon redefinition (Howard et al. 2005) and rescue of stalled ribosomes at the 3'-end of mRNAs without a termination codon (Haebel et al. 2004; Moore and Sauer 2005, 2007; Nonin-Lecomte et al. 2006). A common feature of pseudoknots are bends in the coaxially stacked helices which are caused by additional unpaired nucleotides. This feature was found to be essential for *in vivo* function (Chen et al. 1996).

Ribose zipper describes the hydrogen bond interactions of the 2'-OH group of a ribose that stabilizes two separate RNA strands and hold them together. Since the 2'-OH group can act as a donor and acceptor, it is able to form a bifurcated hydrogen bond to a 2'-OH of the other strand. The ribose zipper is an element of GAAA (special case of an GNRA tetraloop) tetraloop-tetraloop receptor interaction. Here, the RNA backbones of two helices are brought together, allowing a ribose of the antiparallel strand to form the typical bifurcated hydrogen bond.

Minor groove triplets – A minor interactions are characterized by an unpaired A residue that interacts with the minor groove of a G:C-rich helix and is stabilized by van der Waals forces and hydrogen bonds. Depending on the interaction pattern, these interactions can be classified into three groups (Type 0-III). Interestingly, conserved adenines are more abundant in non-helical regions than other nucleotides, making minor groove triplets probably the most important structural motif for stabilizing the 23S rRNA structure (Nissen et al. 2001).

Kink Turn (K-turn) is a motif that consists of a helix followed by a loop and closed by an additional helix (helix-loop-helix motif) and results in a bend of the RNA helix axis by 120°. Typically, the loop is asymmetric, consists of three nucleotides, and is purine rich. An example for a K-turn is found in the SAM-I riboswitch structure. Here, the canonical Watson-Crick base paired helix is followed by three unpaired nucleotides and closed by two G:A base pairs in the second helix. These G:A base pairs twist the backbone of the helix in a characteristic manner, allowing the adenines to stack on each other and participating in A minor interactions within the junction. Just like the tetraloop-receptor motif, the K-turn can be recognized and stabilized by proteins.

G-Quadruplexes are a very complex and stable tertiary structure element that is found in both RNA and DNA structures. They compose of four strands that are guanine rich and form

stacking guanine tetrads. Those tetrads compose of four guanines that symmetrically interact in a hydrogen bond network by forming a hydrogen bond between the Watson-Crick face of one guanine to the Hoogsteen edge of the neighboring guanine. One monovalent cation (potassium cation, K^+) intercalates between two stacking tetrads, thereby stabilizing the overall tetrad structure, and compensating the repulsive forces between the spatially close oxygens. Those structure elements were prominently found e.g., in telomers.

1.3. PROTONATED NUCLEOBASES INCREASE NUMBER OF POTENTIAL INTERACTIONS AND STABILITY

In respect to the formation of structures, RNA is limited to four chemically very similar building blocks to fulfill all functions. Therefore, RNA uses a wide variety of hydrogen bonding pattern in addition to the canonical Watson-Crick face of the nucleobases for structure and function. In contrast, proteins use 22 chemically and structurally different building blocks to fulfill all structural and functional needs. Protonated residues, such as histidine, play a functional role in proteins and were found to affect stability (Malevanets et al. 2017; Quirk et al. 1998), control conformational changes (Kalani et al. 2013), playing key roles in catalytic mechanisms (Quirk and Raines 1999) and modulate binding affinities to substrates (Ballin et al. 2010). In RNA, adenosine and cytosine can be protonated at the N1 and N3 atoms, respectively (Figure 3). It was shown that in gas phase the N3 protonation is more favorable than the N1 protonation (Touboul et al. 2008). Generally, a N7 protonation is also possible. The two nucleotides are poor bases with low pK_a values in solution ($pK_{a,AMP} = 3.5$, $pK_{a,CMP} = 4.2$) which are normally neutral at pH 7 (Saenger 1984). Nevertheless, recent investigations of RNA crystal and solution structures showed that RNAs also use ionized nucleobases to add to the portfolio of possible hydrogen bond acceptors and donors (Singh et al. 2015).

The functional use of protonated bases can be grouped into two classes: structural feature and functional feature. The former uses the protonation as a building block and increases the number of potential hydrogen bonds build. Importantly, in this case the pK_a value is typically drastically shifted from neutrality and stabilized by its surrounding RNA structure. The latter one, on the other hand, uses the protonation in the functional process. Here, the pK_a is only shifted slightly from physiological pH and protonation or deprotonation is modulated as a reaction to a functional need.

An impressive example for a structural protonation is the artificial neomycin binding riboswitch. Here, a U-to-C mutation in the canonical U-turn motif (5'-UNR-3') results in the

same binding abilities as in the wild type (Gottstein-Schmidtke et al. 2014). In the canonical UNR U-turn motif, the first U forms two long-range interactions. Firstly, the 2'OH group hydrogen bonds with the purine N7 of the +3 nucleotide. Secondly, the U H3N3 and the phosphate group of the +2 nucleotide form an additional hydrogen bond. In a U-to-C-mutant at the crucial position, the latter hydrogen bond would be demolished. It was directly shown by NMR spectroscopy that in the artificial neomycin binding riboswitch, the C substitution mimics the canonical U by being protonated at the N3 atom and is hereby able to undergo the same long-range interactions as the canonical U-turn. Another example for the stabilizing function of protonated nucleotides is the plant luteoviral P1-P2 frameshifting signal found in the beet western yellows virus (BWYV). The 28 nucleotides (nt) P1-P2 pseudoknot induces a -1 programmed ribosomal frameshift at a level of 5-15% (Su et al. 1999). The pseudoknot forms into a compact triple-stranded structure in which the loop nucleotides make many non-canonical hydrogen bonds to the two stem regions. Most interestingly, the protonated C8⁺ from loop L1 forms a Hoogsteen base pair to G12 that is in a Watson-Crick base pair with C26 resulting in a stable base triplet C8⁺:(G12-C26). This base triplet is situated at the top of the helical junction S2 and stabilizes the major groove interaction (Egli et al. 2002). This motif was found to be conserved in all luteoviral RNA pseudoknots (Nixon et al. 2002; Nixon and Giedroc 2000) emphasizing its functional importance.

The functional protonated nucleotides were found to e.g., control conformational changes. One example is the spliceosome which is a complex formed out of over 70 proteins and five small nuclear RNAs (snRNA) (U1, U2, U4 U5, and U6) together with the pre-mRNA. The large and highly dynamic complex catalyzes the splicing of pre-mRNA into mature mRNA. During the splicing catalysis, the U4-U6 complex is disrupted and a U2-U6 complex is formed. It was shown by NMR spectroscopy that the structure of the U6 RNA internal stem loop (ISL) internal loop changes depending on the environmental pH. At lower pH, a wobble C:A⁺ base pair is stabilized, and the adjacent uracil is looped out. In contrast, at higher pH values a C-U base pair is newly formed and the uracil stacks into the stem structure (Venditti et al. 2009). While at physiological pH values both conformations are in an equilibrium, it could be shifted by the assembly of the spliceosome and its RNA, protein, or metal ion cofactors. The conformational change is assumed to be required due to the multiple functions of the RNA when in complex with the spliceosome cofactors.

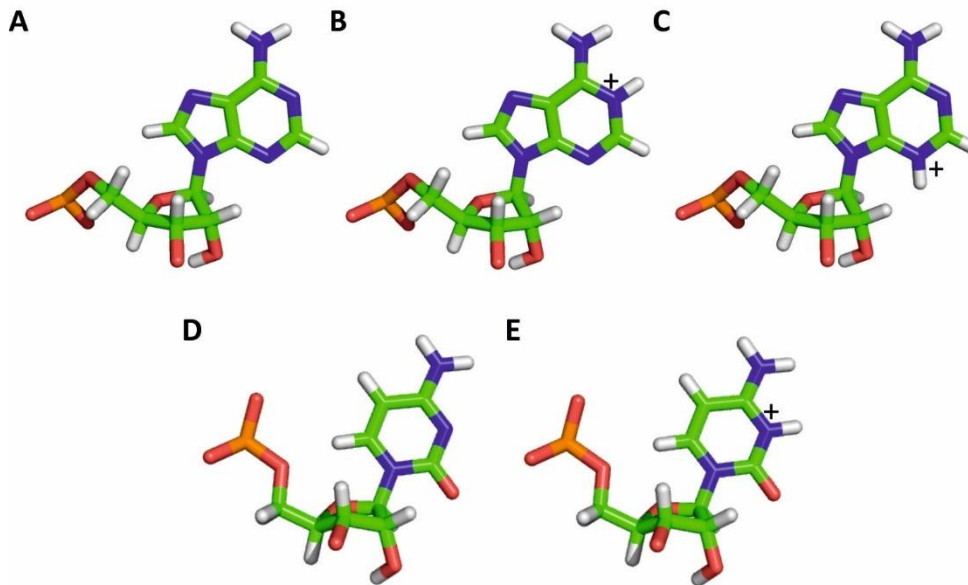


Figure 3: Protonated and neutral states of adenosine and cytosine. Adenosine (neutral, A) can either be protonated at N1 (B) or N3 (C). Cytosine (neutral, D) can be protonated at N3 (E).

Another example, and one of the most intensively studied enzymatic reaction in the last decade, is the self-scission of the HDV ribozyme (Chen et al. 2010). The genomic RNA of HDV is a single-stranded circular RNA that is transcribed in a rolling-circle mechanism using host machinery. The RNA is later cleaved by an intrinsic catalytic sequence, the HDV ribozyme. The ribozyme composes of five helical stem regions (P1, P2, P3, P1.1 and P4) that form a nested double pseudoknot (P1/P2 and P3/P1.1). The catalytic core is in the pseudoknot between P1.1 and P3 and composes of loop L3 and junction J4/2. While loop L3 has some variability, junction J4/2 must contain a “CNRA” motif. The amino group of the conserved C75 of this motif forms two hydrogen bonds: firstly, to non-bridging oxygens of the phosphate backbone between C21 and C22 which are located at the 5’ end of P1.1 and secondly, to the 2’OH group of U20. These interactions lock the junction into the active form and brings the protonated H3N3 of C75 in hydrogen bond distance to the phosphate backbone of G1 to act as a general acid in the cleavage reaction. Here, the pK_a for C75 is shifted to neutrality to perform its designated function (Nakano et al. 2000).

Like proteins, RNAs use a wide range of biochemical interactions and patterns to secure stable structures and efficiently execute their tasks despite having a limited number of available building blocks to perform the needed functions.

2. APTAMERS – THE PROOF THAT RNA HAS HIGH VERSATILITY

In the early 1970, it was found that RNA could form large, static structures and with the discovery of catalytic RNAs, such as the RNase P and the group I self-splicing intron, it was shown that RNA plays a more significant part in the biochemical processes of the cell than previously assumed. With the discovery of the ribosome as a large RNA enzyme, it became clear that RNA could be more than the mediator molecule for protein synthesis. In addition to the limited building block options of choice, all four nucleotides are chemically very similar. In contrast, amino acids are chemically very different making the resulting diversity of proteins structures self-evident.

2.1. SELEX – SELECTING ACTIVE RNA *IN VITRO*

In 1990, it was shown that RNA (and DNA) is capable of direct binding to a variety of compounds (Ellington and Szostak 1990; Robertson and Joyce 1990; Tuerk and Gold 1990). *In vitro* selection experiments (also SELEX - Systematic evolution of ligands by exponential enrichment) use a library of artificial RNA (or DNA) that are exposed to a designated ligand. In an iterative process, RNA sequences with the best binding properties are selected. The found RNA strands are mainly single stranded and unstructured in solution, but when exposed to the designated ligand, they form into highly stable structures incorporating the ligand into the three-dimensional structure of the complex and can discriminate related ligands. The binding affinities of the found RNA are often in the nanomolar range.

The RNA sequences capable of ligand binding are named aptamers (Latin *aptus*, suited and Greek *méros*, part). The SELEX method was further employed to find aptamers for ligands with diverse chemical properties, such as ions, enzyme co-factors, amino acids, nucleotides, oligosaccharides, carbohydrates, and antibiotics. The method is not limited to ions or relatively small molecules. Aptamers which bind to proteins and complex structures like viruses (Zou et al. 2019) or whole cells were isolated, as well (Zamay et al. 2017). RNA structures for aptamers binding to flavin mononucleotide (FMN) (Burgstaller and Famulok 1994; Fan et al. 1996), adenosine triphosphate (ATP) (Dieckmann et al. 1996; Sassanfar and Szostak 1993) and Arginine/Citrulline were solved in the following years (Burgstaller et al. 1995; Yang et al. 1996). Since the aptamers bind to any used ligand with a high enough affinity, a wide range of medical applications is possible. It was shown that the use of modified nucleotides increase the stability of the structure, making modified aptamers a flexible candidate for synthesized therapeutics.

Aptamers are often highly selective and discriminate against chemically related ligand. An example of the very selective discrimination is the L-citrulline binding aptamer. In a SELEX process an aptamer was selected that was highly affine to L-citrulline. The sequence was used as a starting point for the selection for an aptamer specific to L-arginine. L-citrulline is a byproduct of nitric oxide production by the nitric oxide synthase (NOS) that uses L-arginine as a substrate (Burgstaller et al. 1995). The two amino acids differ in the guanidinium group where the imino nitrogen present in L-arginine is exchanged by a carbonyl group in L-citrulline. The two isolated sequences capable of binding either L-citrulline or L-arginine are 33 nucleotides in length. They compose of three stems that are interrupted by two internal loops and capped by a GNRA tetraloop. The discriminating ability for the two closely related compounds is mediated by three nucleotides in the binding pocket (residues 13, 29, 31) that differ in the two sequences. The structures of the two aptamers were determined in a combined NMR-MD approach (Yang et al. 1996). They adopt a similar overall structure but significantly differ in the ligand recognition site explaining the shifted selectivity (Figure 4A). Both ligands are hydrogen bonded by their side chains, while the zwitterion points towards the minor groove. In both complexes, the second internal loop is closed by a noncanonical G:G base pair on which the aliphatic tail of the ligand side chain stacks, as well as the pyrimidine at position 13 and residue G30 form hydrogen bonds to the amino groups of both ligands. In the arginine aptamer, C13 forms a hydrogen bond between its N3 as acceptor and the imino proton of arginine as donor (Figure 4C). The ligand binding hydrogen bond in the citrulline aptamer is formed between the imino group of U13 and to the citrulline carboxyl oxygen (Figure 4D). In both complexes, a water-mediated hydrogen bond is possible between residue 13 O2 and N_ε of the two ligands. The positioning of pyrimidine 13 is supported by purine 29 in both binding pockets. A29 of the arginine aptamer hydrogen bonds the C13 amino group via its N3. Similarly, G29 of the citrulline aptamer forms a hydrogen bond between its imino proton and O4 of U13. An additional hydrogen bond between the G29 amino proton and the carbonyl oxygen of citrulline is proposed. Finally, G31 of the arginine aptamer hydrogen bonds both arginine amino groups via its N7 and O6, while U31 of the citrulline aptamer hydrogens bonds the amino group of citrulline via its O4. Since both ligands were connected to the column material via its peptide amine moiety it is not surprising that the zwitterion is solvent accessible in both aptamer complexes.

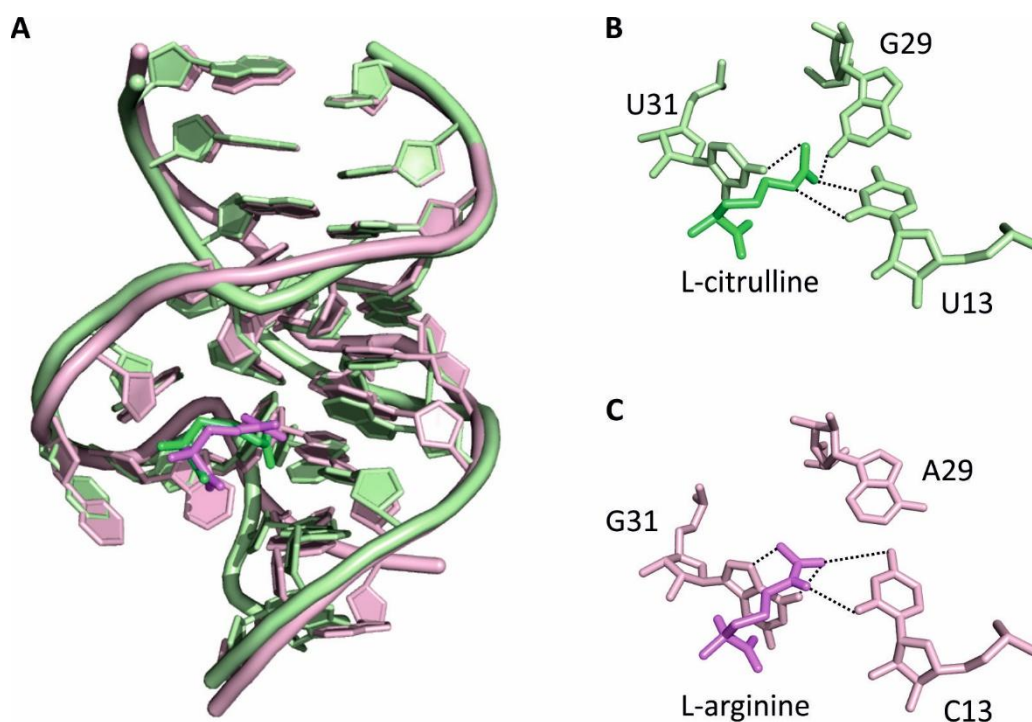


Figure 4: L-Citrulline binding aptamer in comparison with L-Arginine binding aptamer. Overlay of both structure, L-citrulline binding aptamer in green (PDB 1KOD) and L-arginine binding aptamer in pink (PDB 1KOC) (A). Ligand recognition of L-citrulline binding aptamer (B) to residues that differ in L-arginine binding aptamer (C). Hydrogen bonds are shown as dashed lines.

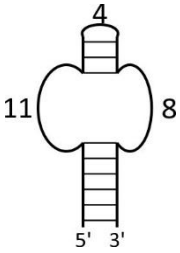
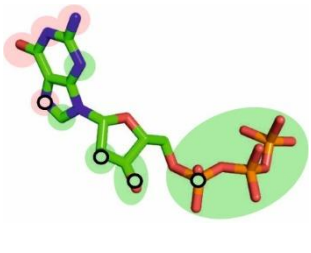
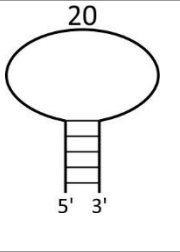
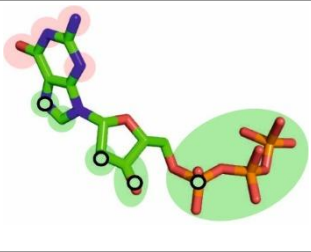
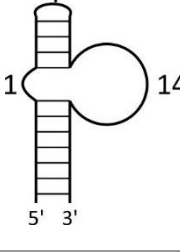
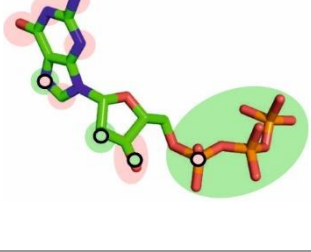
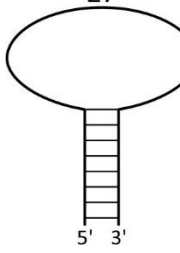

2.2. GTP BINDING APTAMERS – MANY DIFFERENT WAYS TO RECOGNIZE THE SAME LIGAND

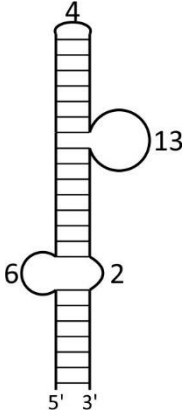
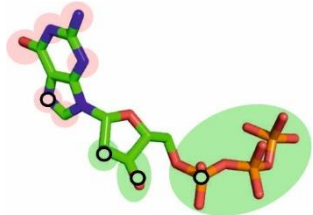
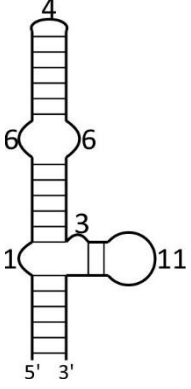
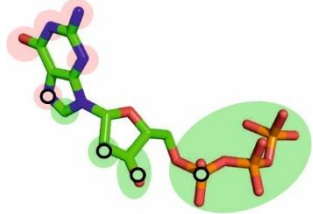
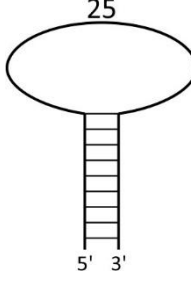
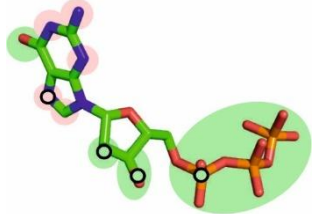
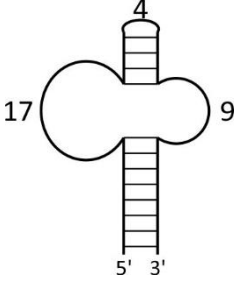
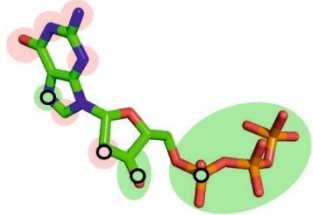
Carothers and coworkers showed the most impressive picture of RNA diversity in recognizing the same ligand, as they purified a collection of guanosine triphosphate (GTP) binding aptamers using SELEX and characterized them (Carothers et al. 2004; Carothers et al. 2006; Davis and Szostak 2002). In their work, they found eleven individual GTP binding aptamer classes with K_d values ranging from 8 μM to 9 nM. 16 chemically different analogues of GTP were used to analyze the binding affinities and specificities. Most interestingly, all RNA aptamers can be grouped into classes by their predicted secondary structures and their resulting ligand binding pockets (Table 1). Varieties of GTP binding strategies were characterized but most likely, all classes differ from each other in the shape and composition of their binding pocket.

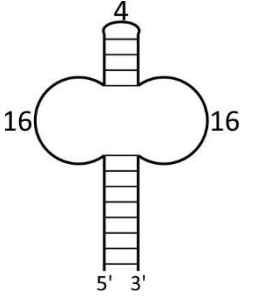
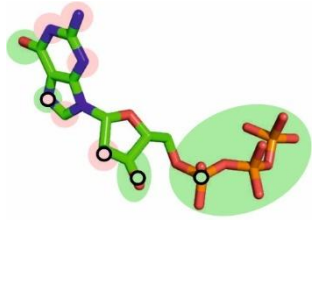
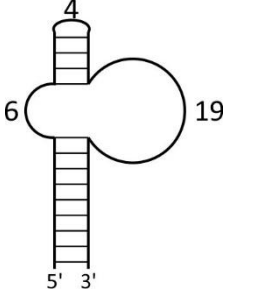
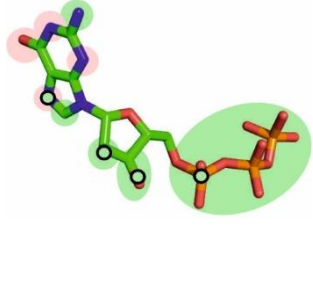
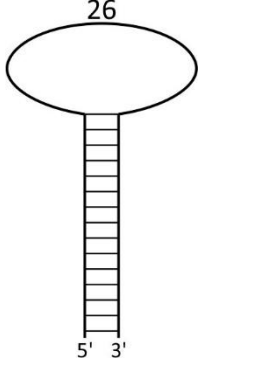
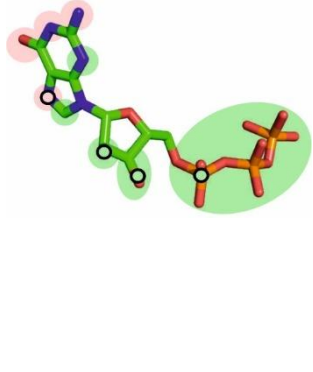
It was long assumed that the selection for higher-affinity binding is always accompanied by binding that is more specific to the target ligand (Eaton et al. 1995). This theory does not consider the contribution of free energy of binding in combination with the intramolecular contacts being formed and broken upon ligand binding and the intermolecular contacts being formed between RNA and ligand. Carothers and coworkers showed that there is no evidence that high affinity binding goes hand in hand with specificity. The hypothesis was proposed that

high affinity binding is mainly influenced by the stability of RNA tertiary structure, namely additional intramolecular RNA-RNA interactions.

Table 1: GTP binding aptamers. Eleven classes of GTP binding aptamers were discovered and characterized. Table indicated class name, their binding affinities, secondary structure (numbers indicate nucleotides per loop) and the ligand's functional group that are recognized. Colored areas indicate positions where bulky residues did (red) or did not (green) alter binding affinities by more than 100-fold. Small circles show positions where the presence of functional group is necessary for binding. Ligand orientation is not necessarily as shown and only GMP variants were investigated (Carothers et al. 2006).

Aptamer	Binding affinity	Secondary Structure	Ligand recognition
Class I	76 nM		
Class II	400 nM		
Class III	8000 nM		
Class IV	900 nM		

Aptamer	Binding affinity	Secondary Structure	Ligand recognition
Class V	17 nM		
9-4	9 nM		
9-12	300 nM		
10-6	300 nM		

Aptamer	Binding affinity	Secondary Structure	Ligand recognition
10-10	30 nM		
10-24	300 nM		
10-59	250 nM		

To support this hypothesis, Carothers and coworker measured the K_d using competition-binding assays with radiolabeled ^{32}P -GTP and unlabeled analogues. Therefore, they not only measured the K_d of the aptamers to the ligand but also the specificity by comparing the ratio of binding to GTP versus binding to analogues. It was shown that the ligand-binding event and RNA folding are tightly coupled in the characterized aptamers. Furthermore, binding affinities of the aptamers to GTP were mainly influenced by the quality of the tertiary intramolecular RNA-RNA interactions rather than by the number and quality of intermolecular RNA-GTP interactions. Even more interestingly, there was no correlation between the specificity for GTP and the binding affinity found in the characterized aptamer classes. Indeed, they found that not the specificity but the calculated free energy change (ΔG) of secondary structure formation positively correlates with the binding affinity for four out of eleven aptamer classes. Additionally, comparison of sequence conservations patterns showed that an increase in

affinity is accompanied by a conservation of bases or base pairs. Nevertheless, further investigations are needed for closer insights into the interplay between ligand binding affinities and structure stability or conservation.

Two binding modes have been further investigated, so far. As described by Carothers and coworkers, the ligand binding strategies could not be more different. The GTP Class I aptamer binds the ligand via a trans WC-Hoogsteen base pair between G31 and the nucleobase moiety of GTP. The resulting loss of binding to 6-thio-GTP and 7-deaza-GTP is evident, since the bases' Hoogsteen faces are modified. Interestingly, the binding affinity to 7-methyl GTP is not lost but diminished by over 100-fold. In contrast, 3-methyl GTP binds to the aptamer with WT affinity since the sugar edge is not involved in specific intermolecular contacts in the GTP-RNA complex. The GTP nucleobase also stacks between G9 and G15. G9 is in a base triplet to G16 and G29, while G15 forms a *syn* WC-sugar edge with a G15-A32 base pair. Furthermore, the contact to G16 is stabilized by a hydrogen bond to the GTP imino proton, even though both bases are not co-planar. 1-methyl GTP binds to the aptamer with over 100-fold decreased affinity due to the steric clash caused by the additional methyl group. The GTP ribose nestles below the first turn of L1. The triphosphate points towards the solvent, since GTP was bound to the column matrix via its phosphate group during the SELEX process.

The GTP Class V aptamer was investigated concerning the ligand binding mode (Nasiri et al. 2016), but no full structure is available, to date. The secondary structure of GTP Class V aptamer composes of three stem regions that are interrupted by bulges and capped by a UUCG tetraloop. The ligand-binding mode was investigated using NMR spectroscopy. The chemical shift of the GTP imino proton indicates that the ligand is bound in a non-Watson-Crick base pair. Furthermore, there were no further hydrogen bond between the imino group and a nitrogen acceptor detectable, indicating an oxygen acceptor. It was also shown that ligand binding depends on the presence of magnesium ions and most interestingly, the presence of potassium ions was crucial for binding of GTP to the aptamer. Sodium ions could moderately compensate the effect of potassium ions, while all ions with similar radii to potassium were sufficient for binding. The sensitivity for different ion radii indicated the formation of a G-quadruplex. By investigation the typical NOE pattern between the RNA and GTP, it was shown that the aptamer incorporates the GTP ligand into an intermolecular G-quadruplex structure upon complex formation. Additionally, seven guanines (at position 12, 13, 42, 43, 46, 47 and most likely 40) of the two bulge regions were shown to be involved in

quadruplet formation. The intermolecular G-quadruplex also explains, why 6-thio-GTP does not bind to the aptamer, since the crucial tetrad hydrogen bond is significantly weakened. As shown previously, the class V aptamer only recognizes the nucleobase moiety tightly, while the RNA most likely does not specifically bind to the sugar and triphosphate moiety.

The ensemble of GTP binding aptamers shows how RNA can bind the same ligand in very different ways. This becomes even more impressive considering the limited structural and chemical diversity of the building blocks at hand. Nevertheless, aptamers are capable of highly selective and specific binding to a variety of ligands.

3. RIBOSWITCHES – FROM *IN VITRO* TO *IN VIVO* (OR THEORY FOUND IN NATURE)

Since the discovery of RNA aptamers, the picture of RNA as a sole mediator molecule in protein synthesis was dismissed. In 2001, the THI box was described (Nou and Kadner 1998). This regulatory element found in the 5'-UTR of specific mRNAs interacts with thiamine, riboflavin, and cobalamin to regulate gene expression of vitamin B₁-, B₂-, and B₁₂-biosynthetic genes. Since no classical modulator (e.g., proteins) could be found, it was proposed that the mRNA might directly bind the effector molecules as do synthetic aptamers. *In vivo* experiments indicated that the mRNA leader sequence adopts different conformations depending on adenosyl cobalamin presence or absence. Nevertheless, no direct binding of cobalamin to the RNA could be proven (Ravnum and Andersson 2001). In 2002, it was shown that three vitamin derivatives indeed directly interact with the mRNA of interest: thiamine pyrophosphate (TPP) (Mironov et al. 2002; Winkler et al. 2002), Flavin mononucleotide (FMN) (Mironov et al. 2002) and adenosyl cobalamin (AdoCbl) (Nahvi et al. 2002). The researchers indicated that the binding of the vitamin derivative stabilizes a conformation of a part of the mRNA sequence which is called aptamer domain. Ligand binding thereby influences the folding of a downstream mRNA region whose structure interferes with transcription or translation, named expression platform. The interplay of the aptamer domain and the expression platform switch gene expression in response to the presence or absence of a ligand is called riboswitch.

Since then, riboswitches have been described in all three domains of life, but the vast majority is found in eubacteria. Here, riboswitches are mostly found in the 5'-untranslated region (UTR) of mRNAs of genes that are mostly associated with the ligand bound to the aptamer domain. Later, even more complex riboswitch constellations have been described. Thus, the glmS

riboswitch regulates gene expression by promoting mRNA self-cleavage upon ligand binding. Additionally, the glycine riboswitch is often located in an arrangement with two aptamer domains and one expression platform (Mandal et al. 2004). The two aptamer domains were shown to cooperatively bind the two glycine molecules to ensure rapid and effective gene regulation. Additionally, tandem riboswitches were described to bind two different ligands. The riboswitch located upstream of the metE gene has two aptamer domains: one binding S-adenosylmethionine (SAM) and the other binding vitamin B₁₂. The riboswitch acts in a two-input Boolean NOR gate logic: The presence of either ligand decreases the full-length transcript concentrations, but the presence of both decrease it even further (Sudarsan et al. 2006).

In the last years, the classes of riboswitches were expanded by additional members, but also new riboswitches for new ligands were discovered. Up to date, 45 riboswitches classes for 31 ligands have been described. The chemical identity of the ligands bound is highly diverse e.g., ions, coenzymes, amino acids, and nucleotides are bound by riboswitches. Even more complex interaction patterns which have not been discovered yet may be possible.

Table 2 Summary of all riboswitch classes known to date. The ligand group contains all ligands with similar functions or structure. The known riboswitch classes for a certain ligand are listed. PDB ID indicates the structures solved for an indicated riboswitch. Also, literature references are listed.

Ligand group	Ligand	Riboswitch (classes)	PDB ID	Ref
Coenzymes and derivatives	Adenosylcobalamin (vitamin B12)	AdoCbl	4GMA, 4GXY	(Johnson et al. 2012; Peselis and Serganov 2012)
	Thiamine pyrophosphate (vitamin B1)	TPP	2GDI	(Serganov et al. 2006)
	Flavin monophosphate (vitamin B2)	FMN	3F4E	(Serganov et al. 2009)
	S-Adenosylmethionine	SAM (SAM-I, SAM-II, SAM-III, SAM-IV, SAM-V,	4KQY, 2QWY, 3E5C,	(Gilbert et al. 2008; Huang and Lilley 2018; Lu et

Ligand group	Ligand	Riboswitch (classes)	PDB ID	Ref
		SAM-VI, SAM-I/IV)	6UET, 6FZO, 6LAS, 4OQU	al. 2008; Lu et al. 2010b, 2010b; Sun et al. 2019; Trausch et al. 2014; Zhang et al. 2019)
	Aquacobalamin	AqCbl	4FRN	(Johnson et al. 2012)
	Molybdenum cofactor	Moco		(Regulski et al. 2008)
	S-Adenosylhomocysteine	SAH	3NPQ	(Edwards et al. 2010)
	Tungsten cofactor	Wco		(Regulski et al. 2008)
	S-Adenosylmethionine/ S-Adenosylhomocysteine	SAM-SAH	6HAG and 6YL5, 6YLB	(Huang et al. 2020)
	Tetrahydrofolate	THF	3SUX	(Huang et al. 2011)
	4-amino-5-hydroxymethyl-2-methylpyrimidine diphosphate	HMP-PP		(Atilho et al. 2019)
	Nicotinamide adenine di nucleotide	NAD ⁺	7D81	(Chen et al. 2020)

Ligand group	Ligand	Riboswitch (classes)	PDB ID	Ref
Nucleotides and derivatives	Adenine, Guanine	Purine (Adenine, Guanine)	1Y26, 1Y27	(Serganov et al. 2004)
	2'Deoxyguanosine	2'-dG (2'-dG-I and 2'-dG-II)	3SKI, 6P2H	(Matyjasik and Batey 2019; Pikovskaya et al. 2011)
	Prequeuosine-1	preQ ₁ (preQ ₁ -I, preQ ₁ -II, preQ ₁ -III)	2L1V, 4JF2, 4RZD	(Kang et al. 2009; Liberman et al. 2013; Liberman et al. 2015)
	Phosphoribosyl pyrophosphate	PRPP	6CK5, 6DLT	(Knappenberger et al. 2018; Peselis and Serganov 2018)
	Nucleoside diphosphate	(d)ADP/(d)CDP		(Sherlock et al. 2019)
Amino acids	L-Lysine	Lysine	3DIL/3 DU	(Garst et al. 2008; Serganov et al. 2008)
	Glycine	Glycine	3OWI	(Huang et al. 2010)
	L-Glutamine	Glutamine	5DDP	(Ren et al. 2015b)
Ions	Magnesium	Mg ²⁺	2QBZ	(Dann et al. 2007)
	Fluoride	Fluoride	4ENC	(Ren et al. 2012)
	Manganese	Mn ²⁺	4Y1I	(Price et al. 2015)
	Nickel or Cobalt	NiCo	4RUM	(Furukawa et al. 2015)

Ligand group	Ligand	Riboswitch (classes)	PDB ID	Ref
Signaling molecules	3'-5'-cyclic-di-guanosine monophosphate	c-di-GMP (c-di-GMP-I and c-di-GMP-II)	3IWN/ 3IRW,3 Q3Z	(Kulshina et al. 2009; Smith et al. 2009; Smith et al. 2011)
	3'-5'-cyclic-di-adenosine monophosphate	c-di-AMP	4QLN/ 4QK9/ 4W90	(Gao and Serganov 2014; Jones and Ferré-D'Amaré 2014; Ren and Patel 2014)
	3'-5'-cyclic-di-adenosine monophosphate guanosine monophosphate	c-GAMP	4YAZ	(Ren et al. 2015c)
	5-amino 4-imidazole carboxamide riboside 5'-triphosphate	ZTP	4ZNP	(Ren et al. 2015a)
	Guanosine-3',5'-bispyrophosphate	ppGpp	6DMC	(Peselis and Serganov 2018)
Other	Glucosamine-6-phosphate	GlmS	2HO7	(Klein and Ferré-D'Amaré 2006)
	Azaaromatics	Azaaromatic		(Li et al. 2016)
	Guanidine	Guanidine (Guanidine-I, Guanidine-II, Guanidine-III)	5T83, 5VJ9, 5NZ6	(Huang et al. 2017; Reiss et al. 2017; Reiss and Strobel 2017)

3.1. CHALLENGES IN RIBOSWITCH RESEARCH

Since RNA was not only discovered later than proteins but also perceived as a mediator molecule with no further function or structure of interest, RNA research runs behind protein research. Furthermore, RNA molecules are unstable due to their catalytical activities and more sensitive to experimental setups. Also, method development started later and is only partly transferable from proteins to RNA.

Missing pieces in the impressive puzzle of cell functions hamper RNA research. As one example, riboswitches (see also Chapter 3) interact with a wide range of ligands to regulate gene expression. To date, the first steps of riboswitch investigation heavily rely on the knowledge of the genes regulated by the riboswitch sequence. The ligands bound by the riboswitch are often associated with the function of the protein encoded on the regulated gene. In many cases, this assumption directly leads to the identification of the native ligand and only then further investigations of the riboswitch can follow. In some cases, the function of the protein expressed by the regulated gene is unclear or the regulation pathway is complex or unknown. Therefore, finding the native ligand is difficult and often misleading.

One example for the challenging investigation of a riboswitch is the *ydaO* sequence. It was first identified as a putative riboswitch in 2004, and until now it is among the most common riboswitches found in diverse range of organisms. The genes associated with this riboswitch sequence are highly diverse. In 2012 Fedor and coworkers proposed that the *ydaO* riboswitch from *Bacillus subtilis* senses adenosine triphosphate (ATP) (Watson and Fedor 2012). They showed that the binding affinity between the purified riboswitch and ATP was in the nanomolar range in *in vitro* experiments. Indeed, the binding was very tight and in the same range found for other riboswitch classes. Additionally, Fedor and coworkers concluded that many genes that are controlled by this riboswitch sequence could potentially be involved in ATP homeostasis. A year later, Breaker and colleagues reinvestigated the same sequence using purification assays with yeast extract (Nelson et al. 2013). They found that the cyclic dinucleotide (CDN) c-di-AMP is recognized with picomolar affinity. The cyclic second messenger was only discovered in 2008 (Pesavento and Hengge 2009; Römling 2008). Not long after, three groups solved the structure of the riboswitch roughly at the same time (Gao and Serganov 2014; Jones and Ferré-D'Amaré 2014; Ren and Patel 2014). Except for small differences in sequences due to structure solving approaches, all three structures are almost identical. They showed that the *ydaO* sequence in fact encodes a c-di-AMP binding riboswitch.

The combination of factors or the putative combination of all make RNA research challenging: Misleading experimental data and blind spots for the ligands and proteins found in a cell, though not uncommon in scientific research, make investigations difficult as shown for the c-di-AMP binding riboswitch. But even more importantly, riboswitch research heavily relies on the knowledge of the genes regulated and the protein function encoded in the genes. Structure prediction could elucidate the ligand binding pocket and give an additional hint how a potential ligand could look like. However, reliable structure predictions always rely on a profound database of solved structures. For many reasons, some are mentioned above, the RNA structure database is smaller than that for proteins. First more structures must be solved to give a solid foundation to structure prediction.

3.2. SAM AND SAH BINDING RIBOSWITCHES

SAM riboswitches belong to the first discovered riboswitch classes and by now seven different subclasses are characterized (SAM-I to SAM-VI, SAM-I/IV) which can be grouped in three super-families. All these riboswitches selectively bind SAM and strongly discriminate against chemically related compounds. Only one riboswitch is known which binds the SAM degradation product SAH. Interestingly, a riboswitch exclusively found in Rhodobacterales binds SAM and SAH with similar affinities. It was therefore named the SAM/SAH-riboswitch (Weinberg et al. 2010).

3.2.1. THE VERSATILE SAM AND THE TOXIC SAH

SAM is a sulfur-containing compound that is essential for many biochemical processes in all living cells. Its essential role is displayed in its abundant use as the second most frequently used enzyme substrate after ATP (Cantoni 1975). SAM is stereo-specifically synthesized out of methionine and ATP by SAM synthetase (Tabor and Tabor 1984) or methionine adenosyl transferase (Figure 6).

SAM is a source of a variety of chemical groups that can be donated (Figure 5 A). In eukaryotes and archaea, the biosynthesis of post-transcriptionally modified tRNA^{Phe} nucleoside wybutosine (yW) shows an impressive picture of its versatility. Here, SAM is used in five enzymatic processes and donates three different groups to the substrate tRNA (Figure 5 C). First, the tRNA-methyltransferase TRM5 transfers the methyl group from SAM to N1 of G37-tRNA (m1G37-tRNA). Second, the radical-SAM enzyme TYW1 produces an Ado radical that helps to incorporate a 2-carbon unit of the second co-substrate pyruvate (Figure 5 B). 4-

demethylwybutosine (imG-14-tRNA) with its characteristic imidazoline ring is produced. This reaction is chemically very interesting and not only involves the two co-substrates but also two oxygen-sensitive $[4Fe-4S]^{2+}/^+$ clusters. Third, the SAM 3-amino-3-carboxypropyl (acp) group is transferred to the tRNA substrate by the tRNA-aminocarboxypropyl-transferase TYW2 and 7-aminocarboxypropyl-demethylwybutosine (yW-72-tRNA) is synthesized. Fourth, the tRNA-methyltransferase TYW3 using SAM as a cofactor methylates the nucleoside at N4. Finally, tRNA-methyl-methoxycarbonyl-transferase TYW4 catalytically modifies the α -amino- α -carboxypropyl side chain in two reactions at the same catalytic site: First, it brings the nucleophilic α -carboxy group in a favorable spatial orientation towards SAM to promote a nucleophilic attack. Second, the α -amino group is deprotonated by Tyr229 and Arg88 activates CO_2 that subsequently is methylated by SAM and added to the α -amino group (Noma et al. 2006).

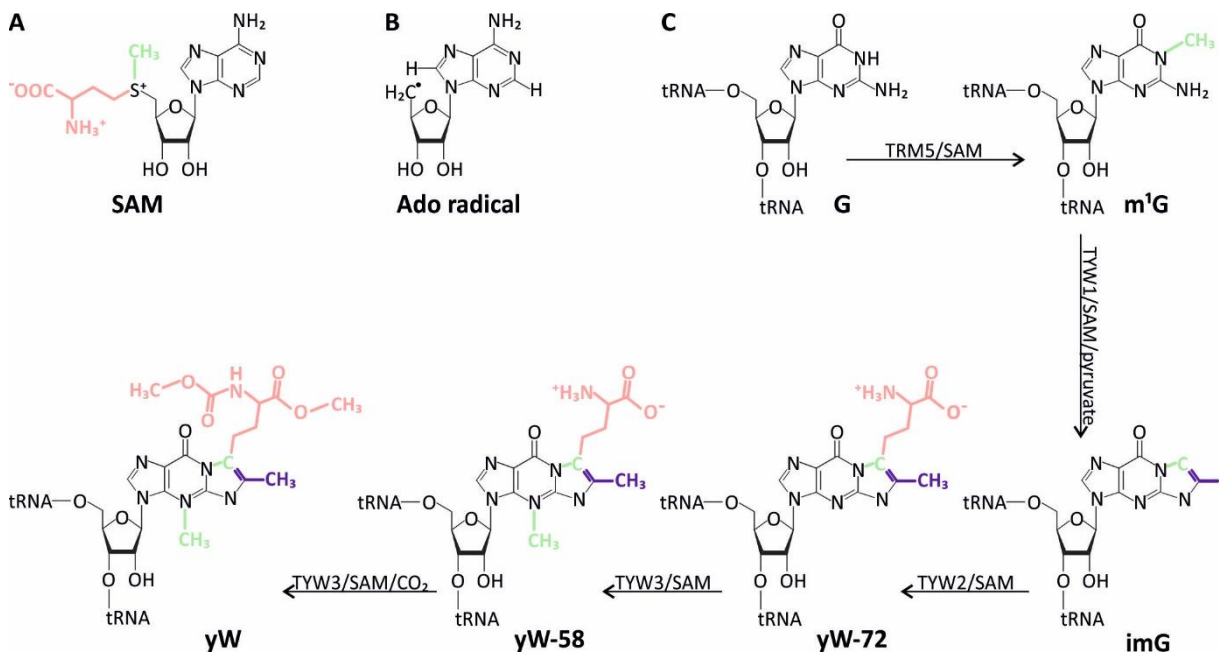


Figure 5: SAM functional groups involved in wybutosine biosynthesis. (A) SAM as donor of its methyl group (green) and acp group (pink). (b) The Ado radical is used by an Fe-S cluster together with pyruvate as cofactors. (C) The SAM functional groups transferred in the wybutosine biosynthesis are collocated accordingly.

In addition to the three mentioned functional groups that are transferred from SAM in the wybutosine biosynthesis, SAM is known to also be a source of amino (Stoner and Eisenberg 1975a, 1975b) and ribosyl groups (Iwata-Reuyl 2003). Nevertheless, the best investigated function of SAM is its role as methyl group donor. The mechanism is found in all kingdoms of

life. Hereby, diverse classes of methyltransferases (MTase) (Cantoni 1975; Chiang et al. 1996) use SAM as a cofactor to methylate various substrates; from small metabolites up to macromolecules like DNA (Schäfer et al. 1997) or RNA strands (Yan and Fujimori 2011), as shown in the wybutosine biosynthesis. During the reaction S-adenosylhomocysteine (SAH) is produced, which was shown to bind to methyltransferases with high affinity and inhibit the essential methyl transfer ability (Figure 6)(Ueland 1982). Due to the cell toxic potential, SAH is rapidly degraded. Most abundant, SAH hydrolases break down SAH into adenosine and homocysteine (Kusakabe et al. 2015; Turner et al. 2000). The latter is either converted into glutathione acting as a cellular antioxidant or remethylated to methionine. The second reaction is catalyzed by the methionine synthase. SAH can also be degraded by SAH nucleosidase which cleaves the glycosidic bond to yield adenine and S-ribosylhomocysteine (Della Ragione et al. 1985; Parveen and Cornell 2011). S-ribosylhomocysteine is then hydrolyzed to homocysteine and 4,5-dihydroxyphenatne-2,3-dione (DPD) by an S-ribosylhomocysteine lyase (Schauder et al. 2001; Winzer et al. 2002). The latter can be remethylated by a cobalamin-dependent methionine synthase. In this reaction using 5-methyl-tetrahydrofolic acid is used as methyl group donor and L-methionine is formed which subsequently can be used with ATP to form SAM (Old et al. 1988).

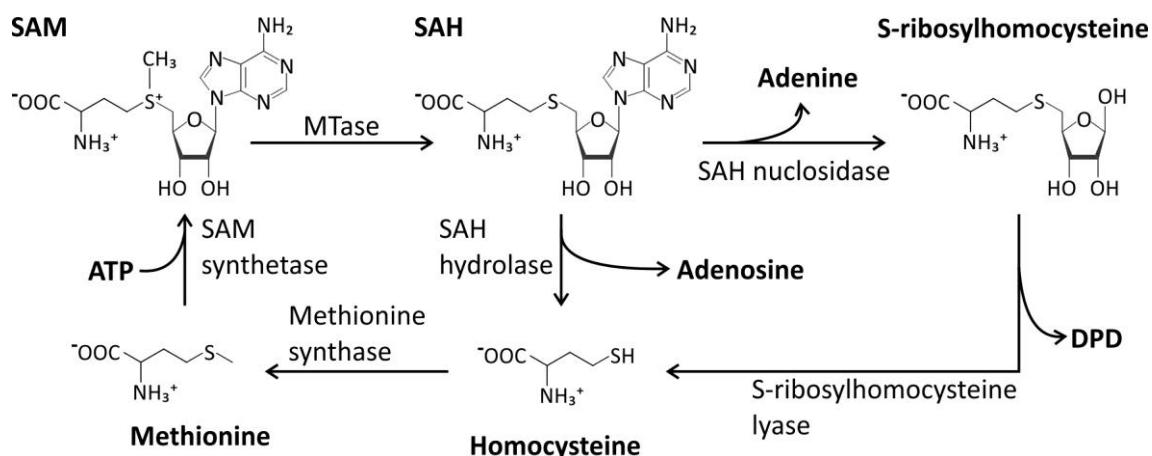


Figure 6: SAM biosynthesis and SAH degradation.

3.2.2. COMPARISON OF KNOWN RIBOSWITCH STRUCTURES AND LIGAND RECOGNITION

The versatile biochemical use of SAM and the cell toxicity of the metabolic byproduct SAH implies the need for a tight regulation of the biosynthesis of SAM and degradation of SAH. In *E. coli*, the repressor protein MetJ regulates the SAM synthetase gene *metK*. Upon SAM accumulation, SAM binding activates MetJ and the complex subsequently binds to the

promotor region of the *met* genes (Holloway et al. 1970; Shoeman et al. 1985; Smith et al. 1985). In *Corynebacterium glutamicum* the regulator protein McbR was found to regulate several genes involved in the sulphur metabolism (Rey et al. 2005). McbR is a TetR family regulator which is transcriptionally active and regulates by binding to the designated DNA in absence of the effector (Rey et al. 2003). Upon effector binding, the protein dissociates from the DNA. McbR also regulates the SAM synthetase gene *metK* (Rey et al. 2003) and was shown to bind to SAH to control *metK* gene expression in a negative feedback mode. (Rey et al. 2005). SAH is a potent inhibitor of methyltransferases as mentioned before. Additionally, SAH levels directly correspond to the consumption levels of SAM: at high transmethylation reaction rates, SAH levels increase. Subsequently, SAH binds to the effector McbR that dissociates from the regulon DNA and *de novo* synthesis of SAM is ensured. Like MetJ, the regulator SahR acts on the gene expression of the SAH hydrolase AhcY in most α - and δ -Proteobacteria, and some γ -Proteobacteria. The SahR protein belongs to the ArsR-family transcription factor and is a negative regulator that binds its DNA operators in the absence of effector, whereas the addition of SAH disrupts the SahR-DNA complex and induces the target gene expression (Novichkov et al. 2014).

In addition to the classic protein repressor model, riboswitches directly sense the SAM and SAH concentration and can act on a transcriptional and translational level. SAM- and SAH-binding riboswitches are known from the beginning of riboswitch investigation and six distinct classes of SAM-binding riboswitches were found and several structures were solved to date. Additionally, SAM is the ligand with the highest number of riboswitches known so far. Just as described for the GTP binding aptamers, nature found diverse ways to specifically bind SAM or SAH and to discriminate against related analogues. A discrimination of 100-1000-fold of SAM over SAH is found in SAM binding riboswitches. Using the structures and comparing secondary structures and the ligand recognition modes to one another lead to the organization of super-families known to date. SAM-binding riboswitches occur in bacteria in different frequencies. While the SAM-I riboswitch is the most abundant riboswitch, SAM-VI is not only the most recently discovered riboswitch, but also the rarest. Nevertheless, all three have in common that the ligand's adenine group is stacking intensively, the sulfonium is strongly recognized by carbonyl contacts, and the methyl group not directly contacted. The last most likely to avoid self-methylation through SAM binding. Finally, the methionine tail is recognized differently in all classes.

SAM-I, SAM-IV and SAM-I/IV Riboswitches

SAM-I is the first identified SAM binding riboswitch and was formerly annotated as the S-leader box which is a genetic control element of a regulon composed of 11 transcriptional units (Winkler et al. 2003). The regulated mRNA encodes 26 different genes in *Bacillus subtilis* that are all involved in sulfur metabolism, methionine, cysteine, and SAM biosynthesis (Winkler et al. 2003). The riboswitch class is found in Gram-positive bacteria and controls gene expression by transcription termination. Hereby, the expression platform forms an anti-terminator upon ligand binding (Winkler et al. 2003). The SAM-I riboswitch is one of the best-investigated riboswitches concerning structure, folding and dynamics. The riboswitch consensus sequence comprises 94 nt and it is the most widely distributed riboswitch (Barrick and Breaker 2007).

The global architecture composes of two sets of coaxially stacked helices, P1/P4 and P2/P3, which are connected by junction J4/1. Three conserved adenines (A83-A84-A85) within the junction connect the helices by stacking on each other. Additionally, A85 base pairs with U64 from junction J3/4 and A24 in loop L2 and thereby anchoring the P1/P4 stack in an 70° angle against the other. A kink-turn motif in P2 bends the helix by 100°. The bend brings stem P2b and L2 in proximity of the P1/P4 stack to form a pseudoknot between L2, J3/4 and J4/1. In-line probing experiments showed that the pseudoknot interaction is not ligand dependent. The helical stacks and the pseudoknot facilitate ligand recognition, while ligand binding stabilizes the overall structure to ensure anti-terminator formation (Hennelly and Sanbonmatsu 2011).

The ligand binds to the minor groove of the P1/P3 helices. In the SAM-RNA complex, SAM is bound in a compact conformation: the acp moiety stacks on the adenine ring. This conformation is stabilized by π -interaction between the aromatic ring and the amino group of the ligand itself. Most proteins are found to bind SAM in an extended *trans* conformation, as other SAM binding riboswitches (Layer et al. 2003). The adenosyl group of SAM is within a base triplet involving A45 and U57 which both are part of an asymmetric internal loop motif in helix P3. Additionally, the adenosyl group stacks on C47 from the base triplet A46:C47-G56. The adenosyl group is also hydrogen bonded via its O4 to the sugar 2'OH of SAM. The acp moiety is hydrogen bonded to G11:C44-G58 base triplet from P3 and J1/2. Furthermore, the carboxylate group is recognized in a pseudo-Watson-Crick base pair with G11. Discrimination between SAM and SAH is accomplished by positioning the carbonyl groups of U7 and U88

within 4 Å of the sulfonium ion to form electrostatic interactions. Notably, the methyl group is not directly bound and points towards the solvent, most likely to avoid methylation of the riboswitch RNA.

While the ligand binding core of SAM-IV and SAM I/IV are very similar to SAM-I they differ in the peripheral architecture (Trausch et al. 2014; Zhang et al. 2019). In SAM-IV and SAM-I/IV form an additional pseudoknot structure with a stem loop of P3 and parts of the additional P5 stem structure. Also, P4 is significantly different. While SAM-I and SAM-IV only share the ligand binding site architecture, SAM-I/IV shows the same ligand binding core as the other two and the same peripheral structure as SAM-IV.

SAM-II and SAM-V Riboswitches

The second SAM-riboswitch super-family composes of riboswitches from the SAM-II (Corbino et al. 2005) and SAM-V classes (Poiata et al. 2009). These classes are evolutionary independent from SAM-I and drastically differ in structure from it, even though they use similar strategies for ligand binding and discrimination against related ligands (Gilbert et al. 2008; Huang and Lilley 2018). SAM-II riboswitches are found in proteobacteria, especially α -proteobacteria. They are located upstream of *metA* and *metC* genes or other methionine and SAM biosynthesis genes. Most SAM-II riboswitches are transcriptionally active, while the riboswitch from which the SAM-II structure was solved regulates translation of the homoserine acetyltransferase *metX*. SAM-II forms an H-type pseudoknot that composes of two Watson-Crick helices (P1 and P2a) and two connecting loops (L1 and L3). Formerly annotated loops L2 and L3 form an additional helix P2b. Each helix stacks upon the next to form a nearly straight structure. Additionally, L1 and L3 interact with the major and minor grooves of P2a/b and P1 to form a compact tertiary structure. The SAM binding pocket is in the center of P2b and the ligand is bound in an extended conformation with its ribose and aminocarboxypropyl moieties stretched along the major groove of the P2b/L2 triplex. The Hoogsteen edge of the SAM adenosyl group base pairs with U44 and in combination with U10 forms a base triplet. The SAM Watson-Crick edge is solvent exposed. Like in SAM-I, the ligand's sulfonium ion is coordinated by two carbonyl groups (U11 and U21) in the major groove of SAM-II. In contrast to SAM-I, both are in an almost 180° angle to the S-C bond, explaining the strong discrimination for SAM over SAH (1000-fold in SAM-II, 100-fold in SAM-I). Just as in SAM-I, all functional groups of SAM are specifically recognized (Lim et al. 2006). While in SAM-I the

pseudoknot formation is ligand-independent, chemical probing on SAM-II showed ligand-dependent stabilization of P2b.

Like SAM-II, SAM-V is found in marine α -proteobacteria and Bacteroidetes (Meyer et al. 2009). The riboswitch is found in the 5'UTR of *metY*, *mmuM*, and *bhmT* mRNAs, which encode O-acetylhomoserine (thiol)-lyase, homocysteine S-methyltransferase, and betaine-homocysteine methyltransferase (Poiata et al. 2009). SAM-V riboswitches were found to translationally regulate gene expression by sequestering the Shine-Dalgarno (SD) sequence in the tertiary structure upon ligand binding. A putative riboswitch from *Candidatus Pelagibacter ubique* (Meyer et al. 2009) was described and in-line probing experiments showed that consensus sequence and secondary structure for SAM-V share several key characteristics with SAM-II. SAM-V has highly conserved nucleotides at the same positions as SAM-II. In both structures two uracil carbonyl groups (U9 and U20 in SAM-V) point towards the sulfonium ion of SAM. Structure superposition of SAM-II and SAM-V binding pockets gave an RMSD of 0.373 Å (Huang and Lilley 2018). In contrast to SAM-II, SAM-VI has an additional stem P3. Additionally, P2a is shorter in SAM-V and exists ligand-independently. Due to these similarities and differences, SAM-II and SAM-V are grouped in different classes but in the same super-family.

Notably, SAM-V often occurs in tandem with SAM-II: the transcriptionally active SAM-II riboswitch is followed by a 20 nt sequence containing a transcription terminator (Yarnell and Roberts 1999) and the tandem is finished by a translationally active SAM-V riboswitch (Meyer et al. 2009). At high SAM concentrations, transcription of the downstream gene is regulated by SAM-II. At low SAM concentrations, the full transcript is made, and SAM-V regulates translation, when SAM concentrations increase later.

SAM-III and SAM-VI Riboswitches

SAM-III riboswitches were formerly annotated as S_{MK} -box until it was found that regulation was achieved directly on RNA level. SAM-III is found in Lactobacillales (Fuchs et al. 2006) and regulates translation of the *metK* gene. The SD base pairs with anti-SD sequence (aSD) upon ligand binding and thereby inhibits translation (Fuchs et al. 2007). The riboswitch (53 nt) forms a Y-shaped structure that composes of three parts: a long arm and two short arms. The short arms are built by stem P1 that holds the SD-anti-SD pair and P4 which is a linker helix. P2 and P3 stack on each other to build the long arm. The adenosyl group of SAM intercalates via π -stacking interaction at the three-way junction and stabilizes continuous base stacking of P1

and P2. The nucleobase adopts a *syn*-conformation, so its Watson-Crick edge base pairs with the sugar edge of G26 connecting P1 and P2. Also, the ligand is stabilized by an additional hydrogen bond to A73 of J2/4. The ribose 2'- and 3'-OH of SAM are hydrogen bonded to N7 and 2'OH of G89, respectively, and held in a 2'-*endo* conformation. Additionally, the 2'OH is in a water-mediated hydrogen bond to A74 of J2/4. As found for the previously mentioned SAM bind riboswitch super-families, the sulfonium ion interacts with carbonyl groups, but here with U72 and G71 and its own ribose O4'. Interestingly, the aminocarboxypropyl moiety has no specific interactions. It was shown that analogues that are missing the methionine moiety do bind to the riboswitch with near native affinity, as do analogues in which an ethyl group replaced the methyl group.

Notably, SAM-III is the first riboswitch that was crystallized with the ligand SAH and gave closer insights into discrimination mechanisms. The binding pocket and structure were found to remain unchanged. However, the SAH conformation within the binding pocket differs from SAM: the ribose rotates by 180°, the adenosyl group is positioned in an *anti*-conformation, the ribose 2'- and 3'-OH contact the phosphate backbone of G90 and the sulfoether of SAH rotates by 180° away from U72 toward G71 losing the key contacts of the sulfur atom in SAM. SAM-III binds the ligand in a lock-and-key mechanism and discrimination is achieved by strong contacts between the carbonyl groups of the RNA and the sulfonium ion of SAM.

The SAM-VI riboswitch is found in actinobacteria, specifically *Bifidobacterium*. The riboswitch was shown to have low similarity to SAM-III and lacks most highly conserved nucleotides found in SAM-III. Additionally, it has a different phylogenetic distribution than SAM-III riboswitches. Nevertheless, all nucleotides that contact SAM in SAM-III are also conserved in SAM-IV and both are found to form three stems (P1, P2 and a short P3). Just as in SAM-II, P3 is built up out of the SD and aSD to give SAM-IV the ability to be translationally active. Despite the severe differences, both riboswitches can be put in the same super-family (Mirihana Arachchilage et al. 2018). The recently solved crystal structure of the SAM-VI riboswitch showed that the two riboswitch classes not only differ in their phylogenetic distribution, but also in ligand recognition mode and folds. Therefore, the two riboswitch classes are grouped in different families. The structure of the SAM-VI riboswitch shows that even with high similarities in secondary structure (prediction), the three-dimensional representations can be highly different. Structure determination remains very important.

SAH Riboswitch

To date, only one SAH binding riboswitch is known which Breaker and coworkers discovered (Wang et al. 2008). Later, Batey and coworkers (Edwards et al. 2010) and coworkers solved its structure. This riboswitch was found in β - and γ -proteobacteria, but also in some δ -proteobacteria and gram-positive actinobacteria. The sequence is located upstream of genes involved in SAH recycling, namely *ahcY*, *methH*, and *metF* genes encoding for the SAH hydrolase, the cobalamin-dependent methionine synthase, and methylenetetrahydrofolate reductase (Wang and Breaker 2008). The riboswitches are shown to be ON switches acting on either transcription or translation of the adjacent gene.

The structure of the SAH riboswitch from *Ralstonia solanacearum* regulating transcription of *ahcY* bound to SAH forms into an LL-type pseudoknot with SAH binding at the junction between the three stems. Hereby, P2 and P4 coaxially stack and P1 is positioned perpendicular to the stack. The additional non-canonical helix P2b is formed from J2/1 and J4/2 and supports stacking of P2 and P4. SAH is bound in the minor groove of the P2b and P1 stack. The SAH adenosyl moiety forms a sheared G:A base pair with G15 and stacks between A29 and C16, like found in the SAM binding riboswitches. An electrostatic interaction between the α -amino of SAH and N3 of G30 stabilizes the aminocarboxypropyl group in the binding pocket. Additionally, the carboxylate is hydrogen bonded to the 2' hydroxyl groups of G47 and G31. The ribose of SAH is solvent assessable and not directly recognized. Finally, the sulfur atom is in 3.2 Å distance to the O4' of A29, forming a chalcogen interaction (Koebel et al. 2016). Ligand specificity is achieved by sterically hindering the binding of SAM: superposition of SAM into the binding pocket showed that the methyl group of SAM and C4' of A29 and N7 of G31 sterically clash.

SAM/SAH Riboswitch

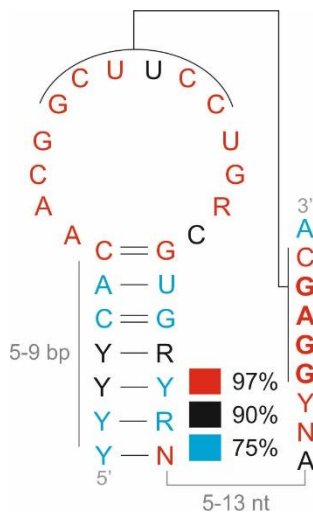


Figure 7: Secondary Structure of the SAM/SAH binding riboswitch consensus sequence. The riboswitch is a pseudoknot with a 5-9 nt long P1 helix and a 5-13 nt long L3 linker. The Shine-Dalgarno sequence is marked in bold letters. Nucleotides are conserved to either 97% (red), 90% (black) or 75% (blue). Non-conserved nucleotides are shown in black. N stands for any nucleotide, P for purine and Y for pyrimidine (adapted from Weinberg et al. 2010).

In 2010, Weinberg and colleagues discovered a riboswitch sequence that was found to bind SAM and SAH with similar affinities (Weinberg et al. 2010). Due to its encoded location upstream of mostly SAM associated genes (namely *metK*) it is assumed that SAM most likely regulates the riboswitch, since cellular concentrations of SAH are held low due to its cell toxicity. The SAM/SAH binding riboswitch is exclusively found in α -proteobacteria.

The riboswitch was found in inline-probing experiments to form an H-type pseudoknot structure that composes of a hairpin structure whose loop interacts with the 3'-end of the sequence to form the pseudoknot (Figure 7). The riboswitch acts as an OFF switch by incorporating the Shine-Dalgarno sequence into a putative tertiary structure upon ligand binding. The hairpin loop and the 3'-end of the sequence are highly conserved while the P1 helix and L3 loop are variable in length. SAM binding was assumed to take place within the unpaired hairpin loop region. The NMR structure determination of the SAM/SAH binding riboswitch is scope of this work.

3.3. CYCLIC DINUCLEOTIDE MESSENGER C-DI-GMP AND cGAMP BINDING RIBOSWITCHES

Cyclic dinucleotides (CDN) are second messengers and compose of two nucleotides which are connected cyclically via the phosphate group of one nucleotide to either the 2'- or 3'-hydroxyl group of the other nucleotide and vice versa. Considering the combinational possibilities of the four nucleotides linked to form a CDN, and the different linkages used, a vast number of possible CDN can be rationally designed. To date, only purine-containing CDN are known, namely c-di-GMP, c-di-AMP and c-AMP-GMP (cGAMP). c-di-GMP and cGAMP were first identified in bacteria (Ross et al. 1987), later it was shown that mammalian cells recognize the signaling molecule as an innate immune signal (Iwasaki and Medzhitov 2010). While in bacteria, the 3'-3'-linkage variants of all three were found to regulate biofilm formation, motility, and virulence, the 2'-3'-linkage isomers of c-AMP-GMP act as agonists of the innate immune response in mammalian cells. In bacteria, c-di-GMP was identified in several biological signaling pathways and is a key bacterial second

messenger. c-di-GMP has a two-fold symmetry. In solution, it is either found as an elongated monomer (at physiological conditions) or an intercalated dimer (high salt conditions, containing potassium ions). Like for the SAM-binding riboswitches, several classes of CDN-binding riboswitches were discovered, and structures of riboswitches binding to each known CDN are solved.

3.3.1. THE MAKE AND BREAK OF C-DI-GMP AND cGAMP

The antagonists diguanylate cyclase (DGC) and phosphodiesterase (PDE) specific to c-di-GMP are known to make and break the CDN, respectively. DGC is a homodimer that forms signal-dependently. It composes of two catalytic GGDEF domains that cooperatively synthesize c-di-GMP by asymmetrically binding two GTP molecules (Chan et al. 2004). The dimerization is regulated by negative feedback with growing c-di-GMP concentrations. cGAMP is found in bacteria containing a 3'-3'-linkage and mammalian cells containing a 2'-3'-linkage. The functional homologues, DncV and cGAS, are synthesizing the molecules 3'-3'-cGAMP and 2'-3'-cGAMP, respectively (Davies et al. 2012; Sun et al. 2013).

Two distinct CDN-degrading PDE classes are known: one contains an EAL, the other a HD-GYP domain. PDE containing an EAL domain degrade d-ci-GMP into the linear 5'-phosphoguananylyl-(3'-5')-guanosine (pGpG) which is later degraded into GMP (Christen et al. 2005). HD-GYP containing PDE directly hydrolyze c-di-GMP to yield two molecules of GMP (Ryan et al. 2006). To date, only HD-GYP containing enzymes are described to degrade cGAMP which analogously yield the linear molecule 5'-phosphoadenylyl-(3'-5')-guanosine (pApG) which later is degraded into AMP and GMP.

The reaction pathway in which c-di-GMP is involved are highly diverse. The dinucleotide c-di-GMP was found to be effector for several protein families and different ligand binding RNA, including riboswitches (Hengge 2010), transcriptional regulators (Krasteva et al. 2010), proteins containing PilZ domains (Boehm et al. 2010), and catalytically inactive GGDEF-EAL tandems (Duerig et al. 2009). Even though, many effectors are known, the full picture of the regulatory pathway remains unclear and potential new interaction are yet to be discovered, especially for the other known and (maybe) unknown CDN.

3.3.2. A COMPARISON OF DIFFERENT RIBOSWITCH STRUCTURES AND LIGAND RECOGNITION

c-di-GMP class I Riboswitches

The sequences for the c-di-GMP class I riboswitch occur upstream of genes encoding for diguanylate cyclase (DGC), phosphodiesterase (PDE), and other genes controlled by c-di-GMP (Sudarsan et al. 2008). In agreement with the ubiquitous use of c-di-GMP as a second messenger, many genes seem to be under the control of the c-di-GMP class I riboswitch.

The structure of the riboswitch aptamer domain was solved for a sequence from *Vibrio cholerae*, and the riboswitch named Vc2. The riboswitch regulates the *tfoX* gene, which is associated for example with cellular responses to starvation (Cameron et al. 2008; Redfield 1991) and transformation competence (Cameron et al. 2008; Sinha et al. 2013). The global riboswitch structure is Y-shaped with two parallel helices that are stabilized by a tetraloop-tetraloop receptor motif between helices P2 and P3 and an additional helix formed between the 5' and 3'-ends (Kulshina et al. 2009; Smith et al. 2009). The ligand-binding site is situated at the three-helix junction where the ligand is asymmetrically bound via its bases G_{α} and G_{β} , and its ribosyl phosphate. The base G_{α} is recognized via its Hoogsteen face to the Watson-Crick edge of G20. Additionally, the 3' oxygen of A48 and the 2'-hydroxyl group of C46 stabilize the N2 of G_{α} . The ribosyl phosphate is positioned in the binding pocket by contacts to the non-bridging phosphate oxygen of A47. Hereby, the G_{α} Watson-Crick edge remains solvent accessible. In contrast, the base G_{β} is in a canonical base pair to C92 and has an additional hydrogen bond between its N7 and the 2' hydroxyl group of A47 and its amino group and N1 of A18. The ribose moiety is stabilized by a hydrogen bond between the 2'-hydroxyl group of G_{β} and the amino group of A18. Metal atoms further coordinate the phosphate backbone. Here, metal ions were shown to have outer-sphere contacts to the 5' phosphate of G_{α} and A47, while an additional magnesium ion coordinates the 5' oxygen of G_{β} in an inner-sphere geometry. Like the c-di-GMP binding proteins, the c-di-GMP class I riboswitch highly stabilizes the ligand binding by stacking interactions. First, A47 intercalates between G_{α} and G_{β} . Second, G21 stacks on G_{α} while being in a canonical base pair with C46. Similarly, G14 stack underneath G_{β} while being in a base pair with C93. The ligand contacts to the terminal residues of P1 were found to initiate the formation of the P1 helix. In the absence of the ligand, only helices P2 and P3 are pre-formed.

The binding affinity of the riboswitch to c-di-GMP is very tight and depending on the used sequence and method ranges from 11 pM to 14 nM (Smith et al. 2009; Wood et al. 2012) and

therefore belongs to the tightest binders described in riboswitch research. Efforts to artificially switch the ligand specificity by mutating the ligand binding site, lead to the discovery of a G20A mutant that promiscuously binds c-di-GMP and its analogue c-GMP-AMP (cGAMP) with similar affinities (Smith et al. 2010).

c-di-GMP class II Riboswitches

Shortly after the discovery of the first c-di-GMP binding riboswitch, a second riboswitch responsive to c-di-GMP was described. The riboswitch was found in a tandem arrangement together with a splicing site junction for alternative RNA processing in *Clostridium difficile*. The riboswitches regulate the expression of a carbohydrate binding domain together with a cellulase and glycosyl hydrolase domain of proteins associated with c-di-GMP synthesis, degradation and signaling. In-line probing experiments indicated that the newly found riboswitch sequence is structurally distinct from the previously described one which was later supported by its crystal structure (Lee et al. 2010; Smith et al. 2011). The ligand binding aptamer domain folds into four helical stems (P1 to P4). The helices P1 and P4 stack coaxially on each other while P2 is orientated perpendicular to the stack. The loops of P3 and P2 interact to form a compact HH-type pseudoknot (intramolecular kissing hairpin) structure. The pseudoknot formation is facilitated by a kink turn in P3 that bends the helix toward P2. The bend is further stabilized by the interaction of the closing base pair of P3 (G46/U31) and the base pair between A34/U45 in a base quadruple, where the latter also directly stacks on the P4 stem. Furthermore, the major groove of stem P4 interacts with J2/4 to form a triple helix. The ligand c-di-GMP is bound in the junction between the helices P1, P2 and P4-J2/4 and is incorporated into the triple helix.

The c-di-GMP nucleobase G_{α} interacts with the sugar edge of A69 which is in a Watson-Crick base pair to U37. G_{β} is coordinated by C73, a magnesium ion and the phosphate of A70. Additionally, the nucleobase of A70 intercalates between G_{α} and G_{β} like the ligand binding mode in the c-di-GMP binding riboswitch of class I. Both, G_{α} and G_{β} are further stabilized by stacking interaction to the closing base pairs of neighboring stems P1 and P4. Interestingly, G_{α} is incorporated in an antiparallel manner compared to the other nucleotides of the stack. The architecture is achieved by a U- and a S-turn that result in a groove reversal. While the other nucleobases interact via the major groove of the stems to form the triple helix, G_{α} forms a minor groove triplet. While the nucleobases of c-di-GMP are extensively recognized by the riboswitch RNA, the phosphodiester linkages are only minorly contacted by a single hydrogen

bond between the amino group of A70 and the non-bridging oxygen of c-di-GMP, while in class I the phosphodiester is contacted by an ionic interaction. In comparison to the c-di-GMP binding riboswitch of class I, the ligand is generally recognized differently, but using the same chemical principles of asymmetrical ligand binding and extensive stacking interaction. Additionally, the riboswitch itself forms a stable tertiary structure and thereby promotes high affinity ligand binding.

c-GAMP Binding Riboswitches

Rational mutation experiments on riboswitches often have a different outcome than expected. The c-di-GMP class II riboswitch Vc2 G20A mutant of the c-di-GMP binding riboswitch has an altered ligand specificity. In contrast to the wild-type riboswitch whose native ligand is c-di-GMP, the G20A mutant binds c-AMP-GMP and c-di-GMP with similar affinities. Genes associated with c-di-GMP regulation are responsible for pili formation, chemotaxis, signaling and synthesis or degradation of c-di-GMP. Some of the riboswitches have an association with genes that are known to be controlled by c-di-AMP (Nelson et al. 2015). This prompted an extensive bioinformatic re-validation of c-di-GMP riboswitch sequences to group the sequences depending on the base identity at position 20. Interestingly, half to a fourth of the sequences harbor an adenosine at the equivalent position for G α recognition, depending on investigated species (Kellenberger et al. 2015; Nelson et al. 2015). Even more surprising, most of the found sequences are promiscuous, while others are selective for either c-di-GMP or c-AMP-GMP, respectively (Kellenberger et al. 2015). For example, the identified riboswitch Gs1761 from *Geobacter sulfurreducens* has a 1,200-fold higher affinity for c-AMP-GMP than for c-di-GMP and do discriminate against linkage isomers just as found for the c-di-GMP class I riboswitch.

Structural investigations of the mutant riboswitch bound to c-AMP-GMP (with an additional mutation in stem P1 due to a sequencing error) revealed a similar tertiary structure as the c-di-GMP class I riboswitch bound to c-di-GMP. The overlay of both structures shows only minor differences (Figure 8). Additionally, the riboswitch structures bound to c-GAMP or c-di-GMP were solved, respectively and they are very similar.

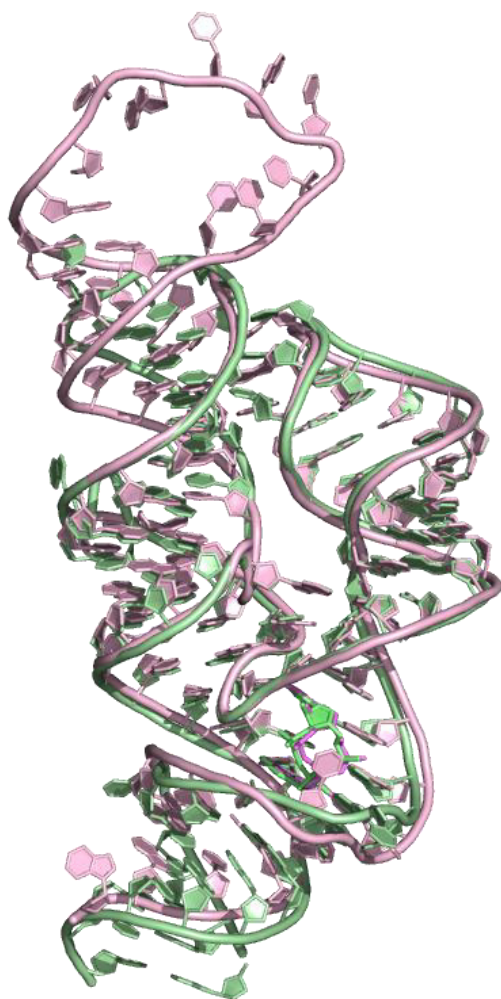


Figure 8: Overlay of c-di-GMP binding riboswitch (pink, 3IWN) and cGAMP binding riboswitch (green, 4YB0) both bound to c-di-GMP (shown in stick representation).

A closer investigation of the binding pockets of the two states showed that ligands are bound similarly. In the c-di-GMP bound state, the base of A48 is not base paired to G_{α} . Instead, it is twisted out of plane to enable hydrogen bonding between its 3'-hydroxyl group and the N2 amino proton of G_{α} . This causes the stems P2 and P3 to be twisted by 15-20°. The different behaviors concerning promiscuity could be caused by the orientation of P2 and P3 in the two states. The structure comparison of the native c-di-GMP class I riboswitch complex and its G20A mutant bound to c-AMP-GMP showed even lower differences than the previously mentioned overlay. The G20A mutant does bind c-di-GMP and c-AMP-GMP with similar affinities.

Furthermore, the distance between G_{α} O6 and N1 of A20 in the GS1761 riboswitch is in the range of 3.0 Å. The distance could be bridged by a proton to connect to two hydrogen bond acceptors and to ensure a hydrogen bond. Two cases are likely, either the guanine base is bound in a rare tautomeric conformation or the ligand binding A20 is protonated to enable binding. However,

the origin of the needed proton remains unclear since crystallization methods do not directly detect protons due to low resolution. Investigating the ligand binding mode and elucidating a putative protonated state is scope of this work.

4. ITC ON RNA – BEYOND AFFINITY MEASUREMENTS

Calorimetric methods have been used in life sciences for decades. The methods rely on the property that most chemical reactions are either associated with the release (exothermic) or accumulation (endothermic) of heat, a value that can be measured and interpreted. Isothermal Titration Calorimetry (ITC) uses this effect. The ITC instrument has two adiabatically shielded cells – a reference and a measurement cell. The reference cell is filled

with solvent while the measurement cell contains an appropriate amount of reactant, none of which must be chemically labeled. During the experiment, an appropriate amount of ligand is added into the measurement cell in a stepwise manner. The increments must be chosen according to the system. Ideally, during the first injection, all titrant molecules bind to the macromolecule releasing or accumulating a relatively high amount of heat. The amount of energy needed to equilibrate the temperature in both cells is recorded by the system. While the experiment proceeds, more titrant is added, and the heat released or accumulated decreases until the system reaches equilibrium. The ligand to macromolecule ratio is plotted against the energy to give a sigmoidal graph from which the binding affinity (K_d), number of bound ligands per macromolecule (stoichiometry, n) and the enthalpy of the reaction (dH) can be directly be calculated; entropy (dS) is calculated indirectly.

The systems investigated using ITC are highly diverse. Most commonly in life sciences, ITC experiments are used to investigate ligand-macromolecule interactions in an aqueous environment. Therefore, water is used as a reference and macromolecules dissolved in buffer is added to the measurement cell to which the (putative) ligand is titrated (Lieberman et al. 2014).

While the thermodynamic parameters already are an interesting information of a system, ITC can be used to further dive into the system and deepen the knowledge about it. As mentioned before, some macromolecules or ligand use protonation for efficient binding or discrimination between ligands. The characterization of protonated molecules is challenging. They are not directly detectable in crystal structure since hydrogens are only modelled into the structure. But even if the existence is known from other experiments, its stability is of interest and subsequently the pK_a value must be determined. Using ITC experiments is an easy way to determine the value while spending a relatively low amount of sample. To this aim, a series of standard ITC experiments is executed using a range of buffers with varying pH. The determined K_d is plotted against the pH. The pK_a is determined from the slope of the resulting graph. Furthermore, e.g., for riboswitches and aptamers ligand binding characterization and associated with its insights into the ligand binding mode are of interest. Especially, when closely related analogue ligands are rejected, and the structure of the riboswitch-ligand complex is unknown. ITC experiments can give closer insights into the binding pocket even before the structure is solved (Feig 2007).

Additionally, in macromolecule structure determination screening for a construct with high binding affinity is one of the key success factors in NMR and crystal structure determination. The higher the binding affinity, the better resolved NMR spectra are. ITC is a method by which the binding affinities can be screened considerably fast to yield the best binding construct to start the structure determination. Also, structure properties like flexible linkers have a high impact on the binding affinity. Rational design considerations combined with an ITC screening can lead to a good starting candidate for promising structure determination.

RNA is unstable and handling is facilitated by using an appropriate buffer. Buffer conditions like pH, salt concentration and presence of certain ions can be screened for resulting in an appropriate system used. As mentioned before, varying the pH can help identify and characterize protonated bases or ligands. Furthermore, in some RNA structure G-quadruplexes were identified as mentioned above. Here, potassium ions are crucial for formation of this structure.

From the ITC thermogram not only the thermodynamic parameters can be calculated but it also gives a good insight into sample quality. Firstly, the stoichiometry is a good indicator. If the n value is significantly below 1 (for a 1:1 binder) the sample or ligand quality is questionable. The sample is inhomogeneous, and some populations are most likely misfolded. In this case, revising the sample preparation process would be advisable (Salim and Feig 2009). ITC also is used to investigate the kinetic properties of systems using single injection mode for enzyme activity or kinITC (Burnouf et al. 2012). The methods in contrast to the classical approach do analyze the slope of the peak returning to equilibrium and derive dynamic data from it.

5. RNA STRUCTURE DETERMINATION USING NMR – FOR WELL-BEHAVED AND CHALLENGING RNA

High-resolution structures of macromolecules are mainly solved by crystallography; to date, about 90% of the structures according to statistics of the protein data bank (PDB). The resulting structures give a picture of the macromolecule in the crystal environment and, unfortunately, lack the exact positions of hydrogen atoms. They are later modeled into the structure. In most cases, the information retrieved from crystal structures fit the scientific needs. Nevertheless, some structures are affected by crystal packing effects and are not in agreement with structures that were obtained under physiological conditions. Especially,

protonated residues are not directly detectable but assumed by close proximities of potential hydrogen bond donors and acceptors. Recently, advances in Cryo-electron microscopy (Cryo-EM) became more interesting in macromolecule structure determination. For complexes about 40 kDa and larger, like the ribosome, impressive insights were obtained and cryo-EM developed to the method of choice in ribosome investigations (Brown and Shao 2018). To overcome the size limitation, methods involving a cargo protein were developed (Liu et al. 2019). The method gives many new possibilities in contrast to crystallography as it is not limited to crystals, needs lower concentrations, and is not as sensible to sample heterogeneities. The obtained resolutions are in the range of crystal structures. With growing method development also RNA structure determination which is challenging due to flexibility and size will be in the scope of Cryo-EM structure determination, as was shown for the structure of SAM-IV (Zhang et al. 2019).

In contrast, solution NMR structures rely on proton-proton distance information measured under physiological conditions. In addition to pure structural information, dynamic parameters can be derived as well as protonated states residues can be directly identified. As all methods, NMR structure determination has its limits. Nucleic acid structures can be obtained for molecules up to 100 nt, high-resolution structure only up to 50 nt. Also, ligand binding molecules should have a reasonably high binding affinity and only a limited degree of freedom and flexibility.

5.1. FROM RESONANCE ASSIGNMENT TO STRUCTURE – THE TRADITIONAL APPROACH

In crystallography, a density map is obtained from the crystal and its scattering. Simply put, the density map is equal to the surrounding of the macromolecule sequence. Consequently, the sequence is modeled into the density map. In contrast, in NMR structure determination a collection of spectra is recorded that contain the resonances for distinct nuclei e.g., ^1H directly bound to either ^{13}C or ^{15}N , respectively. The chemical shift of these nuclei-to-nuclei correlation and the combination of the different correlations gives insights into its putative identity. One of the first one-bond spectra recorded during the structure determination process, the $^1\text{H},^{13}\text{C}$ -HSQC, is either optimized to the aromatic or aliphatic resonances. In the aromatic $^1\text{H},^{13}\text{C}$ -HSQC spectra of an RNA, the A-H2C2 signals occur downfield while the G-H8C8 signals are upfield shifted. Between the two resonance regions, the signals for A-H8C8, C/U-H5C5 and C/U-H5C5 can be identified. Since for the pyrimidine residues a C5-C6 coupling is observed, the A-H8C8 signals are distinguishable from the C/U-H5C5 and C/U-H5C5 signals due to the signal splitting.

This and the other one-bond spectrum types show comparable chemical shift differences and allow conclusions about nucleotide identity. Already in these first experiments, signal overlap is a severe problem especially for increasing number of nucleotides. To overcome this problem, selectively ^{15}N - or ^{13}C , ^{15}N -labeled samples can be used that only contain one or two ^{13}C - and/or ^{15}N -labeled nucleotide type. Proceeding in the assignment process, different nuclei within each nucleotide can be correlated which leads to a collection of spin systems. One-bond correlations are standard experiments while the success of long-range experiments strongly depends on intrinsic RNA properties (e.g., relaxation, binding affinity, size) and therefore in some cases are difficult or even unsuccessful (Barnwal et al. 2017; Fürtig et al. 2003).

The resulting spin systems must be put in a context of the RNA sequence. The sequential assignment is either achieved by through-bond correlation via the phosphate backbone or by identification of hydrogen-hydrogen distances that imply intra-residual distances. The first option needs a full assignment of the phosphate signals which in many cases is only successful for comparably small and tightly binding riboswitches. The latter option is well established and was used before the sophisticated through-bond methods were developed. It is suitable for riboswitches at the high size limit but is more biased by the interpretation of signal intensities. In both cases, the inter-residual distance information is mandatory for the following structure calculation.

In addition to the distance information, the structure determination benefits from information on the sugar conformation of the RNA backbone. The sugar pucker can be calculated from the full assignment and the resulting chemical shift information of the aliphatic carbon atoms resulting in the canonical coordinates that are interpreted accordingly (Ebrahimi et al. 2001; Fürtig et al. 2003; Ohlenschläger et al. 2008).

To achieve a structure determination from NMR data, resonance assignment is followed by the extraction of NOE distance information from several different NOESY spectra. The distances derived from NOE intensities are group into weak, intermediate, and strong intensities and accordingly into long, intermediate, and close proximities. The structure bundle resulting from the structure calculation process represents the possible structures within the limits of the distance information and most likely also the flexibility of the structures under physiological conditions in solution.

5.2. WHEN YOUR RNA DOESN'T BEHAVE AS YOU WANT IT TO

For RNA with high degree of freedom (e.g., structures containing flexible linker) in combination with a low binding affinity, some of the traditional approaches do not yield high resolution spectra or spectra from which not all spin systems can be assigned. As mentioned above (Chapter 4), a mutant screening approach should be considered using ITC or related methods. In some cases, even point mutations can increase the binding affinity significantly without alternating the structure and thereby simplify the structure determination.

Due to the limited number of distinct building blocks and spectral crowding associated with it, NMR signal overlap is a large problem. Therefore, base-type specific labeling approaches became a standard procedure in the last years (Fürtig et al. 2003). Additionally, chemically synthesized site-specifically labeled samples that contain only one specifically labeled residue help identifying signals unambiguously (Wunderlich et al. 2012; Wunderlich et al. 2015).

Depending on the size and structure of the RNA, divide and conquer approaches can be considered. Here, the full structure is divided in smaller fragments and their structures are solved individually. The full structure is derived from the smaller parts. The method heavily relies on a reliable secondary structure investigation. Anyway, structurally independent fragments are crucial for the success when considering a divide and conquer approach. Nevertheless, even large structures can be solved using this technique as was shown e.g., for the HCV IRES domain II (Lukavsky et al. 2003).

Similarly, it has been shown that structure determination benefits from sequential labelling (Lu et al. 2010a). Here, individual strands of the RNA structure are synthesized labeled or unlabeled and chemically linked or annealed to each other. While it might result in drastically reduced binding affinity when an annealing process is chosen, the method reduces signal overlap and can help to identify specific parts of the sequence. Furthermore, the combination of unlabeled and labeled fragments as in base-specific labeled samples give the possibility for filtering experiments. The method was e.g., successfully used in the investigation of the GTP class V aptamer (Nasiri et al. 2016). Here, the helical part of the aptamer was investigated using a labeled strand that was annealed to the unlabeled corresponding strand.

6. AIM OF THIS WORK

Insights into RNA structures are upcoming in the last decades. Nevertheless, they lack behind protein structures. The knowledge and putative structure predictions heavily rely on the data

base of known structures. Therefore, structures with uncommon binding pattern and intriguing use of the nucleobases' chemical properties are of interest.

In crystal structures the hydrogen atoms are only modeled into the resulting structures and hydrogen bonds are assumed by spatial proximity of known hydrogen donor and acceptor. The protonation of nucleobases is often not considered due to their unfavorable pK_a values that are far from neutrality. Therefore, the discovery and description of putative protonation heavily relies on the assumption of the investigator. For riboswitch ligand binding events, ITC experiments with a range of pH buffer can give indications for a protonated macromolecule. The direct, unambiguous detection of protonated nucleobases to this date is only possible by NMR. In this work, ITC experiments were used as a foundation for the characterization of two different ligand binding RNA structures that used protonated nucleobases.

SAM binding riboswitches belong to the most abundant riboswitches in bacteria. They are reported to be highly selective against related analogues as is the only known SAH binding riboswitch. The ligand recognition strategies are thereby highly diverse. In contrast to the well described SAM binding riboswitches, the SAM/SAH binding riboswitch was described to bind SAM and SAH with similar affinities. Given the biological function of the two analogues and the gene association of the riboswitch the reason for promiscuous binding remained unclear. Structural insights and the bioinformatical research could explain the found phenomena. Solving the structure of the SAM/SAH binding riboswitch was aim of this work.

III PUBLICATIONS

1. WOLTER ET AL – STRUCTURE OF THE GTP CLASS II APTAMER

Title: A Stably Protonated Adenine Nucleotide with a Highly Shifted pKa Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer

Authors: Wolter AC, Weickmann AK, Nasiri AH, Hantke K, Ohlenschläger O, Wunderlich CH, Kreuz C, Duchardt-Ferner E, Wöhnert J

Published in: Angew. Chem. Int. Ed. 2017, 56, 401-404

Contributions

(1) Entwicklung und Planung	Wolter AC (25%), Weickmann AK (15%), Duchardt-Ferner E (20%), Wöhnert J (40%)
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Wolter AC: RNA-Präparation, NMR-Messungen; Weickmann AK: ITC-Messungen; Nasiri AH: Initiale Klonierung, RNA-Präparation, NMR-Messungen; Hantke K: Klonierung, RNA-Präparation; Wunderlich CH: Präparation selektiv markierter Proben; Duchardt-Ferner E: NMR-Messungen
(3) Erstellung der Datensammlung und Abbildungen	Wolter AC: Erstellung von Abbildungen; Weickmann AK: Erstellung von Abbildungen; Wöhnert J: Erstellung von Abbildungen
(4) Analyse/ Interpretation der Daten	Wolter AC: Auswertung der NMR-Messungen, Strukturrechnung; Weickmann AK: Auswertung der ITC-Messungen; Duchardt-Ferner E: Auswertung der NMR-Messungen; Wöhnert J: Auswertung der NMR-Messungen; Ohlenschläger O: Strukturminimierung
(5) Übergeordnete Einleitung/ Ergebnisse/Diskussion	Wolter AC (30%), Weickmann AK (15%), Duchardt-Ferner E (10%), Wöhnert J (45%)

A Stably Protonated Adenine Nucleotide with a Highly Shifted pK_a Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer

Antje C. Wolter, A. Katharina Weickhmann, Amir H. Nasiri, Katharina Hantke, Oliver Ohlenschläger, Christoph H. Wunderlich, Christoph Kreutz, Elke Duchardt-Ferner, and Jens Wöhnert*

Dedicated to Professor Harald Schwalbe on the occasion of his 50th birthday

Abstract: RNA tertiary structure motifs are stabilized by a wide variety of hydrogen-bonding interactions. Protonated A and C nucleotides are normally not considered to be suitable building blocks for such motifs since their pK_a values are far from physiological pH. Here, we report the NMR solution structure of an *in vitro* selected GTP-binding RNA aptamer bound to GTP with an intricate tertiary structure. It contains a novel kind of base quartet stabilized by a protonated A residue. Owing to its unique structural environment in the base quartet, the pK_a value for the protonation of this A residue in the complex is shifted by more than 5 pH units compared to the pK_a for A nucleotides in single-stranded RNA. This is the largest pK_a shift for an A residue in structured nucleic acids reported so far, and similar in size to the largest pK_a shifts observed for amino acid side chains in proteins. Both RNA pre-folding and ligand binding contribute to the pK_a shift.

The complex tertiary structures of functional RNAs such as rRNA, ribozymes, and riboswitches are stabilized by extensive hydrogen-bonding interactions. The structural diversity observed in these RNAs is based on a wide variety of hydrogen-bonding interactions, including not only Watson–Crick hydrogen bonds but all putative hydrogen-bond donor and acceptor groups. Protonated, positively charged A and C nucleotides can further increase the diversity of hydrogen bonding interactions in RNA. While the pK_a values for the

protonation of A (ca. 3.7) and C (ca. 4.5) in single-stranded RNA are far away from neutrality, the structural environment in folded RNAs can shift these pK_a values significantly.^[1] Thus, a number of biologically important RNAs indeed contain protonated C or A nucleotides (Table S1 in the Supporting Information).^[2] For instance, the hepatitis delta virus (HDV), hairpin, and Varkud satellite (VS) ribozymes feature C or A nucleotides with pK_a values between approximately 6 and 7 in their active sites, where these nucleotides serve as the catalytic acid in the ribozyme reaction.^[3] Viral pseudoknot RNAs that contribute to ribosomal frameshifting also harbor protonated C or A residues with pK_a values close to 7.^[4] Here, these protonated nucleotides apparently tune conformational equilibria between frame-shifting and non-frame-shifting conformations.

Furthermore, A residues of A:C or A:G mismatches embedded in stable Watson–Crick double helices are sometimes protonated.^[5] Under physiological conditions, dynamic equilibria between protonated and deprotonated states are present in all these RNAs. However, owing to the small number of examples of RNA structures containing protonated adenines described so far, a full appreciation of their structural and functional potential is still lacking.

Structures of RNA aptamers in complex with their ligands have revealed fundamental principles of RNA folding and RNA–ligand interactions.^[6] GTP (Figure 1a) has been a particularly versatile target for raising RNA aptamers in SELEX experiments. Szostak and co-workers reported no less than 11 different families of GTP-binding aptamers.^[7] For one of them, a complete three-dimensional structure has been reported, and for another, it was shown that the aptamer binds GTP through an intermolecular G-quadruplex.^[8] Herein, we report the NMR solution structure of the class II aptamer bound to GTP. The bound structure features an unusual base quartet stabilized by a stably protonated A residue.

The predicted secondary structure of the class II aptamer is deceptively simple (Figure 1b). An asymmetric internal bulge (A8–A13, A26, A27) separates a terminal helix from a short central helix that is capped by a formally single-stranded apical loop (A16–C23). Imino proton spectra indicate that GTP binding induces tertiary-structure formation in the RNA (Figure 1c; also see Figure S1 in the

*] A. C. Wolter, A. K. Weickhmann, A. H. Nasiri, K. Hantke, Dr. E. Duchardt-Ferner, Prof. Dr. J. Wöhnert
Institut für Molekulare Biowissenschaften and
Zentrum für Biomolekulare Magnetische Resonanz (BMRZ)
Goethe-Universität Frankfurt
Max-von-Laue Str. 9, 60438 Frankfurt/Main (Germany)
E-mail: woehnert@bio.uni-frankfurt.de

Dr. O. Ohlenschläger
Leibniz Institut für Altersforschung (Fritz-Lipmann-Institut)
Beutenbergstrasse 11, 07745 Jena (Germany)

Dr. C. H. Wunderlich, Prof. Dr. C. Kreutz
Institute of Organic Chemistry, Centre for Molecular Biosciences
(CMBI), University of Innsbruck
Innrain 80/82, 6020 Innsbruck (Austria)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under <http://dx.doi.org/10.1002/anie.201609184>.

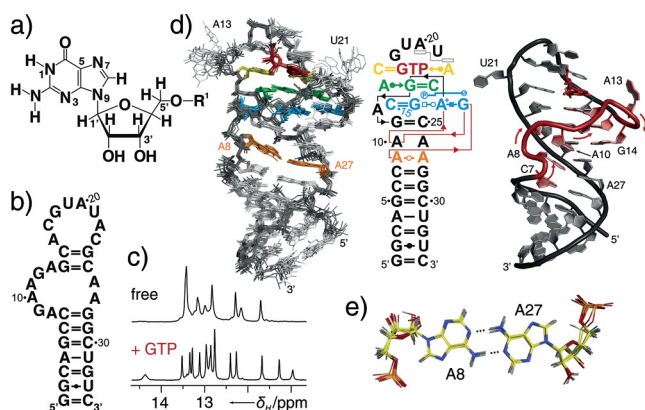


Figure 1. Overall structure of the class II aptamer/GTP complex. a) Structure of GTP (R^1 = triphosphate). b) Predicted secondary structure of the class II aptamer. c) Comparison of $1D$ - 1H -imino proton spectra (2 mM Mg^{2+} , pH 6.3, 20°C) for the RNA in the absence (top) and presence (bottom) of GTP. d) Left: Superimposition of the 10 RNA/GTP complex structures with the lowest target function. GTP is shown in red, C17 and A22 in yellow, the A8:A27 base pair in orange, a base quartet in blue, and a base triplet in green. Middle: Structure diagram of the complex structure using the Leontis–Westhof notation with the same coloring. Right: Average structure with the backbone segment from C7 to C15 shown in red and arrows indicating its local direction. e) The *trans*-A8:A27 base pair in the internal bulge in a bundle representation.

Supporting Information). The high-resolution NMR structure of the GTP-bound class II aptamer presented here (Figure 1 d and Table S2) reveals a compactly folded rigid RNA structure (Figure S2). Nucleotides G9, A11, G12 and A13 are extruded from the internal bulge and are involved in long-range tertiary interactions with nucleotides from the central helix or the apical loop. This leads to a very intricate path for the backbone in the RNA segment spanning nucleotides A8 to G14 (Figure 1 d), which allows, for instance, the formation of an unusual *trans*-A8:A27 base pair (Figure 1 e and Figure S3).

GTP is bound in the apical loop in an intermolecular base triplet with C17 and A22 (Figure 2, Figure S4). C17 and GTP form a Watson–Crick base pair, while A22 binds the sugar edge of GTP. The floor of the GTP-binding site consists of an intramolecular base triplet comprising G12, A16, and C23 (Figure 2 and Figure S4). Directly below lies a base quartet of

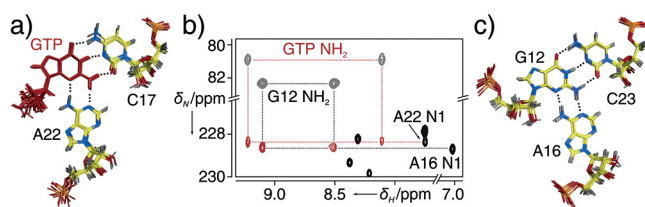


Figure 2. The GTP-binding site. a) GTP recognition in a base triplet with C17 and A22. Hydrogen bonds are indicated by dashed lines. b) Overlay of the N7/N3/N1 region of a long-range 1H , ^{15}N -HSQC (black) and an HNN-COSY-spectrum (grey, red) recorded at 20°C, pH 6.3, demonstrating the presence of hydrogen bonds between the GTP and G12 amino groups and the N1 nitrogen atoms of A22 and A16, respectively. c) Structure of the G12:C23:A16 base triplet at the floor of the GTP-binding site.

hitherto unprecedented geometry. It includes the C15:G24 Watson–Crick base pair of the central helix, as well as G9 and A11 from the internal bulge (Figure 3 a). A11 is protonated at its N1 nitrogen atom (Figure 3 b and Figure S5). Remarkably, the NMR signal for the A11⁺ imino proton is directly

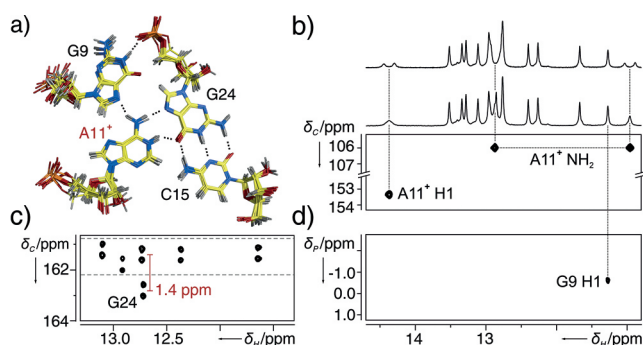


Figure 3. The G9:A11:C15:G24 base quartet contains a protonated adenine nucleotide. a) Structure of the base quartet. Hydrogen bonds are indicated by dashed lines. b) Imino region of ^{15}N -coupled (top) and decoupled (middle) $1D$ - 1H spectra and a 1H , ^{15}N -HSQC spectrum (bottom) for a ^{15}N -A-labeled RNA, revealing the signals for the A11⁺ imino and amino groups. c) An H(N)C-correlation experiment shows a downfield shift for the G24 C6 carbonyl group, which is in line with its role as a hydrogen-bond acceptor. d) A cross-peak in a long-range 1H , ^{31}P -HSQC between the G9 imino group and the 5'-phosphate group of G24 demonstrates the presence of a hydrogen bond between these groups.^[9] All spectra shown were recorded at pH 6.3, 20°C, 2 mM Mg^{2+} .

observable at 14.3 ppm. The N1 imino and N6 amino groups of A11⁺ are hydrogen bonded to the C6 carbonyl group (Figure 3 c) and the N7 nitrogen atom of G24, respectively. The A11 amino group also interacts with the N7 nitrogen atom of G9 (Figure S5). The G9 imino group is hydrogen bonded to the 5'-phosphate group of G24 (Figure 3 d). The hydrogen-bonding features in the base quartet were derived directly from dedicated NMR experiments (Figure 3 and Figure S5).

NMR signals for A⁺ imino protons are normally not observable at ambient temperatures and pH as used here (20°C, pH 6.3, 2 mM Mg^{2+}). In the case of the class II aptamer/GTP complex, however, the A11⁺ imino proton resonance remains observable up to 40°C (Figure 4 a). Furthermore, it was detected even in a 1H -spectrum recorded at pH 8.3 and 20°C, and the 1H , ^{13}C -HSQC-spectra at pH 6.3 and 8.3 are almost indistinguishable (Figure 4 b), thus showing that A11 remains protonated even at a pH well above neutrality. Apparently, the unique structural environment in the base quartet strongly stabilizes the protonated A11⁺. The binding affinity (K_D) of the class II aptamer for GTP as measured by isothermal titration calorimetry (ITC) shows a clear pH dependence (Figure 4 c, Table S3, and Figure S6). By using the thermodynamic cycle shown in Figure 4 c, the pK_a for A11 protonation in the free aptamer as well as in the GTP-bound RNA was estimated. This approach yielded a pK_a value in the free RNA of 6.7 and a lower limit for the pK_a in the GTP-bound state of 8.9 (Figure S6). Both pK_a values are significantly elevated compared to those for A residues in single-

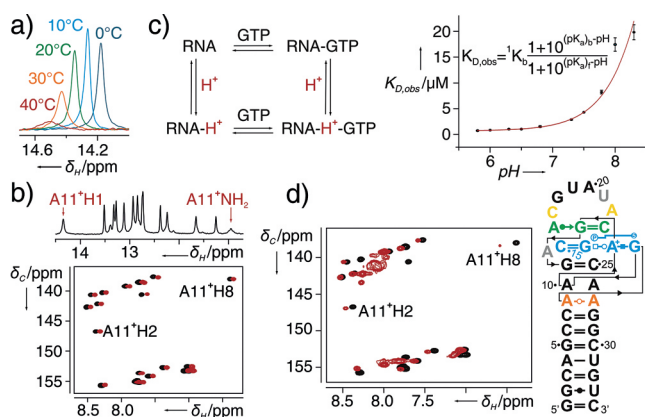


Figure 4. Properties of A11⁺. a) The A11⁺ imino proton is observable even at elevated temperatures in 1D-¹H imino proton spectra of the RNA/GTP complex at pH 6.3. b) 1D-¹H- (top) and ¹H,¹³C-HSQC (bottom, red) spectra of the complex at pH 8.3 and 20°C show stable A11 protonation at elevated pH values. The ¹H,¹³C-HSQC-spectrum at pH 6.3 is also shown (black). The two spectra are intentionally shifted by one line-width with regard to each other to allow a comparison. c) Thermodynamic cycle describing GTP binding and protonation of the aptamer (left). The pH dependence of the K_D for GTP of the class II RNA aptamer (right). The K_D changes from 0.7 μM at pH 5.8 to 19.8 μM at pH 8.3. The inset shows the formula to derive the pK_a values for the protonation of the free and the GTP-bound RNA.^[11] d) At pH 5.3 and in the presence of Mg²⁺, the class II aptamer is partially prefolded and A11 is already protonated, even in the absence of GTP. Left: Overlay of ¹H,¹³C-HSQC-spectra of ¹³C-A selectively labeled RNA at pH 5.3, 20°C, 2 mM Mg²⁺ in the absence of GTP (red) and at pH 6.3, 20°C, 2 mM Mg²⁺ in the presence of GTP (black). Right: Structure diagram for the pre-folded ligand-free RNA at pH 5.3.

stranded RNA. While the pK_a of A11 in the free RNA is similar to values found for protonated A residues in the U6 ISL spliceosomal RNA, viral frameshifting pseudoknots, and the leadzyme, for example, a pK_a value for adenine protonation of more than 8.9 has so far not been reported.^[4,5,10] These data also imply that A11 in the free RNA should be protonated at pH < 6.7. Accordingly, the ¹H,¹³C-HSQC-spectrum in the absence of GTP at pH 5.3 (2 mM Mg²⁺) does show the H2C2 signal characteristic for protonated A11. Furthermore, the chemical shifts of additional adenine H2C2 and H8C8 groups of the free RNA resemble those in the GTP/RNA complex at pH 6.3 (Figure 4d and Figure S7a,e). In particular, this includes A8, A10, A26, and A27 from the central bulge and A16 as part of the G12:A16:C23 base triplet. In contrast, the chemical shifts for A20, which is directly involved in ligand binding, and A22, which is found in the apical loop, are different between the free RNA at pH 5.3, 2 mM Mg²⁺, and the GTP-bound complex. In addition, imino group signals are observable in the free RNA at pH 5.3, 2 mM Mg²⁺ for A11⁺ and G12 from the G12:A16:C23 base triplet; G9, which forms the hydrogen bond to the phosphate backbone in the base quartet; and G14 and G24 from the central helix (Figure S7b). Taken together, these data suggest that the central bulge, the base quartet, and the base triplet at the floor of the ligand-binding pocket are already preformed in the free RNA at pH 5.3 in the presence of Mg²⁺. At pH 6.3, spectra of the free RNA show broad signals (Figure S7). This

indicates the presence of a conformational equilibrium involving unfolded/unprotonated and folded/A11-protonated conformations, which might then be selected by GTP for binding.

In summary, our structure of a class II RNA aptamer bound to GTP reveals the structural basis for GTP recognition, as well as the integral architectural role of a protonated adenine in the context of a unique base quartet. This demonstrates that protonated adenine nucleotides can be important building blocks of RNA tertiary-structure elements even at a pH values well above neutrality. The observed pK_a shift of more than 5 pH units for A11 is comparable to the largest pK_a shifts reported for amino acid side chains in the active sites of proteins.^[12] Protonated cytidines can also stabilize RNA tertiary structure elements.^[13] Therefore, the occurrence of protonated nucleotides should be more routinely considered in efforts to predict RNA tertiary structures from sequences, as well as during the refinement of experimental RNA structures. The class II GTP aptamer can also be considered as an RNA element capable of reacting to two different external stimuli (pH and GTP concentration) and integrating them into a single output signal (stable folding) using AND logic. Thus, this RNA might be useful as a building block for switching and sensor elements in synthetic biology and as a pH-sensitive building block for RNA nanoarchitectures.

Experimental Section

RNA synthesis, NMR sample preparation, and NMR assignment and data analysis were described in detail previously, and NMR signal assignments were deposited in the BMRB (entry 25661).^[14] The ITC and the NMR experiments for structure determination, as well as structure calculation methods, are described in the Supporting Information. Structure coordinates are available from the PDB (PDB ID: 5LWJ).

Acknowledgments

We are grateful to Sina Kazemi, Jan Philip Wurm, and Boris Fürtig for helpful discussions, and Kerstin Yacoub for help with sample preparation. This work was supported by the Center for Biomolecular Magnetic Resonance (BMRZ) at the Goethe University Frankfurt and the Deutsche Forschungsgemeinschaft (WO901/1-1 to J.W. and the Collaborative Research Center (SFB) 902 “Molecular principles of RNA based regulation”). The Fritz Lipmann Institute is financially supported by the Federal Government of Germany and the State of Thuringia.

Keywords: aptamers · GTP · NMR spectroscopy · protonated adenine · RNA structures

[1] a) W. Saenger, *Principles of nucleic acid structure*, Springer, New York, 1988, p. 108; b) L. E. Kapinos, B. P. Opershall, E. Larsen, H. Sigel, *Chem. Eur. J.* 2011, 17, 8156–8164; c) A. Ren, N.

- Vusurovic, J. Gebetsberger, P. Gao, M. Juen, C. Kreutz, R. Micura, D. J. Patel, *Nat. Chem. Biol.* **2016**, *12*, 702–708.
- [2] a) A. Huppler, L. J. Nikstad, A. M. Allmann, D. A. Brow, S. E. Butcher, *Nat. Struct. Biol.* **2002**, *9*, 431–435; b) D. D. Cash, O. Cohen-Zontag, N.-K. Kim, K. Shefer, Y. Brown, N. B. Ulyanov, Y. Tzfati, J. Feigon, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10970–10975; c) M. Pechlaner, D. Donghi, V. Zelenay, R. K. O. Sigel, *Angew. Chem. Int. Ed.* **2015**, *54*, 9687–9690; *Angew. Chem.* **2015**, *127*, 9823–9826.
- [3] a) B. Gong, J.-H. Chen, E. Chase, D. M. Chadalavada, R. Yajima, B. L. Golden, P. C. Bevilacqua, P. R. Carey, *J. Am. Chem. Soc.* **2007**, *129*, 13335–13342; b) Z. Cai, I. Tinoco, Jr., *Biochemistry* **1996**, *35*, 6026–6036; c) S. E. Butcher, F. H.-T. Allain, J. Feigon, *Biochemistry* **2000**, *39*, 2174–2182; d) T. J. Wilson, N.-S. Li, J. Lu, J. K. Frederiksen, J. A. Piccirilli, D. M. J. Lilley, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11751–11756.
- [4] a) P. V. Cornish, M. Hennig, D. P. Giedroc, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12694–12699; b) J. L. Wilcox, P. C. Bevilacqua, *J. Am. Chem. Soc.* **2013**, *135*, 7390–7393; c) B. Houck-Loomis, M. A. Durney, C. Salguero, N. Shankar, J. M. Nagle, S. P. Goff, V. M. D'Souza, *Nature* **2011**, *480*, 561–564.
- [5] J. L. Wilcox, P. C. Bevilacqua, *Biochemistry* **2013**, *52*, 7470–7476.
- [6] a) T. Hermann, D. J. Patel, *Science* **2000**, *287*, 820–825; b) J. Feigon, T. Dieckmann, F. W. Smith, *Chem. Biol.* **1996**, *3*, 611–617.
- [7] a) J. H. Davis, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11616–11621; b) J. M. Carothers, S. C. Oestreich, J. H. Davis, J. W. Szostak, *J. Am. Chem. Soc.* **2004**, *126*, 5130–5137.
- [8] a) J. M. Carothers, J. H. Davis, J. J. Chou, J. W. Szostak, *RNA* **2006**, *12*, 567–579; b) A. H. Nasiri, J. P. Wurm, C. Immer, A. K. Weickmann, J. Wöhnert, *RNA* **2016**, *22*, 1750.
- [9] E. Duchardt-Ferner, J. Ferner, J. Wöhnert, *Angew. Chem. Int. Ed.* **2011**, *50*, 7927–7930; *Angew. Chem.* **2011**, *123*, 8073–8076.
- [10] a) P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1997**, *119*, 6621–6628; b) P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1994**, *116*, 8390–8391.
- [11] D. Xie, S. Gulnik, L. Collins, E. Gustchina, L. Suvorov, J. W. Erickson, *Biochemistry* **1997**, *36*, 16166–16172.
- [12] F. C. Kokesh, F. H. Westheimer, *J. Am. Chem. Soc.* **1971**, *93*, 7270–7274.
- [13] S. R. Gottstein-Schmidtke, E. Duchardt-Ferner, F. Groher, J. E. Weigand, D. Gottstein, B. Suess, J. Wöhnert, *RNA* **2014**, *20*, 1163–1172.
- [14] A. C. Wolter, E. Duchardt-Ferner, A. H. Nasiri, K. Hantke, C. H. Wunderlich, C. Kreutz, J. Wöhnert, *Biomol. NMR Assignments* **2016**, *10*, 101–105.

Manuscript received: September 19, 2016

Revised: November 1, 2016

Final Article published: ■■■■■■, ■■■■■■

2. KELLER ET AL - ADENINE PROTONATION FOR LIGAND PROMISCUITY

Title: Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches

Authors: Keller H*, Weickhmann AK*, Bock T, Wöhnert J *equal contribution

Published in: RNA 24:1390-1402;

Contributions

(1) Entwicklung und Planung	Keller H (30%), Weickhmann AK (30%), Wöhnert J (40%)
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Keller H: Klonierung, RNA-Präparation, NMR-Messungen; Weickhmann AK: RNA-Präparation, ITC-Messungen, Bock T: Klonierung, RNA-Präparation
(3) Erstellung der Datensammlung und Abbildungen	Keller H: Erstellung von Abbildungen, Weickhmann AK: Erstellung von Abbildungen
(4) Analyse/ Interpretation der Daten	Keller H: Auswertung der NMR-Messungen; Weickhmann AK: Auswertung der ITC-Messungen; Wöhnert J: Auswertung der NMR-Messungen
(5) Übergeordnete Einleitung/ Ergebnisse/Diskussion	Keller H (40%), Weickhmann AK (20%), Wöhnert J (40%)

Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches

HEIKO KELLER,¹ A. KATHARINA WEICKHMANN,¹ THOMAS BOCK, and JENS WÖHNERT

Institute for Molecular Biosciences and Center of Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

ABSTRACT

In certain structural or functional contexts, RNA structures can contain protonated nucleotides. However, a direct role for stably protonated nucleotides in ligand binding and ligand recognition has not yet been demonstrated unambiguously. Previous X-ray structures of *c*-GAMP binding riboswitch aptamer domains in complex with their near-cognate ligand *c*-di-GMP suggest that an adenine of the riboswitch either forms two hydrogen bonds to a G nucleotide of the ligand in the unusual enol tautomeric form or that the adenine in its N1 protonated form binds the G nucleotide of the ligand in its canonical keto tautomeric state. By using NMR spectroscopy we demonstrate that the *c*-GAMP riboswitches bind *c*-di-GMP using a stably protonated adenine in the ligand binding pocket. Thereby, we provide novel insights into the putative biological functions of protonated nucleotides in RNA, which in this case influence the ligand selectivity in a riboswitch.

Keywords: protonated adenine; riboswitch; *c*-di-GMP; *c*-GAMP; NMR; hydrogen bonds

INTRODUCTION

The pK_a values of the nucleotide building blocks in RNA and DNA are far away from neutrality (G, U \sim 9.2, A \sim 3.9, C \sim 4.2) in unstructured nucleic acids (Saenger 1988). Watson–Crick base-pairing in double helical structural elements shifts these pK_a values even further away from neutrality (Saenger 1988; Thaplyal and Bevilacqua 2014). Therefore, nucleotides in RNAs mostly adopt their canonical neutral protonation states and tautomeric forms. However, many RNAs incorporate non-Watson–Crick base pairs in their structures or adopt intricate tertiary structures. These structural environments can induce significant pK_a shifts of RNA nucleotides toward neutrality leading to the occasional presence of nucleotides with altered protonation states in RNA structures. Very simple examples are A:C mismatches embedded in regular A-form double helical structural elements (e.g., Puglisi et al. 1990; Cai and Tinoco 1996; Huppler et al. 2002; Pechlaner et al. 2015). Adenine N1 protonation stabilizes these mismatches since two hydrogen bonds can be formed between A⁺ and C resulting in a base pair with the same geometry as the classical G:U wobble pair. The pK_a for the A in such base pairs is shifted by \sim 3 units and

can be as high as 8.2 in a particularly stable helical context (Wilcox and Bevilacqua 2013a,b). Similarly, N1 protonated A's are able to stabilize G:A mismatches in Watson–Crick helical contexts (Pan et al. 1999) as well as the parallel double helices formed by poly(rA) (Gleghorn et al. 2016). Protonated C's occur for instance in base triples and triple helices. Both protonated C's and A's have been observed in frame-shifting pseudoknots (e.g., Cornish et al. 2005; Houck-Loomis et al. 2011; Wilcox and Bevilacqua 2013a) and pK_a values between 6.2 and 8.2 have been reported for these nucleotides. At physiological pH values these pseudoknots therefore exist as conformational ensembles containing protonated and deprotonated species. Thus, A or C protonation might play a role in tuning conformational equilibria and thereby frameshifting efficiency. Of particular functional importance is the occurrence of nucleotides with shifted pK_a values in the active site of ribozymes. The hammerhead, the hairpin, the VS, the twister and the pistol ribozyme all are reported to have a G with a pK_a shifted toward neutrality in the active site (Wilcox et al. 2011; Kath-Schorr et al. 2012; Liu et al. 2014; Ren et al. 2014, 2016). There, it supposedly acts as the general base and activates

¹These authors contributed equally to this work and should be regarded as joint first authors.

Corresponding author: woehnert@bio.uni-frankfurt.de

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.067470.118>.

© 2018 Keller et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

the attacking 2'-OH group in the first step of the phosphodiester cleavage reaction. For both the hairpin and the VS ribozyme an N1 protonated A supposedly is the general acid in the cleavage reaction (Rupert et al. 2002; Wilson et al. 2010; Kath-Schorr et al. 2012). For the twister and the pistol ribozyme an N3 protonated A is suggested to play the same functional role (Ren et al. 2016; Wilson et al. 2016). A protonated C is used as the general acid in the HDV ribozyme (Gong et al. 2007). However, in all cases described so far, the shifted pK_a is near neutrality (7.0), suggesting that at physiologically relevant pH values these RNAs exist in an equilibrium with significant populations of protonated and deprotonated species. These equilibria are most likely functionally important for switching between active and inactive conformations and/or by allowing the catalytic base of a ribozyme to function as the catalytic acid in the reverse reaction and vice versa. Recently, examples for nucleotides in stable RNA tertiary

structures with very strongly shifted pK_a values (>5 pH units) were reported where the protonation enables these nucleotides to function as stabilizing building blocks of such structural elements (Wilcox and Bevilacqua 2013b; Gottstein-Schmidtke et al. 2014; Wolter et al. 2017). What has not been reported so far at least to our knowledge, however, is an example where a protonated nucleotide plays a direct role in ligand binding and recognition by a functional RNA.

However, a putative example for such an RNA with a protonated nucleotide directly binding to a ligand is represented by the riboswitches binding to the cyclic dinucleotide 3',3'-cyclic-GMP-AMP (*c*-GAMP, Fig. 1A). *c*-GAMP has been discovered recently as a regulatory cyclic dinucleotide in bacteria and not much is known about its associated signaling pathways (Davies et al. 2012). In *Geobacter* species it regulates the expression of numerous genes involved in exoelectrogenesis by binding to *c*-GAMP

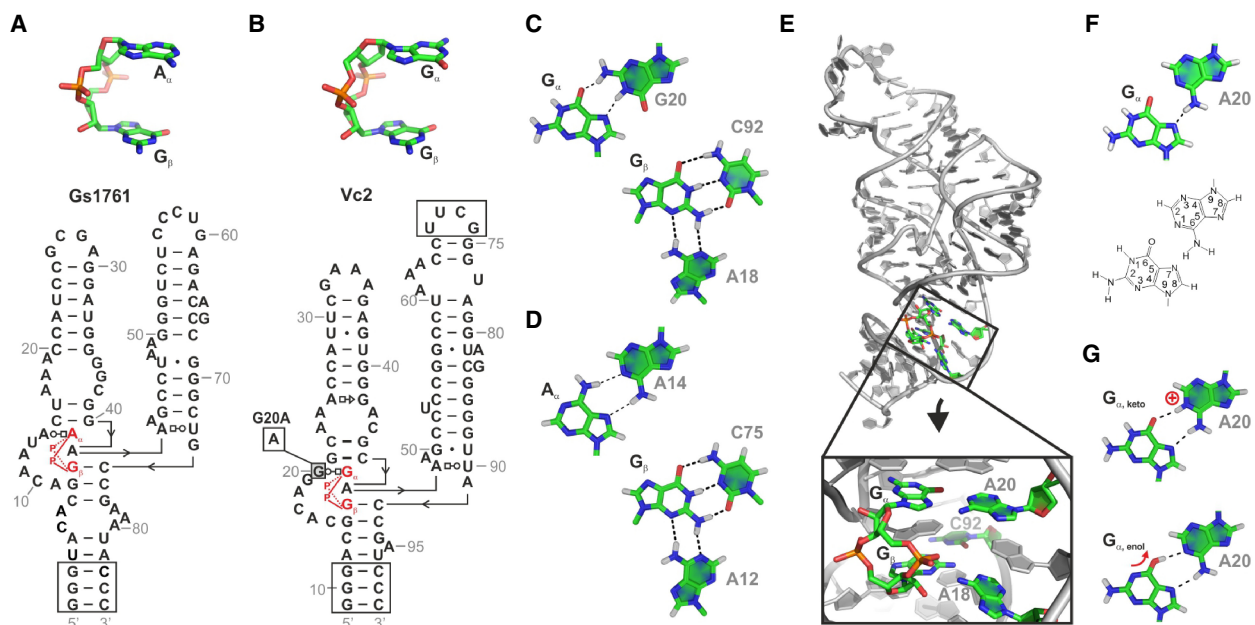


FIGURE 1. Ligand recognition by *c*-GAMP and *c*-di-GMP binding riboswitches. (A) Structure of 3',3'-cyclic-GMP-AMP (*c*-GAMP, top) and secondary structure of a *c*-GAMP-binding riboswitch (bottom) from *Geobacter sulfurreducens* (Gs 1761, Ren et al. 2015). Nucleotides added to the terminus of the P1 stem in order to increase transcription efficiency and P1 stability in the NMR construct used in this study are boxed. The ligand is colored red. (B) Structure of *c*-di-GMP (top) and a *c*-di-GMP riboswitch from *Vibrio cholerae* (Vc2, Smith et al. 2009). Nucleotides added to the terminus of the P1 stem in order to increase transcription efficiency and an artificial stable UUCG tetraloop closing P3 in the NMR construct are boxed. The ligand is colored red. The position of the point mutation G20A that renders this riboswitch bispecific for *c*-GAMP and *c*-di-GMP is indicated by a shaded box. (C) Intermolecular base-pairing between the two ligand nucleotides G_{α} and G_{β} and nucleotides of the *c*-di-GMP riboswitch as seen in the X-ray structure of the Vc2-RNA/*c*-di-GMP complex (PDB ID 3irw, Smith et al. 2009). (D) Intermolecular base-pairing between the two ligand nucleotides A_{α} and G_{β} and nucleotides of the *c*-GAMP riboswitch as seen in the X-ray structure of the Gs1761-RNA/*c*-GAMP complex (PDB ID 4yaz, Ren et al. 2015). A14 here corresponds to G20 in the Vc2-riboswitch. (E) X-ray structure of the G20A-mutant of the Vc2 *c*-di-GMP-binding riboswitch bound to *c*-di-GMP (PDB ID 3mum, Smith et al. 2010). The ligand and riboswitch nucleotides involved directly in *c*-di-GMP binding are colored by atom type. The ligand-binding mode is shown as a close-up view. Nucleotides that recognize the ligand and the ligand itself are highlighted and shown as stick model. (F) Interaction between G_{α} of the ligand and nucleotide A20 as seen in E. In order to facilitate the description of possible hydrogen bonding interaction in this base pair the atom numbering for both the G_{α} and the A20 is shown below. (G) Two possible hydrogen-bonding patterns allowing the formation of two hydrogen bonds between G_{α} and A20. A20 could be protonated at N1 (top) with G_{α} adopting the standard keto tautomeric form or G_{α} could occur as the 6-enol tautomer with A20 in the neutral nonprotonated state (bottom).

riboswitches (Kellenberger et al. 2015; Nelson et al. 2015). These riboswitches are surprisingly similar in terms of sequence, secondary and tertiary structure to a previously described riboswitch class binding to cyclic-di-GMP (Fig. 1B) called GEMM-I (Sudarsan et al. 2008). The c-GAMP binding riboswitches are therefore designated as members of the GEMM-Ib riboswitch class. A remarkable feature of the GEMM-I riboswitch aptamer domains is their generally high affinity for their cognate ligand c-di-GMP with K_D values sometimes in the picomolar range (Sudarsan et al. 2008; Smith et al. 2009). The high affinity of these riboswitches is most likely the consequence of the rather low intracellular concentrations of c-di-GMP found to be in the high nanomolar to low micromolar range depending on the organism and growth conditions (Kader et al. 2006; Simm et al. 2009) and a kinetic control of their activity (Wickiser et al. 2005). Notably, at these low concentrations c-di-GMP does not yet form kinetically stable G-quadruplex-like and other oligomeric structures that would compete with its signaling functions (Gentner et al. 2012) despite the presence of high intracellular concentrations of potassium ions. The intracellular concentrations of c-GAMP have not yet been systematically quantified in different organisms and growth conditions. However, in *Geobacter sulfurreducens* c-GAMP can reach concentration levels comparable to those of c-di-GMP under certain conditions (Kellenberger et al. 2015).

GEMM-I riboswitches bind their ligand c-di-GMP in a bipartite binding site (Kulshina et al. 2009; Smith et al. 2009, 2010). The two guanine bases (designated G_α and G_β) of the ligand are recognized differently (Fig. 1C). The Hoogsteen-edge of G_α is recognized by the Watson-Crick edge of G20 from the riboswitch (nucleotide numbering in the RNA corresponds to the *Vibrio cholera* Vc2 sequence presented in Smith et al. 2009). Two hydrogen bonds are formed between the C6 carbonyl group of G_α and the G20 amino group as well as between the N7 nitrogen of G_α and the G20 imino group. G_β forms a standard Watson-Crick base pair with C92 of the riboswitch. In the c-GAMP-binding riboswitches G20 is replaced by an adenine nucleotide (A14 in the *Geobacter sulfurreducens* Gs1761 riboswitch, Fig. 1A). A_α of c-GAMP now binds to A14 of the GEMM-Ib riboswitch through a base-pairing interaction isosteric to the G_α :G base pair in the GEMM-I riboswitch (Fig. 1D, Ren et al. 2015, PDB ID 4yaz). The A_α amino group is hydrogen bonded to the N1 nitrogen of A14 and the A_α N7 nitrogen is hydrogen bonded to the A14 amino group (Fig. 1D). Surprisingly, the G20A mutation of the Vc2 GEMM-I riboswitch yields an RNA capable of binding to both c-di-GMP and to c-GAMP whereas the WT discriminates against c-GAMP (Smith et al. 2010; Kellenberger et al. 2013). Furthermore, a number of naturally occurring c-GAMP binding GEMM-Ib riboswitches are apparently bispecific for c-di-GMP and c-GAMP (Kellenberger et al. 2015; Nelson et al. 2015). X-ray struc-

tures of the Vc2 G20A mutant (Smith et al. 2010, PDB ID 3mum) and the Gs1761 c-GAMP riboswitch (Ren et al. 2015, PDB ID 4yb0) bound to c-di-GMP revealed a ligand recognition mode (Fig. 1E) with a hydrogen bond between the N7 nitrogen of G_α and the A20 amino group (Fig. 1F) similar to what was observed for c-GAMP binding (Fig. 1D). Furthermore, in both of these structures there is a short distance (2.8 Å in pdb 3mum, Smith et al. 2010 and 3.0 Å in pdb 4yb0, Ren et al. 2015) between the C6 carbonyl group of G_α and the N1 nitrogen of A20 (Fig. 1F). When assuming standard protonation patterns for the nucleotides, both of these functional groups are hydrogen bond acceptors carrying a negative partial charge. Therefore, the observed close contact should be energetically unfavorable. However, if as suggested previously (Smith et al. 2010; Ren et al. 2015) A20 is protonated at the N1 nitrogen, it could form a stabilizing hydrogen bonding interaction with the C6 carbonyl group of G_α (Fig. 1G, top). Alternatively, it was suggested (Ren et al. 2015) that G_α could adopt the enol tautomeric form with a hydroxyl group at C6. Then the C6 hydroxyl group could form a hydrogen bond to the N1 nitrogen of A20 (Fig. 1G, bottom). Both possibilities would be equally exciting because neither the direct participation of a protonated nucleotide in ligand binding in an RNA nor the stable induction of a rare tautomeric state for a standard nucleotide in an RNA structure have been demonstrated unequivocally. However, unusual tautomeric states in G:U mismatches are adopted transiently in low populations, as recently demonstrated by NMR spectroscopy (Kimsey et al. 2015). In addition, X-ray structures of near-cognate tRNAs bound to the decoding center of the ribosome and of mismatch-containing duplexes suggested that they play a role in miscoding (Demeshkina et al. 2012; Rozov et al. 2015, 2016; Rypniewski et al. 2016). However, X-ray structures with a resolution >1 Å do not allow the direct observation of hydrogen positions and thereby the unambiguous assignment of hydrogen bonding patterns in cases where the positions of the heavy atoms allow alternative hydrogen bonding patterns. For the complexes of c-di-GMP bound to Vc2 G20A and Gs1761 the resolution is 2.9 Å (pdb 3mum) and 2.1 Å (pdb 4yb0), respectively. On the other hand, solution NMR spectroscopy is a very useful tool for the direct unambiguous elucidation of hydrogen bonding patterns in nucleic acids (Dingley and Grzesiek 1998; Wöhnert et al. 1999; Duchardt-Ferner et al. 2011; Duchardt-Ferner and Wöhnert 2017). In particular, ^{13}C and ^{15}N chemical shifts are faithful reporters of changes in protonation states (Legault and Pardi 1994, 1997). In favorable cases, NMR signals for protons at the protonation site can be detected directly (e.g., Macaya et al. 1991; Brodsky et al. 1998; Nixon et al. 2002; Cash et al. 2013; Gottstein-Schmidtke et al. 2014; Wolter et al. 2017). Some of the NMR methods for the identification of protonation events are applicable even for larger RNAs in cases where initial structural information from other methods is

already available as in the case of the c-GAMP riboswitches and their complexes with c-di-GMP and c-GAMP.

Here we set out to delineate the hydrogen bonding patterns between c-di-GMP and the c-GAMP riboswitches by NMR. We find that an N1 protonated adenine nucleotide of the riboswitch forms a hydrogen bond to the guanine base (G_{ω}) of the ligand in its standard neutral imino tautomeric state. The adenine N1 protonation is induced by c-di-GMP binding. The protonated state of the riboswitch in the c-di-GMP complex is persistent even at pH values above 8.3 and most likely limits the specificity of the c-GAMP-binding riboswitches by allowing stable binding of c-di-GMP.

RESULTS AND DISCUSSION

The GEMM riboswitch variants with sizes of >80 nucleotides and their complex tertiary structures are challenging for NMR studies. In particular, NMR signal assignment strategies using the standard heteronuclear, multidimensional NMR approaches are no longer applicable for RNAs of this size with highly complex tertiary structures that are not amenable to “divide and conquer” approaches. Thus, our NMR studies had to rely on the available structural information from X-ray structures, the comparison of the NMR properties of different variants of the same riboswitch and chemical shift comparison in conjunction with base-type selective labeling. A particularly well-suited starting point for our investigations was therefore the G20A-mutant of the c-di-GMP binding GEMM-I riboswitch from *V. cholerae* (Vc2, Fig. 1B). There are X-ray structures available for the WT-Vc2 riboswitch bound to c-di-GMP (PDB ID 3mxx, 2.3 Å) as well as for the G20A-mutant bound to c-di-GMP (PDB ID 3mum, 2.9 Å) and c-GAMP (PDB ID 4yb1, 2.1 Å) revealing very similar overall structures (Smith et al. 2010; Ren et al. 2015). Therefore, the WT Vc2 riboswitch bound to c-di-GMP and the G20A-mutant bound to c-GAMP serve as reference states. Since they only differ by a single nucleotide from the G20A-mutant bound to c-di-GMP, they allow a meaningful comparison of their NMR spectra.

We initially characterized c-di-GMP and c-GAMP binding to the Vc2 WT riboswitch and the G20A mutant by 1D imino proton NMR spectra and ITC (Fig. 2). It should be noted that all NMR and ITC measurements reported in this study were carried out in buffers (25 or 50 mM BisTris pH 6.5, 5 mM magnesium acetate if not explicitly noted otherwise) completely lacking potassium ions. ^1H - and ^{31}P -NMR experiments showed that in this buffer c-di-GMP does not form G-quadruplex or oligomeric structures (Supplemental Fig. S1) at the c-di-GMP concentrations needed for ITC or NMR experiments. In contrast, the presence of 25 mM potassium phosphate or 25 mM potassium phosphate and 250 mM potassium chloride induced large changes in the NMR-spectra of free c-di-GMP

(Supplemental Fig. S1). These changes indicate the formation of G-quadruplex and other oligomeric structures at the c-di-GMP concentrations required for the NMR and ITC experiments (>100 μM) in agreement with previous reports (Gentner et al. 2012). The formation of these structures would directly compete with c-di-GMP binding to RNA and thereby masking the effects of c-di-GMP binding in our experiments. Under in vivo conditions, however, these oligomeric c-di-GMP structures are apparently not relevant since the measured intracellular c-di-GMP concentrations are well below 100 μM (Kader et al. 2006; Simm et al. 2009). At such low concentrations of c-di-GMP no quadruplex formation and/or oligomerization was observed even in the presence of potassium ions (Gentner et al. 2012).

Both the Vc2 WT and the G20A RNA showed significant changes in their imino proton NMR spectra upon ligand addition in the presence of 5 mM Mg^{2+} . The appearance of novel imino proton signals upon ligand addition suggested ligand-induced folding of the RNA in all cases. Importantly, the spectral changes observed upon addition of c-di-GMP or c-GAMP to the G20A-mutant are similar to the spectral changes observed for the WT-RNA upon addition of c-di-GMP (Fig. 2A) in agreement with a very similar ligand binding mode in all three cases. In contrast, the spectra of the WT-RNA in the presence of c-GAMP show a reduced number of imino proton signals compared to the other complexes in line with incomplete folding of the RNA upon binding c-GAMP. The WT-RNA binds c-di-GMP with a picomolar K_D (Sudarsan et al. 2008; Smith et al. 2009, 2010). ITC measurements with c-GAMP showed that the affinity of the WT-RNA for this ligand is much lower, revealing a K_D of 10 μM (Fig. 2B). In contrast, the G20A-mutant binds tightly to both c-di-GMP and c-GAMP with K_D values of 53 and 120 nM, respectively (Fig. 2B; Supplemental Table 1). Thus, as expected based on previous results, the G20A-mutant is bispecific and binds both ligands with similar affinity and a similar overall binding mode.

A very sensitive and well-established reporter for a putative protonation of adenines is their C2 chemical shift (Legault and Pardi 1994, 1997). The C2 chemical shifts for adenines are normally found in a range between 152 and 157 ppm. Upon protonation the C2 chemical shifts are observed ~8–10 ppm upfield at 146–147 ppm. Thus, if c-di-GMP binding induces A20 protonation in the Vc2-G20A mutant, an H2C2 signal with an upfield C2 chemical shift should be found in a ^{13}C -HSQC spectrum that would be absent in the spectra of both the c-di-GMP bound WT-RNA as well as the c-GAMP bound G20A-mutant RNA. In order to reduce spectral crowding ^{13}C -HSQC-spectra were recorded for ^{13}C , ^{15}N -adenine labeled RNAs and unlabeled ligand. Thus, all other nucleotides of the RNA and the ligand are spectroscopically silent. A comparison of the ^{13}C -HSQC-spectra of ^{13}C -adenine labeled G20A

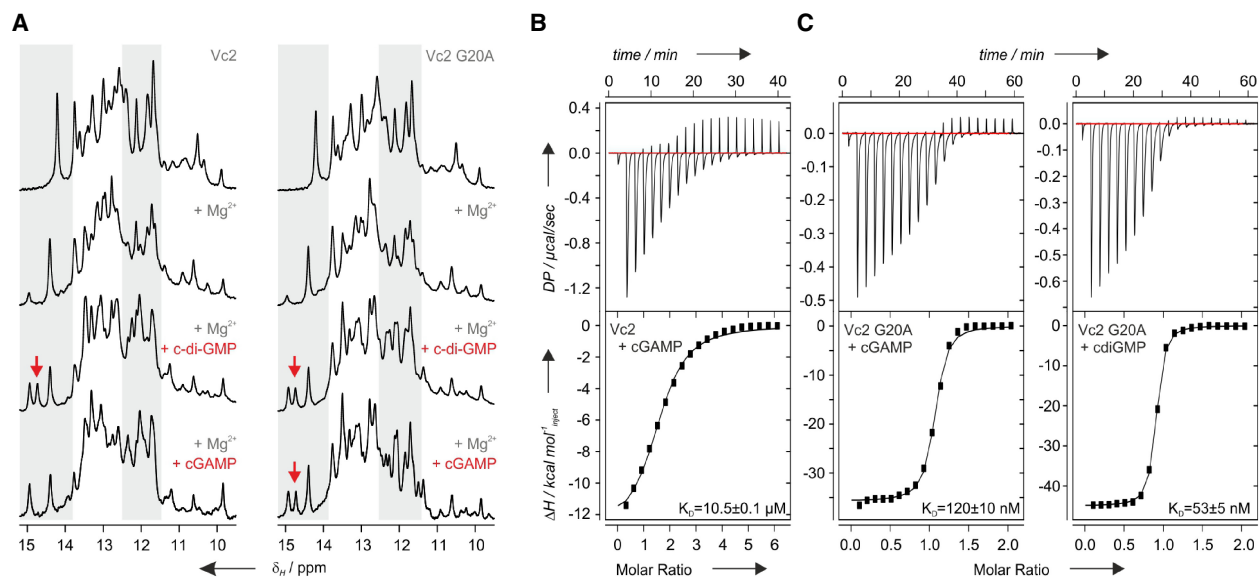


FIGURE 2. Ligand binding to the Vc2-WT and the Vc2-G20A-mutant riboswitch. (A) Comparison of the imino proton spectra of the Vc2-WT (left) and the Vc2-G20A-mutant (right) riboswitch RNA in the ligand-free state (top), in the presence of 5 mM Mg^{2+} (middle), in the presence of 5 mM Mg^{2+} and c-di-GMP (middle) and in the presence of 5 mM Mg^{2+} and c-GAMP (bottom). Shaded areas highlight chemical shift regions with pronounced spectral changes. Note that the imino proton spectra of the G20A-mutant in the presence of c-di-GMP and c-GAMP are very similar to each other and to the spectra of the WT bound to c-di-GMP. In contrast, in the spectrum of the WT-riboswitch bound to c-GAMP imino proton signals are missing compared to those of the c-di-GMP complex and the mutant (red arrows) suggestive of incomplete folding. (B) ITC thermogram and fit for c-GAMP binding to the WT Vc2 riboswitch showing a K_D of only 10 μ M. c-di-GMP is bound with a K_D in the picomolar range (Smith et al. 2009, 2010). (C) ITC thermograms and fits for c-GAMP (left) and c-di-GMP (right) binding to the G20A-mutant of the Vc2 riboswitch. C-GAMP is bound with a K_D of 120 nM while c-di-GMP is bound with a K_D of 53 nM. All ITC-experiments shown here were carried out in a buffer containing 50 mM Bis-Tris pH 6.5 and 5 mM magnesium acetate at 25°C.

RNA in its free form and bound to c-di-GMP (Fig. 3A) reveals massive spectral differences. In particular, signal dispersion in the presence of c-di-GMP increases significantly in line with the expected ligand-induced folding of the RNA. Importantly, there is one signal present in the spectrum of the G20A-RNA/c-di-GMP complex with a chemical shift of 147.4 ppm in the ^{13}C -dimension and 8.3 ppm in the 1H -dimension, respectively, which is not present in the free RNA and is a candidate for the H2C2 signal of a protonated adenine. In the ^{13}C -HSQC-spectrum of the WT RNA containing a G at position 20 bound to c-di-GMP this signal is absent (Fig. 3B). This signal is also absent in the ^{13}C -HSQC spectrum of the G20A mutant RNA bound to c-GAMP where the nonprotonated nucleotide A20 forms hydrogen bonds to the Hoogsteen-edge of A_α of the ligand (Fig. 3C). A second signal that is only found in the ^{13}C -HSQC spectrum of the G20A-RNA/c-di-GMP complex but not in those of the other two complexes appears at a chemical shift of 7.5 ppm (1H) and \sim 142 ppm (^{13}C). Comparison with spectra of other RNAs containing protonated adenines (Legault and Pardi 1997; Wolter et al. 2017) suggests that this signal might correspond to the C8 of a protonated adenine. Thus, the comparison of the ^{13}C -HSQC spectra for the G20A-RNA/c-di-GMP complex with those of the WT-RNA bound to the same ligand and the G20A-RNA bound to c-GAMP shows that the former

contains a protonated adenine not occurring in the latter two complexes. Due to the sequential and structural similarities of the three systems, A20 is the only logical candidate for the protonated adenine. Importantly, the signal corresponding to the protonated A20 in the G20A-RNA/c-di-GMP complex is still observable with similar intensity and at the same position in ^{13}C -HSQC spectra recorded at pH 8.3, suggesting a stable protonation even at elevated pH (Fig. 3D). Unfortunately, the absence of chemical shift, line widths or intensity changes for the A20 H2C2 resonance upon increasing the pH from 6.5 to 8.3 prevented us from determining the pK_a for A20 protonation in the G20A-RNA/c-di-GMP complex by pH titrations. We avoided increasing pH values above pH 8.3 since under these conditions G and U nucleotides of the RNA, which are not in Watson-Crick base pairs, would start to become deprotonated to a significant amount ($pK_a \sim 9.2$). Since such nucleotides are found in the direct vicinity of the binding site (Smith et al. 2009; Ren et al. 2015), chemical shift changes potentially observable at higher pH values could no longer be attributed to A20 deprotonation alone. Thus, we can only conclude that the pK_a for A20 deprotonation in the complex must be significantly larger than 8.3.

The imino protons of protonated adenines are normally not directly observable at physiologically relevant pH due to fast exchange with the bulk solvent and/or rapid

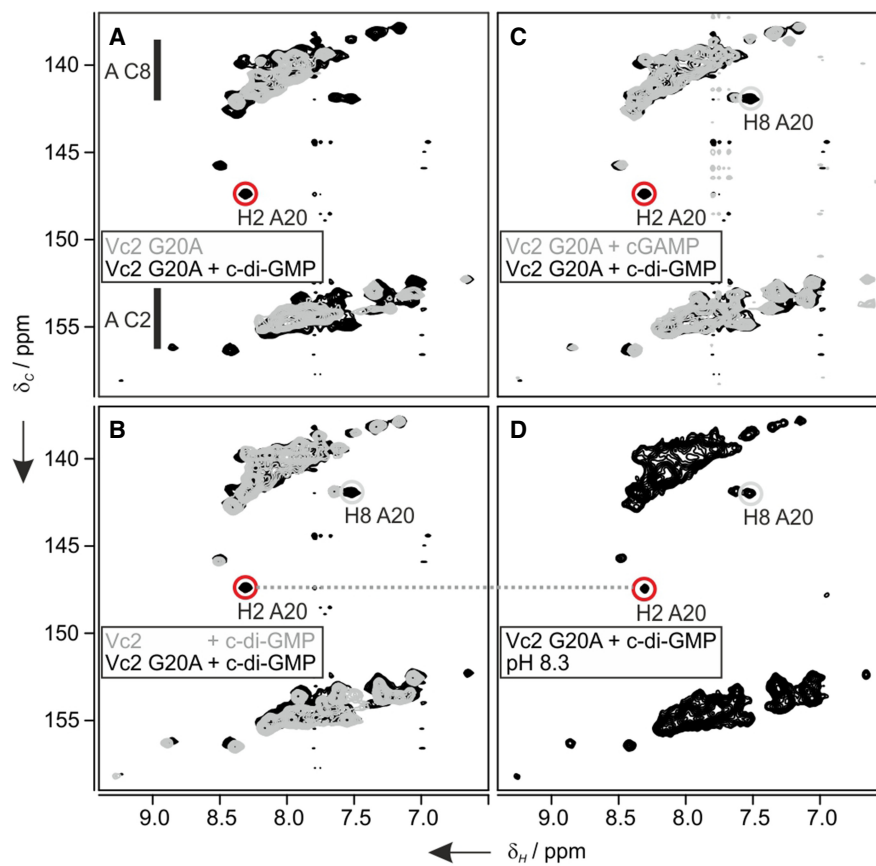


FIGURE 3. Evidence for a protonated A20 in the complex of the G20A-mutant Vc2 riboswitch bound to *c*-di-GMP. (A) Overlay of ^{13}C -HSQC spectra of ^{13}C , ^{15}N -adenine labeled ligand-free G20A RNA (gray) and G20A RNA bound to *c*-di-GMP (black). Large spectral changes are observed as expected for ligand-induced RNA-folding. Importantly, a signal is observed with a ^{13}C chemical shift of 147.4 ppm and a ^1H chemical shift of 8.3 ppm, respectively (red circle)—the chemical shift range associated with C2 carbon nuclei of protonated adenine nucleotides. The typical chemical shift ranges for adenine carbon nuclei are indicated by bars on the left side of the spectrum. (B) Overlay of the ^{13}C -HSQC-spectra of the G20A-mutant (black) and the WT-RNA (gray) both bound to *c*-di-GMP. The H2C2 signal indicative of adenine protonation occurs only in the G20A-mutant. (C) Overlay of the ^{13}C -HSQC-spectra of the G20A-mutant RNA bound to *c*-di-GMP (black) or *c*-GAMP (gray). The H2C2 signal indicative of adenine protonation occurs only in the *c*-di-GMP complex. (D) ^{13}C -HSQC spectrum recorded for the G20A-mutant bound to *c*-di-GMP recorded at pH 8.3. The H2C2 signal typical for the protonated adenine unique to this complex is still present at high intensity (red circle) and has the same ^1H and ^{13}C chemical shifts as in the spectrum recorded at pH 6.5.

opening of base-pairing interactions. The only instances of NMR observable adenine imino protons in nucleic acids to our knowledge were reported under conditions of either low pH or for RNAs where the protonated nucleotide was an integral part of the tertiary structure (Macaya et al. 1991; Wolter et al. 2017). A ^{15}N -HSQC-spectrum recorded at 10°C with ^{15}N -adenine labeled G20A-mutant RNA bound to *c*-di-GMP showed a single adenine imino group resonance with chemical shifts of ~ 13.7 (^1H) and ~ 152.1 (^{15}N) ppm, respectively (Fig. 4A). This imino resonance corresponds to the imino group of the protonated A20 that serves as a hydrogen bond donor to the C6 carbonyl group of the G_α of the bound *c*-di-GMP. As expected, it is not observable in the G20A-RNA/*c*-GAMP complex (Fig. 4B) since in the *c*-GAMP complex the A20 adenine N1 nitrogen is not protonated and serves as a hydrogen bond acceptor for the A_α amino group of the

c-GAMP ligand (Fig. 1D). In a two-bond ^{15}N -HSQC spectrum the H2-proton at 8.3 ppm corresponding to the protonated A20 is connected to its N3 and the N1 nitrogen chemical shifts (Fig. 4C). The N1 chemical shift connected to the H2 proton of A20 is the same as the one observed in the standard ^{15}N -HSQC and ~ 60 ppm upfield from the N1 nitrogen chemical shifts of the unprotonated adenine nucleotides as expected. Furthermore, temperature-dependent ^{15}N -HSQC-spectra show that the imino group signal is observable at temperatures up to 25°C (Fig. 4D). Thus, the protonation of A20 at N1 in the complex with *c*-di-GMP is very stable.

In a recent example of an RNA containing a $G:A^+$ interaction with a geometry resembling the one of the $G_\alpha:A20$ interaction, hydrogen bonding between the G C6 carbonyl group and the N1 imino group of the protonated A lead to a pronounced downfield shift of the guanine C6 carbon

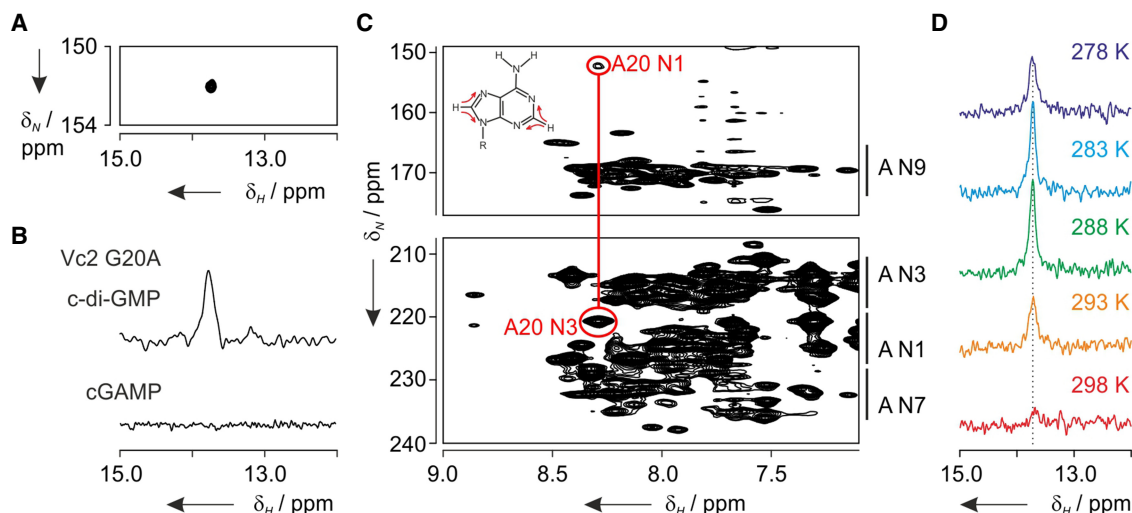


FIGURE 4. A20 is stably protonated at the N1 nitrogen in the G20A-mutant riboswitch in complex with c-di-GMP. (A) A 2D- ^{15}N -HSQC spectrum recorded at 10°C for a sample containing ^{15}N -adenine labeled G20A-RNA bound to unlabeled c-di-GMP shows a single imino group signal. (B) A comparison of 1D- ^{15}N -HSQC spectra at 10°C recorded with 256 scans in the c-di-GMP complex and 1024 scans in the c-GAMP complex show that while this imino proton signal is present in the c-di-GMP complex it is absent in the c-GAMP complex of the G20A mutant RNA. (C) A two-bond 2D- ^{15}N -HSQC spectrum links the H2 proton of A20 at 8.3 ppm to the N1 and N3 nitrogens of the same nucleotide. An inset shows the magnetization transfer pathway for this experiment. The A20 N1 nitrogen chemical shift of 152.3 ppm is ~ 60 ppm upfield compared to the N1 chemical shifts for unprotonated adenines and is the same as the one observed for the imino group nitrogen in the standard ^{15}N -HSQC experiment. Vertical bars on the right side of the spectrum indicate the typical ^{15}N chemical shift ranges for different types of nitrogen nuclei in adenine nucleotides. (D) The imino proton signal of A20 $^+$ is observable up to 25°C in 1D- ^{15}N -HSQC-spectra at pH 6.5 suggesting a stable protonation.

compared to all other guanine C6 resonances (Wolter et al. 2017). In order to measure the C6 chemical shifts of c-di-GMP bound to either WT-RNA or the G20A-mutant, we enzymatically prepared and purified ^{13}C -labeled c-di-GMP and measured ^{13}C -1D-spectra of the labeled ligand bound to both RNAs in their unlabeled form (Fig. 5A). The C6 carbon of the c-di-GMP G_{α} bound to the G20A-mutant which interacts with the protonated A20 is shifted downfield compared to the G_{α} C6 when bound to the WT RNA, where it interacts with the neutral G20 (Figs. 5B, 1G,C). On the other hand, the G_{β} C6 chemical shift is very similar in both complexes, in agreement with a similar binding mode for G_{β} in both complexes.

The X-ray structures of the Vc2 and Vc2 G20A-complexes with their ligands revealed that Mg^{2+} -ions played an integral role for RNA ligand interactions. In particular, both phosphate groups of the ligand are coordinated by Mg^{2+} -ions. The closest distance between a bound Mg^{2+} and the A20 protonation site is 7.4 Å. Furthermore, comparison of the NMR-spectra of the free RNA in the absence and the presence of Mg^{2+} revealed differences in the imino proton region that are suggestive of a certain degree of Mg^{2+} -induced structural preorganization of the RNA. In other RNAs containing protonated nucleotides, it was demonstrated that protonation and Mg^{2+} -binding are anti-cooperative (Huppler et al. 2002; Wilcox and Bevilacqua 2013b). In contrast, in the HDV ribozyme the simultaneous protonation of C41 in a base triple and the

binding of a structural Mg^{2+} ion cooperatively promote ribozyme folding and cleavage activity (Nakano and Bevilacqua 2007). In order to further characterize the interplay between A20 protonation and Mg^{2+} -binding in the Vc2 G20A riboswitch/c-di-GMP complex, we compared ^1H -NMR spectra and measured the ligand affinity at pH 6.5 and 8.3 at different Mg^{2+} concentrations (Supplemental Fig. S2; Supplemental Table 2). At both pH values no ligand binding and no ligand-induced RNA folding was observable when the Mg^{2+} concentration was below ~ 2 mM. At both pH values an increased Mg^{2+} concentration enhanced the affinity of the riboswitch for c-di-GMP. However, the c-di-GMP affinity is always higher at pH 6.5 than at pH 8.3. Thus, both a low pH favoring A20 protonation as well as high Mg^{2+} -concentrations favoring Mg^{2+} -binding contribute positively to c-di-GMP binding and RNA-folding in this system.

We next investigated binding of both c-GAMP and c-di-GMP to the Gs1761 riboswitch from *Geobacter sulfurreducens* (Kellenberger et al. 2015; Nelson et al. 2015). This riboswitch was described as being highly selective for c-GAMP. Based on in-line-probing experiments, Hammond and coworkers (Kellenberger et al. 2015) reported K_D s of this riboswitch for c-GAMP and c-di-GMP of 0.53 and 660 nM, respectively. Using ITC, Patel and coworkers (Ren et al. 2015) measured K_D s of 70 nM and 930 nM for c-GAMP and c-di-GMP, respectively, in a buffer containing 50 mM potassium acetate, pH 6.8, 100 mM KCl and 20 mM

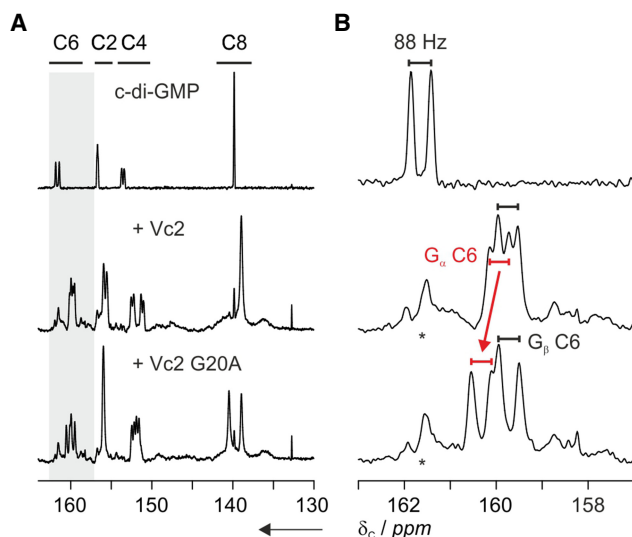


FIGURE 5. Adenine A20 N1 protonation induces a downfield shift of the G_{α} C6 carbon resonance. (A) Comparison of 1D- ^{13}C spectra of ^{13}C , ^{15}N -labeled *c*-di-GMP alone (top) and in complex with Vc2 (middle) or Vc2 G20A mutant RNA (bottom). The bars at top indicate typical chemical shift ranges for guanine carbon nuclei. The spectral region typical for C5 nuclei in guanines (115–120 ppm) is not shown. (B) Zoom of the spectral region that is shaded in A to highlight the effect of the A20 protonation on the ligand C6 chemical shifts. Signals for *c*-di-GMP C6 carbon nuclei appear as doublets due to the C-C scalar coupling with the C5 carbons (88 Hz). Adenine N1 protonation in the Vc2-G20A mutant leads to a downfield shift of the *c*-di-GMP G_{α} C6 chemical shift compared to the Vc2-RNA (red) whereas the *c*-di-GMP G_{β} C6 chemical shift is similar in both complexes (black). Natural abundance ^{13}C -signals from the RNA are highlighted with an asterisk.

MgCl_2 at an elevated temperature of 35°C probably in order to reduce *c*-di-GMP oligomerization. However, Patel and coworkers used an RNA sequence that accidentally differed by two point mutations from the WT-RNA (U72C, C73U) for X-ray structure determination and ITC-experiments (Fig. 1A; Supplemental Fig. S3). We initially used the same RNA sequence as Patel and coworkers with an additionally stabilized P1 stem (named Gs1761 throughout this paper) in our experiments (Fig. 1A). As expected from the earlier findings, our NMR titration experiments using 1D- ^1H -imino proton spectra showed that both ligands bound to the Gs1761 riboswitch in the presence of Mg^{2+} and caused very similar spectral changes (Fig. 6A). Overall, the general quality of the proton spectra for the Gs1761 riboswitch complexes with both ligands is lower compared to the Vc2-derived RNAs. ITC-experiments with our Gs1761 construct under buffer conditions chosen to suppress intermolecular G-quadruplex formation by *c*-di-GMP (50 mM Bis-Tris, pH 6.5, 5 mM Mg^{2+} acetate at 25°C) showed that Gs1761 bound *c*-GAMP only approximately threefold tighter than *c*-di-GMP with K_D values of 290 nM and 710 nM, respectively (Fig. 6B). Restoring the two point mutations to the wild-type se-

quence (Gs1761 WT) did not change the affinities of the riboswitch for *c*-GAMP (236 nM) and *c*-di-GMP (589 nM) significantly (Supplemental Fig. S3; Supplemental Table 1). When Bis-Tris is replaced by 50 mM potassium phosphate in a buffer (pH 6.5) containing 5 mM Mg^{2+} at 25°C the Gs1761 riboswitch binds *c*-GAMP with a K_D of 302 nM (Supplemental Fig. S4), which is very similar to the K_D we measured in Bis-Tris buffer (292 nM). However, no binding of *c*-di-GMP to the RNA is detectable by ITC at 25°C (Supplemental Fig. S4), most likely due to the formation of competing G-quadruplex and oligomeric structures by *c*-di-GMP (Supplemental Fig. S1). Adding 250 mM potassium chloride to this buffer significantly enhances the affinity of the Gs1761 RNA for *c*-GAMP ($K_D = 33$ nM, Supplemental Fig. S4) in agreement with the results of Patel and coworkers (Ren et al. 2015) but binding to *c*-di-GMP is again not detectable by ITC at 25°C (Supplemental Fig. S4).

The comparison of ^{13}C -HSQC spectra for the ^{13}C , ^{15}N -adenine labeled Gs1761 RNA either in its free form or in the presence of *c*-di-GMP showed the appearance of only one signal with the chemical shift characteristics for a H2C2-group of a protonated adenine with chemical shifts of ~ 147.7 ppm (^{13}C) and 8.46 ppm (^1H , Fig. 6C). This signal is absent when the Gs1761 RNA is titrated with its cognate ligand *c*-GAMP (Fig. 6D). Thus, the appearance of this signal in the *c*-di-GMP complex therefore reports on the formation of the base-pairing interaction between G_{α} of the ligand and the protonated A14 (equivalent to A20 in the Vc2 G20A-mutant RNA) of the riboswitch.

Finally, we tested a *c*-GAMP riboswitch aptamer domain from *Clostridium beijerinckii* (Cbe 1–2) (Fig. 7A) for *c*-GAMP and *c*-di-GMP binding that was described as being bispecific for both ligands (Nelson et al. 2015). In agreement with the previous observations for the other *c*-GAMP binding riboswitch variants, both *c*-GAMP and *c*-di-GMP induced significant changes in the imino proton spectra of the RNA indicative of stable binding (Fig. 7B). ITC-experiments show that this RNA binds *c*-di-GMP with a lower K_D (30 nM) than *c*-GAMP (463 nM) in 50 mM Bis-Tris buffer, pH 6.5, 5 mM Mg^{2+} acetate at 25°C (Fig. 7C). The comparison of ^{13}C -HSQC spectra for the ^{13}C , ^{15}N -adenine labeled Cbe 1–2 RNA either in its free form or in the presence of *c*-di-GMP again showed the appearance of one signal with the chemical shift characteristics for a H2C2-group of a protonated adenine (Fig. 7D). In analogy to the observations with the other riboswitches, this signal is absent when the Cbe 1–2 RNA is titrated with *c*-GAMP (Fig. 7E). Thus, Cbe 1–2 binds *c*-di-GMP via a protonated adenine residue as well. Protonation of the adenine nucleotide can therefore be regarded as a common feature for *c*-di-GMP binding in artificial riboswitch variants (Vc2 G20A), naturally occurring *c*-GAMP riboswitches with reportedly high specificity (Gs 1761) and those previously reported to be bispecific (Cbe 1–2). Thus, in these systems

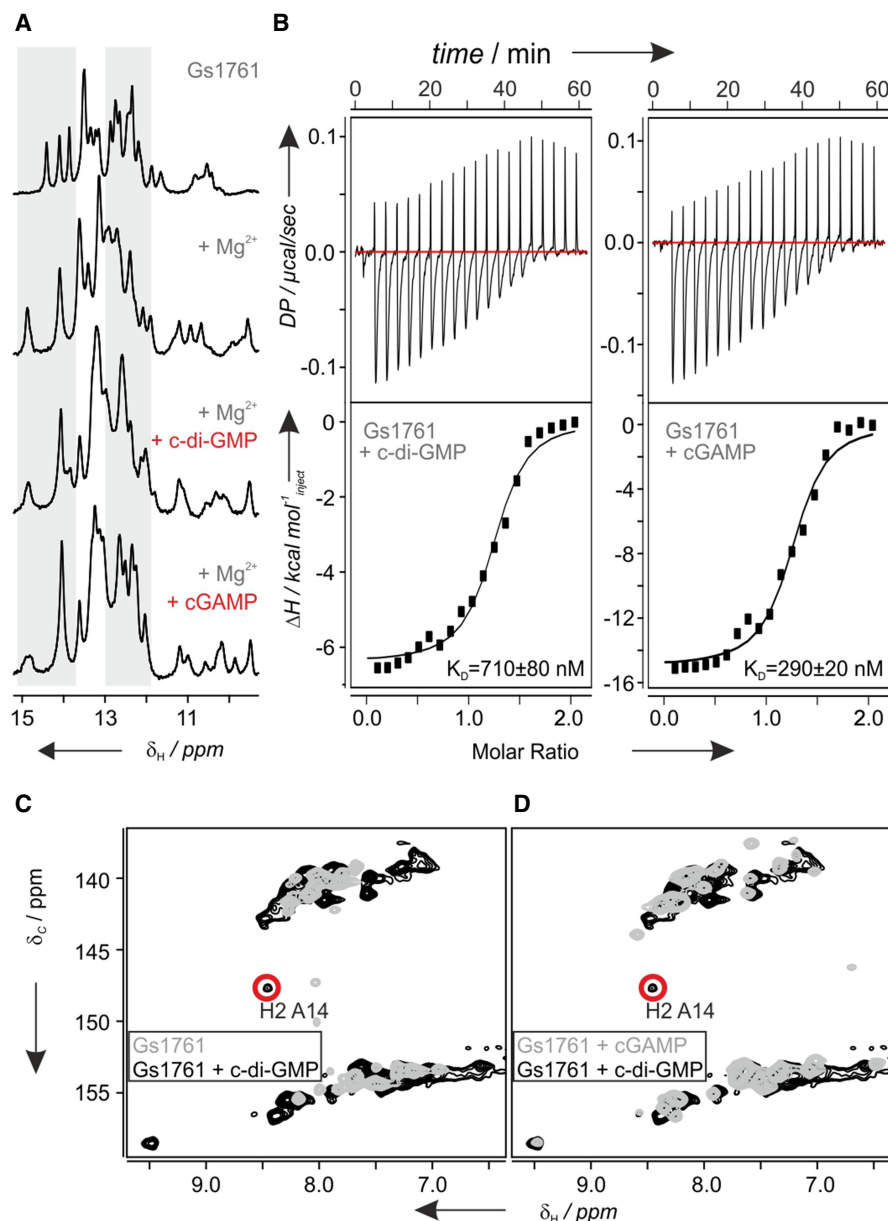


FIGURE 6. Ligand binding to the Gs1761 riboswitch and evidence for adenine protonation in the complex with c-di-GMP. (A) Imino proton spectra of the Gs1761 riboswitch RNA in the ligand-free state (top), in the presence of Mg^{2+} (middle), in the presence of Mg^{2+} and c-di-GMP (middle) and in the presence of Mg^{2+} and c-GAMP (bottom). Chemical shift regions with pronounced spectral changes are shaded in gray. (B) Representative ITC thermograms and fits for Gs1761 riboswitch RNA binding to c-di-GMP (left) and c-GAMP (right), respectively. (C) Overlay of ^{13}C -HSQC spectra of ^{13}C , ^{15}N -adenine labeled Gs1761 riboswitch RNA in its ligand-free state (gray) or bound to c-di-GMP (black). RNA-folding is induced by ligand binding as observed by large spectral changes. Interestingly, one signal appears with a ^{13}C chemical shift of 147.7 ppm and a 1H chemical shift of 8.46 ppm (red circle)—the chemical shift range associated with C2 carbon nuclei of N1 protonated adenine nucleotides. (D) Overlay of ^{13}C -HSQC spectra of ^{13}C , ^{15}N -adenine labeled Gs1761 riboswitch RNA bound to c-di-GMP (black) or c-GAMP (gray). The signal indicative of the adenine protonation is absent in the c-GAMP bound complex.

adenine protonation does not contribute to but probably rather limits the maximally achievable ligand selectivity since it supports c-di-GMP binding to c-GAMP riboswitches by allowing the formation of two hydrogen bonds between G_a and the relevant riboswitch adenine residue. The significant differences in ligand affinities measured for the Gs1761 constructs with a natural P1 stem used by

Hammond and coworkers (Kellenberger et al. 2015) and for the Gs1761 variants used in this work and by Patel's group (Ren et al. 2015) with a strongly stabilized P1 stem suggest that the composition and stability of the P1 stem might contribute to ligand specificity by promoting c-GAMP binding. Furthermore, high potassium ion concentrations apparently also support high affinity c-GAMP

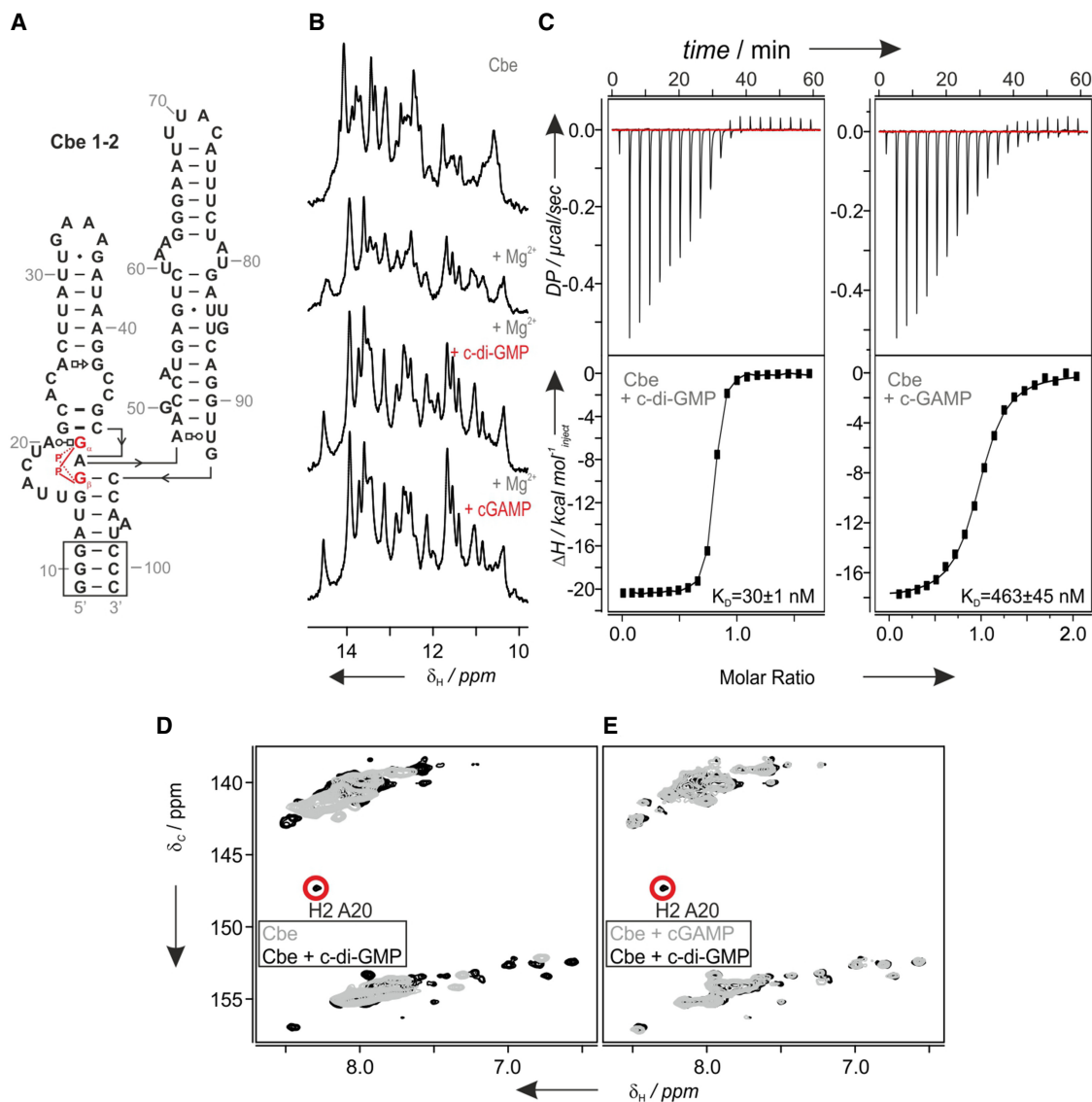


FIGURE 7. Ligand binding to the *Clostridium beijerinckii* (Cbe 1–2) riboswitch and evidence for adenine protonation in the complex with *c*-di-GMP. (A) Secondary structure of the Cbe 1–2 riboswitch aptamer domain. (B) Imino proton spectra of the Cbe 1–2 riboswitch RNA in the ligand-free state (top), in the presence of Mg^{2+} (middle), in the presence of Mg^{2+} and *c*-di-GMP (middle) and in the presence of Mg^{2+} and *c*-GAMP (bottom). (C) Representative ITC thermograms and fits for Cbe 1–2 riboswitch RNA binding to *c*-di-GMP (left) and *c*-GAMP (right), respectively. The resulting K_D values are given for both complexes. (D) Overlay of ^{13}C -HSQC spectra of ^{13}C ^{15}N -adenine labeled Cbe 1–2 riboswitch RNA in its ligand-free state (gray) or bound to *c*-di-GMP (black). The signal at a ^{13}C chemical shift of 147.3 ppm and a 1H chemical shift of 8.29 ppm (red circle) appears in the chemical shift range that is associated with C2 carbon nuclei of N1 protonated adenine nucleotides. (E) Overlay of ^{13}C -HSQC spectra of ^{13}C , ^{15}N -adenine labeled Cbe 1–2 riboswitch RNA bound to *c*-di-GMP (black) or *c*-GAMP (gray). The signal indicative of the adenine protonation is present only in the *c*-di-GMP bound complex.

binding. However, a more complete analysis of the structural determinants for ligand specificity in this riboswitch class is beyond the scope of this work.

Overall, the experiments described here clarify the hydrogen bonding patterns between the GEMM-1b riboswitches with their near-cognate ligand *c*-di-GMP. Interatomic distances between heavy atoms observed in X-ray crystallographic studies suggested that either a protonated A nucleotide of the riboswitch forms two hy-

drogen bonds with G_{α} of the ligand in its canonical protonation state or that a standard A forms hydrogen bonds to the G_{α} of the ligand assuming a rare tautomeric state. NMR spectroscopy unambiguously shows a protonated A of the riboswitch when bound to *c*-di-GMP and thereby establishes the *c*-di-GMP/GEMM-1b complex as the first example for an RNA where a protonated nucleotide is directly involved in ligand binding. Remarkably, the N1 protonated state of the A in this complex is persistent

even at pH values well above neutrality. Thus, environmental pH changes apparently do not contribute to the modulation of ligand specificity of the GEMM-Ib riboswitch. Instead, adenine protonation might be an important contribution to the widespread bispecific binding behavior for c-GAMP and c-di-GMP observed for members of this riboswitch class.

MATERIALS AND METHODS

Templates for in vitro transcription

We used PCR generated double-stranded DNA fragments or linearized plasmid DNA as template for T7 in vitro transcription. Overlapping oligonucleotides encoding a modified version of the c-di-GMP sensing riboswitch from *Vibrio cholera* (Vc2, Smith et al. 2009) and a mutant of this riboswitch harboring a guanine to adenine mutation at position 20 (Vc2 G20A) were used to generate double-stranded DNA fragments using PCR. pUC18-derived plasmids encoding the template sequences for Vc2, Vc2 G20A and modified versions of the cGAMP sensing riboswitches from *Geobacter sulfurreducens* (Gs1761, Ren et al. 2015) and *Clostridium beijerinckii* (Cbe 1–2, Nelson et al. 2015) were generated by Gibson cloning (Gibson et al. 2009).

RNA preparation

All RNAs were synthesized by run-off in vitro transcription using T7 RNA polymerase as previously described (Stoldt et al. 1998). Unlabeled, ^{15}N -adenine or ^{13}C , ^{15}N -adenine labeled Vc2, Vc2 G20A, Gs1761, Gs1761WT as well as Cbe 1–2 were transcribed using *Sma*I linearized plasmid DNA or PCR generated double-stranded DNA fragments as templates and purified by denaturing PAGE as previously described (Duchardt-Ferner et al. 2010). ^{13}C , ^{15}N - and ^{15}N -labeled rNTPs were purchased commercially (Silantes).

RNAs were folded under conditions favoring monomeric hairpin structures in a low salt buffer (2.5 mM Bis-Tris, pH 6.5) by denaturing at 70°C for 5 min, rapid 10-fold dilution with the same ice-cold buffer and subsequent annealing on ice. They were exchanged into 25 mM Bis-Tris buffer, pH 6.5, 5 mM magnesium acetate for NMR spectroscopy using ultracentrifugation devices (VivaSpin 2, MWC 3 kDa) and multiple cycles of concentration and dilution with NMR buffer. The RNAs in the final NMR samples were monomeric and homogeneous as judged from a native gel (Supplemental Fig. S5).

Unlabeled c-di-GMP and c-GAMP were purchased commercially as sodium salts (Sigma Aldrich) and dissolved in NMR buffer.

Synthesis and purification of ^{13}C , ^{15}N -labeled c-di-GMP

^{13}C , ^{15}N -labeled c-di-GMP was enzymatically synthesized as described in Rao et al. (2009) using the stand-alone GGEEF domain of the *Thermotoga maritima* diguanylate cyclase (TM1788) harboring the R158A mutation, except that we used ^{13}C , ^{15}N -labeled GTP as substrate for synthesis. Reactions were run at 45°C in TM1788 reaction buffer (50 mM Tris/HCl, pH 7.5, 250 mM

NaCl, 20 mM MgCl_2 , 1 mM DTT). The progress of the reaction was monitored by ^{31}P NMR and indicated complete conversion of GTP to c-di-GMP (Supplemental Fig. S6).

Magnesium pyrophosphate that precipitated during synthesis was separated by centrifugation. ^{13}C , ^{15}N -labeled c-di-GMP was purified from the reaction by phenol–chloroform extraction of the supernatant followed by ethanol precipitation of the aqueous phase. The pellet was dissolved in water and the concentration was determined spectroscopically using $26,100\text{ M}^{-1}\text{ cm}^{-1}$ as the extinction coefficient (ϵ_{260}) for c-di-GMP. 1D- ^1H , 1D- ^{13}C , 1D- ^{31}P , and 2D- ^{13}C -HSQC-spectra of the product showed no impurities (Fig. 5A, top; Supplemental Fig. S6) and most importantly the absence of any remaining GTP. The molecular mass of the product was confirmed by high-resolution MALDI mass spectrometry.

NMR spectroscopy

NMR experiments were recorded on 600 MHz Bruker AV, 600 MHz Bruker AVIII-HD, and 950 MHz Bruker AVII NMR spectrometers, all equipped with 5 mm, z-axis gradient ^1H [^{13}C , ^{15}N]-TCI cryogenic probes using standard pulse sequences (Fürtig et al. 2003). 1D imino proton spectra were recorded at 10°C using jump-return water suppression whereas ^{13}C -HSQC-spectra were recorded at 25°C. ^{13}C -1D spectra were recorded on an 800 MHz Bruker AV NMR spectrometer equipped with a 5 mm, z-axis gradient ^{13}C , ^{15}N [^1H]-TXO cryogenic probe at 25°C. Unless stated otherwise, samples were measured in 25 mM Bis-Tris (pH 6.5), 5 mM magnesium acetate containing 7.5% D_2O . Spectra were processed and analyzed in TOPSPIN 3.2 (Bruker Biospin). RNA concentrations ranged from 100 μM for unlabeled RNA, 300 μM for ^{13}C , ^{15}N -adenine labeled RNA to 900 μM for ^{15}N -adenine labeled RNA. Samples with isotopically labeled RNA contained 1.2 equivalents of unlabeled c-di-GMP. The sample used for recording the 1D- ^{13}C -spectrum of ^{13}C , ^{15}N -labeled c-di-GMP bound to the unlabeled G20A-riboswitch contained 400 μM c-di-GMP and 1.1 equivalents of RNA.

Isothermal titration calorimetry (ITC)

Unlabeled RNAs and both ligands (c-di-GMP, c-GAMP) were prepared in 50 mM Bis-Tris (pH 6.5) containing 5 mM magnesium acetate, if not indicated otherwise. Additional experiments to test for the influence of potassium ions on the binding affinities were carried out in buffer containing 25 mM potassium phosphate buffer, pH 6.5 and 5 mM magnesium acetate with or without 250 mM potassium chloride, respectively. ITC experiments at pH 8.3 were carried out in 50 mM Tris/HCl buffer containing 0, 2, 5, or 10 mM Mg^{2+} . The ligand (200–450 μM) was injected into a solution of 20–45 μM RNA. All measurements were performed at 25°C using a MicroCal iTC₂₀₀ instrument (Malvern Instruments). After an initial waiting time of 120 sec, the first injection of 0.2 μL was followed by 19 serial injections of 2 μL , separated by intervals of 180–1440 sec. For each experiment, the reference power was set to 11 μcal^{-1} , stirring speed to 750 rpm and the high feedback mode was selected for experiments with c-di-GMP, while for c-GAMP experiments the low feedback mode was chosen. Three independent titrations were performed and the reported K_D values are the average from these titrations. The thermograms were

processed using Origin7.0 (OriginLab) assuming a one-site binding model. An overview of all binding parameters derived from the ITC measurements is given in Supplemental Tables 1 and 2.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by the Center of Biomolecular Magnetic Resonance (BMRZ), the Deutsche Forschungsgemeinschaft (WO 901/6-1 as part of the special focus program SPP 1879 "Nucleotide Second Messenger Signaling in Bacteria" and project B10 of the CRC [SFB] 902 "Molecular Principles of RNA-based Regulation"). The ^{13}C , ^{15}N [^1H]-TXO cryogenic probe for ^{13}C -direct detection NMR experiments was obtained with support from the DFG (DFG INST 161/817-1 FUGG). We are grateful to members of the Wöhnert group for helpful discussions, to Kerstin Yacoub for help with sample preparations, to Elke Duchardt-Ferner for advice about NMR methods, and to Zhao-Xun Liang for a gift of the plasmid for the expression of a c-di-GMP synthase mutant from *T. maritima*.

Received May 26, 2018; accepted July 9, 2018.

REFERENCES

- Brodsky AS, Erlacher HA, Williamson JR. 1998. NMR evidence for a base triple in the HIV-2 TAR C-G.C+ mutant-argininamide complex. *Nucleic Acids Res* **26**: 1991–1995.
- Cai Z, Tinoco Jr. 1996. Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry* **35**: 6026–6036.
- Cash DD, Cohen-Zontag O, Kim N-K, Shefer K, Brown Y, Ulyanov NB, Tzfati Y, Feigon J. 2013. Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo. *Proc Natl Acad Sci* **110**: 10970–10975.
- Cornish PV, Hennig M, Giedroc DP. 2005. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Proc Natl Acad Sci* **102**: 12694–12699.
- Davies BW, Bogard RW, Young TS, Mekalanos JJ. 2012. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* **149**: 358–370.
- Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. 2012. A new understanding of the decoding principle on the ribosome. *Nature* **484**: 256–259.
- Dingley AJ, Grzesiek S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2\text{J}_{\text{NN}}$ couplings. *J Am Chem Soc* **120**: 8293–8297.
- Duchardt-Ferner E, Wöhnert J. 2017. NMR experiments for the rapid identification of P=O...H-X type hydrogen bonds in nucleic acids. *J Biomol NMR* **69**: 101–110.
- Duchardt-Ferner E, Weigand JE, Ohlenschläger O, Schmidtke SR, Suess B, Wöhnert J. 2010. Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angew Chem Int Ed Engl* **49**: 6216–6219.
- Duchardt-Ferner E, Ferner J, Wöhnert J. 2011. Rapid identification of noncanonical RNA structure elements by direct detection of OH...O=P, NH...O=P, and NH2...O=P hydrogen bonds in solution NMR spectroscopy. *Angew Chem Int Ed Engl* **50**: 7927–7930.
- Fürtig B, Richter C, Wöhnert J, Schwalbe H. 2003. NMR spectroscopy of RNA. *Chembiochem* **4**: 936–962.
- Gentner M, Allan MG, Zaehring F, Schirmer T, Grzesiek S. 2012. Oligomer formation of the bacterial second messenger c-di-GMP: reaction rates and equilibrium constants indicate a monomeric state at physiological concentrations. *J Am Chem Soc* **134**: 1019–1029.
- Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CAIII, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343–345.
- Gleghorn ML, Zhao J, Turner DH, Maquat LE. 2016. Crystal structure of a poly(rA) staggered zipper at acidic pH: evidence that adenine N1 protonation mediates parallel double helix formation. *Nucleic Acids Res* **44**: 8417–8424.
- Gong B, Chen JH, Chase E, Chadalavada DM, Yajima R, Golden BL, Bevilacqua PC, Carey PR. 2007. Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc* **129**: 13335–13342.
- Gottstein-Schmidtke SR, Duchardt-Ferner E, Weigand JE, Groher F, Suess B, Wöhnert J. 2014. Building a stable RNA U-turn with a protonated cytidine. *RNA* **20**: 1163–1172.
- Houck-Loomis B, Durney MA, Salguero C, Shankar N, Nagle JM, Goff SP, D'Souza VM. 2011. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480**: 561–564.
- Huppler A, Niekstad LJ, Allmann AM, Brow DA, Butcher SE. 2002. Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nat Struct Biol* **9**: 431–435.
- Kader A, Simm R, Gerstel U, Morr M, Römling U. 2006. Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* **60**: 602–616.
- Kath-Schorr S, Wilson TJ, Li NS, Lu J, Piccirilli JA, Lilley DM. 2012. General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *J Am Chem Soc* **134**: 16717–16724.
- Kellenberger CA, Wilson SC, Sales-Lee J, Hammond MC. 2013. RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc* **135**: 4906–4609.
- Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, et al. 2015. GEMM-I riboswitches from *Geobacter* sense the bacterial second messenger cyclic AMP-GMP. *Proc Natl Acad Sci* **112**: 5383–5388.
- Kimsey IJ, Petzold K, Sathyamoorthy B, Stein ZW, Al-Hashimi HM. 2015. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**: 315–320.
- Kulshina N, Baird NJ, Ferré-D'Amaré AR. 2009. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol* **16**: 1212–1217.
- Legault P, Pardi A. 1994. *In situ* probing of adenine protonation in RNA by ^{13}C NMR. *J Am Chem Soc* **116**: 8390–8391.
- Legault P, Pardi A. 1997. Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *J Am Chem Soc* **119**: 6621–6628.
- Liu Y, Wilson TJ, McPhee SA, Lilley DM. 2014. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat Chem Biol* **10**: 739–744.
- Macaya RF, Gilbert DE, Malek S, Sinsheimer JS, Feigon J. 1991. Structure and stability of X.G.C mismatches in the third strand of intramolecular triplexes. *Science* **254**: 270–274.
- Nakano S, Bevilacqua PC. 2007. Mechanistic characterization of the HDV genomic ribozyme: a mutant of the C41 motif provides

- insight into the positioning and thermodynamic linkage of metal ions and protons. *Biochemistry* **46**: 3001–3012.
- Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR. 2015. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci* **112**: 5389–5394.
- Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, Hennig M, Giedroc DP. 2002. Solution structure of a luteoviral P1-P2 frame-shifting mRNA pseudoknot. *J Mol Biol* **322**: 621–633.
- Pan B, Mitra SN, Sundaralingam M. 1999. Crystal structure of an RNA 16-mer duplex R(GCAGAGUAAAUCUGC)₂ with nonadjacent G (syn):A⁺(anti) mispairs. *Biochemistry* **38**: 2826–2831.
- Pechlaner M, Donghi D, Zelenay V, Sigel RK. 2015. Protonation-dependent base flipping at neutral pH in the catalytic triad of a self-splicing bacterial group II intron. *Angew Chem Int Ed Engl* **54**: 9687–9690.
- Puglisi JD, Wyatt JR, Tinoco Jr. 1990. Solution conformation of an RNA hairpin loop. *Biochemistry* **29**: 4215–4226.
- Rao F, Pasunooti S, Ng Y, Zhuo W, Lim L, Liu AW, Liang ZX. 2009. Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. *Anal Biochem* **389**: 138–142.
- Ren A, Košutić M, Rajashankar KR, Frener M, Santner T, Westhof E, Micura R, Patel DJ. 2014. In-line alignment and Mg²⁺ coordination at the cleavage site of the env22 twister ribozyme. *Nat Commun* **5**: 5534.
- Ren A, Wang XC, Kellenberger CA, Rajashankar KR, Jones RA, Hammond MC, Patel DJ. 2015. Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep* **11**: 1–12.
- Ren A, Vušurović N, Gebetsberger J, Gao P, Juen M, Kreutz C, Micura R, Patel DJ. 2016. Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. *Nat Chem Biol* **12**: 702–708.
- Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. 2015. Structural insights into the translational infidelity mechanism. *Nat Commun* **6**: 7251.
- Rozov A, Demeshkina N, Khusainov I, Westhof E, Yusupov M, Yusupova G. 2016. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat Commun* **7**: 10457.
- Rupert PB, Massey AP, Sigurdsson ST, Ferré-D'Amaré AR. 2002. Transition state stabilization by a catalytic RNA. *Science* **298**: 1421–1424.
- Rypniewski W, Banaszak K, Kuliński T, Kiliszek A. 2016. Watson-Crick-like pairs in CCUG repeats: evidence for tautomeric shifts or protonation. *RNA* **22**: 22–31.
- Saenger W. 1988. *Principles of nucleic acid structure*. Springer, New York.
- Simm R, Morr M, Remminghorst U, Andersson M, Römling U. 2009. Quantitative determination of cyclic diguanosine monophosphate concentrations in nucleotide extracts of bacteria by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. *Anal Biochem* **386**: 53–58.
- Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA. 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat Struct Mol Biol* **16**: 1218–1223.
- Smith KD, Lipchock SV, Livingston AL, Shanahan CA, Strobel SA. 2010. Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch. *Biochemistry* **49**: 7351–7359.
- Stoldt M, Wöhnert J, Görlach M, Brown LR. 1998. The NMR structure of *Escherichia coli* ribosomal protein L25 shows homology to general stress proteins and glutamyl-tRNA synthetases. *EMBO J* **17**: 6377–6384.
- Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**: 411–413.
- Thaplyal P, Bevilacqua PC. 2014. Experimental approaches for measuring pKa's in RNA and DNA. *Methods Enzymol* **549**: 189–219.
- Wickiser JK, Winkler WC, Breaker RR, Crothers DM. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18**: 49–60.
- Wilcox JL, Bevilacqua PC. 2013a. A simple fluorescence method for pK_a determination in RNA and DNA reveals highly shifted pK_a's. *J Am Chem Soc* **135**: 7390–7393.
- Wilcox JL, Bevilacqua PC. 2013b. pKa shifting in double-stranded RNA is highly dependent upon nearest neighbors and bulge positioning. *Biochemistry* **52**: 7470–7476.
- Wilcox JL, Ahluwalia AK, Bevilacqua PC. 2011. Charged nucleobases and their potential for RNA catalysis. *Acc Chem Res* **44**: 1270–1279.
- Wilson TJ, Li NS, Lu J, Frederiksen JK, Piccirilli JA, Lilley DM. 2010. Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proc Natl Acad Sci* **107**: 11751–11756.
- Wilson TJ, Liu Y, Domnick C, Kath-Schorr S, Lilley DM. 2016. The novel chemical mechanism of the Twister ribozyme. *J Am Chem Soc* **138**: 6151–6162.
- Wöhnert J, Dingley AJ, Stoldt M, Görlach M, Grzesiek S, Brown LR. 1999. Direct identification of NH...N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res* **27**: 3104–3110.
- Wolter AC, Weickmann AK, Nasiri AH, Hantke K, Wunderlich C, Kreutz C, Duchardt-Ferner E, Wöhnert J. 2017. A stably protonated adenine nucleotide with an extremely shifted pKa stabilizes the tertiary structure of a GTP-binding RNA-aptamer. *Angew Chemie Int Ed Engl* **56**: 401–404.

3. WEICKHMANN ET AL – NMR RESONANCE ASSIGNMENT OF SAM/SAH RIBOSWITCH

Title: NMR resonance assignments for the SAM/SAH-binding riboswitch RNA bound to S-adenosylhomocysteine

Authors: Weickhmann AK, Keller H, Duchardt-Ferner E, Strebiter E, Juen MA, Kremser J, Wurm JP, Kreutz C, Wöhnert J

Published in: Biomolecular NMR Assignments 12, 329–334 (2018)

Contributions

(1) Entwicklung und Planung	Weickhmann AK (30%), Keller H (15%), Duchardt-Ferner E (25%), Wöhnert J (30%)
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Weickhmann AK: Klonierung, RNA-Präparation, NMR-Messungen, Herstellung und Aufreinigung von markiertem SAM; Keller H: Klonierung, initiale NMR-Messungen; Duchardt-Ferner E: NMR-Messungen; Strebiter E: Herstellung selektiv markierter Proben; Juen MA: Herstellung selektiv markierter Proben; Kremser J: Herstellung selektiv markierter Proben; Wurm JP: Herstellung und Aufreinigung von markiertem SAM
(3) Erstellung der Datensammlung und Abbildungen	Weickhmann AK: Erstellung von Abbildungen, Datensammlung für BMRB
(4) Analyse/ Interpretation der Daten	Weickhmann AK: Auswertung der NMR-Messungen; Keller H: initiale NMR-Auswertungen
(5) Übergeordnete Einleitung/ Ergebnisse/Diskussion	Weickhmann AK (60%), Wöhnert J (40%)



NMR resonance assignments for the SAM/SAH-binding riboswitch RNA bound to S-adenosylhomocysteine

A. Katharina Weickhmann¹ · Heiko Keller¹ · Elke Duchardt-Ferner¹ · Elisabeth Strebitzer² · Michael A. Juen² · Johannes Kremser² · Jan Philip Wurm^{1,3} · Christoph Kreutz² · Jens Wöhnert¹

Received: 22 May 2018 / Accepted: 25 July 2018
© Springer Nature B.V. 2018

Abstract

Riboswitches are structured RNA elements in the 5'-untranslated regions of bacterial mRNAs that are able to control the transcription or translation of these mRNAs in response to the specific binding of small molecules such as certain metabolites. Riboswitches that bind with high specificity to either S-adenosylmethionine (SAM) or S-adenosylhomocysteine (SAH) are widespread in bacteria. Based on differences in secondary structure and sequence these riboswitches can be grouped into a number of distinct classes. X-ray structures for riboswitch RNAs in complex with SAM or SAH established a structural basis for understanding ligand recognition and discrimination in many of these riboswitch classes. One class of riboswitches—the so-called SAM/SAH riboswitch class—binds SAM and SAH with similar affinity. However, this class of riboswitches is structurally not yet characterized and the structural basis for its unusual bispecificity is not established. In order to understand the ligand recognition mode that enables this riboswitch to bind both SAM and SAH with similar affinities, we are currently determining its structure in complex with SAH using NMR spectroscopy. Here, we present the NMR resonance assignment of the SAM/SAH binding riboswitch (env9b) in complex with SAH as a prerequisite for a solution NMR-based high-resolution structure determination.

Keywords RNA · Riboswitch · SAH · SAM · NMR assignment · Triple resonance experiments

Introduction

Riboswitches are regulatory RNA elements located mainly in the 5'-untranslated regions (UTR) of bacterial mRNAs (Roth and Breaker 2009). Ligand binding to these RNA elements induces a conformational rearrangement resulting

in changes of gene expression on the level of either transcription or translation (Grundy and Henkin 2006). Typical riboswitch ligands are small molecules such as metabolites, enzyme cofactors, amino acids, nucleobases, second-messenger molecules or ions. Riboswitches can exploit different molecular architectures to recognize the same ligand. In particular, for the physiologically important coenzyme S-adenosylmethionine (SAM) three structurally distinct riboswitch superfamilies have been characterized. They all regulate the expression of enzymes involved in the SAM biosynthesis pathway and are highly selective for SAM (Wang and Breaker 2008). However, their secondary and tertiary structures as well as their ligand binding pocket architectures are strikingly different from each other. Additionally, a riboswitch has been identified which is highly specific for S-adenosylhomocysteine (SAH), the product of SAM-mediated methylation reactions. It discriminates strongly against SAM (Wang et al. 2008) and regulates the expression of enzymes involved in SAH recycling. In contrast to these riboswitch classes which discriminate strongly between SAM and SAH, the recently identified SAM/SAH

✉ Jens Wöhnert
woehnert@bio.uni-frankfurt.de

A. Katharina Weickhmann
weickhmann@bio.uni-frankfurt.de

¹ Institute for Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe-University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany

² Institute of Organic Chemistry, Centre for Molecular Biosciences (CMBI), University of Innsbruck, Innrain 80/82, 6020 Innsbruck, Austria

³ Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany

binding riboswitches found in α -proteobacteria are bispecific for SAM and SAH (Weinberg et al. 2010) and bind both ligands with K_{DS} in the low μM range. This riboswitch class contains a 5'-terminal hairpin structure that is predicted to interact with six nucleotides 3' of the hairpin to form a pseudoknot structure upon ligand binding (Fig. 1a). The hairpin and the pseudoknot are connected by a linker of variable length (9–16 nt). Upon ligand binding the ribosomal binding site is likely to be included into the tertiary structure and thereby ribosomal access is hindered. The hairpin loop and the pseudoknot-forming sequence are highly conserved, while the hairpin stem and the linker residues are not.

The riboswitch RNA *env9* is one of 176 RNA sequences reported so far to be part of the SAM/SAH riboswitch class (Rfam: RF01727). It is derived from an environmental bacterial DNA sample where it is usually associated with a SAM synthase (*metK*) gene. *Env9* is 43 nt in length and binds both SAM and SAH with low micromolar affinities as expected (data not shown). Prior to our NMR-experiments aimed at structure determination, we mutated the hairpin stem P1 of *env9* (Fig. 1a) to yield an RNA construct with either two GC closing base pairs to increase P1 stability (designated *env9b*) or an GU and an AU base pair to facilitate NMR assignment (designated *env9b_P1*). The ligand affinities were not altered by these mutations (data not shown) compared to the wild type *env9* RNA. Here, we report the ^1H , ^{13}C and ^{15}N resonance assignments of *env9b_P1* in complex with the ligand SAH as a prerequisite for an NMR-based structure determination.

Methods and experiments

Unlabeled, uniformly ^{15}N -, uniformly $^{13}\text{C}/^{15}\text{N}$ -, uniformly 5-D₁/ribose-3',4',5',5"-D₄-labeled and three different nucleotide-type selectively labeled samples ($^{13}\text{C}/^{15}\text{N}$ -A/C-, $^{13}\text{C}/^{15}\text{N}$ -G- or $^{13}\text{C}/^{15}\text{N}$ -U) of the *env9b_P1* riboswitch were prepared by in vitro transcription with T7 polymerase and purified by denaturing polyacrylamide gel electrophoresis as described in detail elsewhere (Duchardt-Ferner et al. 2010). Linearized plasmid DNA was used as template for in vitro transcription (Milligan and Uhlenbeck 1989). In order to obtain RNAs with homogeneous 3'-ends, the primary RNA-transcript contained a hepatitis delta virus (HDV) ribozyme fused to the 3'-end of the target RNA (Ferré-D'Amaré and Doudna 1996). All labeled nucleotide triphosphates were purchased commercially (Silantes GmbH, Germany and Cambridge Isotope Laboratories, Inc.). Samples with site-specifically pyrimidine ^{13}C -C6, purine ^{13}C -C8, uridine $^{15}\text{N}_1$, $^{15}\text{N}_3$ ($^{15}\text{N}_2$) or adenine $^{13}\text{C}_2$, $^{13}\text{C}_8$ labeled nucleotides were chemically synthesized (Wunderlich et al. 2012; Juen et al. 2016). The site-specifically labeled samples were either in an *env9b* ($^{15}\text{N}_2$ -U12, $^{15}\text{N}_2$ -U16, $^{15}\text{N}_2$ -U37,

^{13}C -C8-A7/ ^{13}C -C6-U12, ^{13}C -C8-G10/ ^{13}C -C6-U16, ^{13}C -C8-A6/ ^{13}C -C6-C15, ^{13}C -C6-C11/ ^{13}C -C8-G17, ^{13}C -C8-G9/ ^{13}C -C6-C14) or in an *env9b_P1* background (^{13}C -C6-U13/ ^{13}C -C6-C19, ^{13}C -C6-C8/ ^{13}C -C8-A18, ^{13}C -C8-A35/ ^{13}C -C6-U36, ^{13}C -C2,C8-A35/ ^{13}C -C6-U37, ^{13}C -C6-U34/ ^{13}C -C2,C8-A35) to minimize signal overlap. The sample concentrations varied from 0.2 to 1.2 mM in NMR buffer containing 25 mM potassium phosphate buffer, pH 6.2 and 50 mM potassium chloride with 5, 10 or 100% (v/v) D₂O, respectively. ^{13}C , ^{15}N -labeled SAH was synthesized using ^{13}C , ^{15}N -labeled SAM obtained as described previously (Ottink et al. 2010) and the methyltransferase PaMTH1 and its substrate myricetin (Chatterjee et al. 2015). ^{13}C , ^{15}N -labeled SAH was purified using RP-HPLC. Unlabeled SAH was obtained commercially (Sigma-Aldrich). NMR spectra were recorded at 5 and 10 °C for experiments involving the detection of exchangeable protons or at 25 °C for the detection of non-exchangeable protons on Bruker AVANCE III 600, 700, 800, 900 or 950 MHz spectrometers with cryogenic triple resonance HCN-probes. The data were processed using TOPSPIN 3.2 or 3.5 software (Bruker BioSpin, Germany) and analyzed using the software CARA (Keller 2004). All samples were titrated with SAH to saturation as observed in 1D ^1H spectra with a final RNA:ligand ratio of ~1:4 in the presence of 2 mM magnesium acetate.

Initial assignment of the exchangeable RNA imino and amino protons relied on ^1H , ^1H -NOESY-experiments with either WATERGATE- or jump-return water suppression recorded for an unlabeled sample in conjunction with WATERGATE- ^1H , ^{15}N -HSQC-experiments optimized for either imino or amino groups and HNN-COSY-experiments (Dingley and Grzesiek 1998; Wöhnert et al. 1999a) recorded with an uniformly ^{15}N -labeled sample. Additionally, the ^{15}N -labeled sample was used to record a 2-bond ^1H , ^{15}N -HSQC (Sklenár et al. 1994) for the correlation of H2 and the H8 protons to the adjacent N1/N3 and N7/N9 nitrogen nuclei, respectively. To correlate the resonances of the slowly exchanging imino protons to guanine C2/C6 or uridine C2/C4 resonances, respectively, an H(N)C experiment (Ohlenschläger et al. 2004) was used. To connect the guanine H8 resonances with the H1 imino proton resonances of the same spin system, an H1/8-C5-HMQC (Phan 2000) was recorded on the selectively G- ^{13}C , ^{15}N -labeled sample. Additionally, uridine and cytidine H5 and C5 signals were connected to the imino or amino resonances of the same spin system, respectively, by recording H5(C5C4N)H (Fig. 1c) spectra on selectively A/C- ^{13}C , ^{15}N - or selectively U- ^{13}C , ^{15}N -labeled samples (Wöhnert et al. 1999b). For uridines with weak imino resonances H5-N3 correlations were obtained using an Ir- ^1H , ^{15}N -HSQC (Sklenár et al. 1994) on a uniformly ^{15}N -labeled sample. ^{13}C -NOESY-HSQC spectra with WATERGATE water suppression (Piotto et al. 1992) were recorded for the aromatic CH moieties using

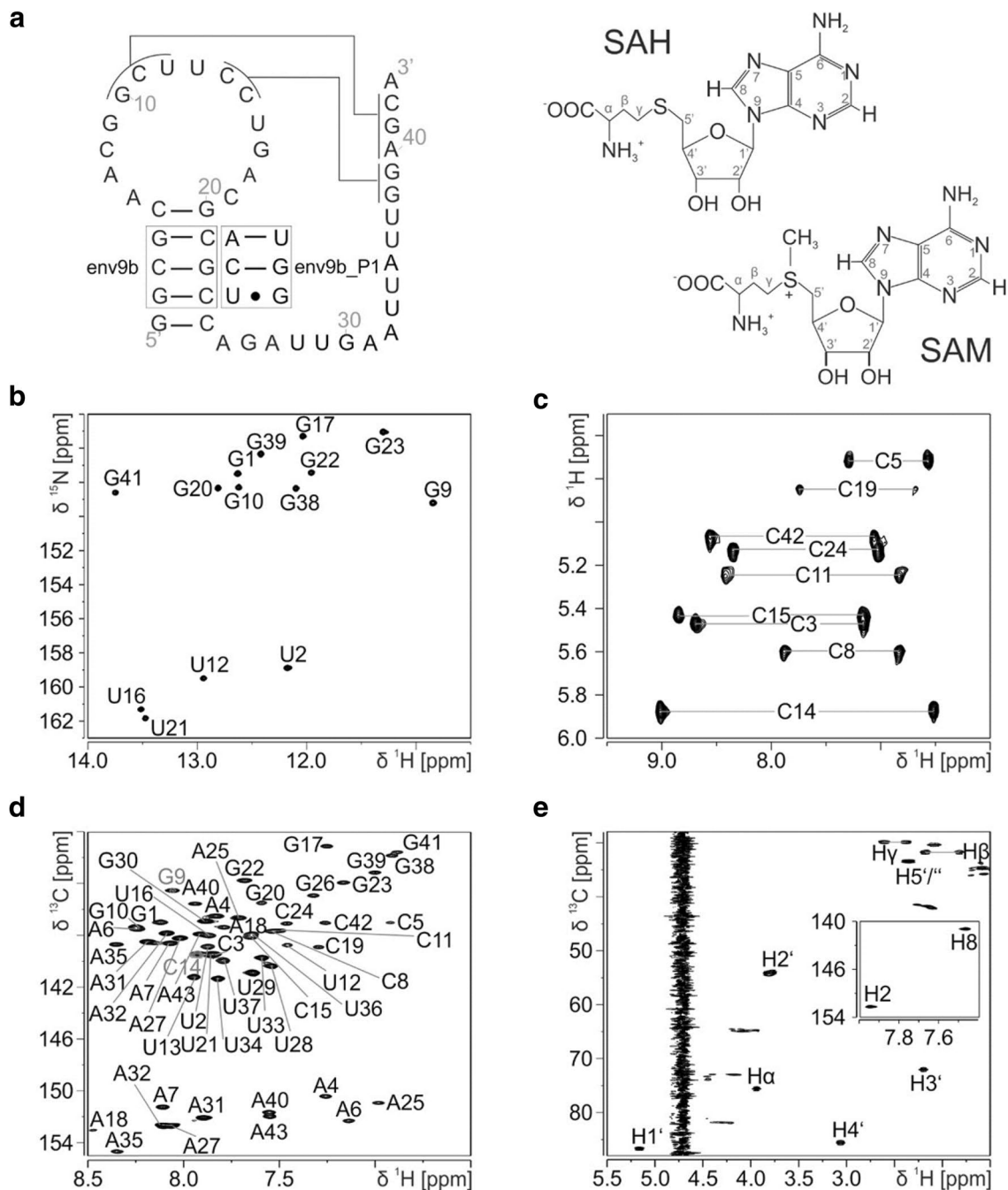


Fig. 1 **a** Sequence and predicted secondary structure (Weinberg et al. 2010) for the SAM/SAH binding riboswitch env9 (left). Mutations in the hairpin stem P1 to increase stem stability (env9b) or to facilitate NMR assignments (env9b_P1) are indicated in boxes. Putative non-Watson-Crick base pairs are indicated by a dot. Structures of S-adenosylhomocysteine (top, right) and S-adenosylmethionine (bottom, right) and their respective atomic numbering schemes. **b** ^1H , ^{15}N -HSQC optimized for imino groups recorded on a uniformly ^{15}N -labeled sample at 10 °C. The assignments of the guanine H1N1- and uridine H3N3- imino groups are given. **c** H5(C5C4N)H experi-

ment recorded on a selectively A/C- ^{13}C , ^{15}N -labeled sample at 25 °C. The respective assignments are given. A grey line indicates amino protons belonging to the same spin system. **d** Overlay of an ^1H , ^{13}C -HSQC spectrum optimized for the aromatic CH moieties recorded for a uniformly ^{13}C , ^{15}N -labeled sample at 25 °C (black) with the spectrum recorded for a site-specific ^{13}C -C8-G9/C6- ^{13}C -C14-labeled sample (red). The assignments for all aromatic CH moieties are given. **e** ^1H , ^{13}C -HSQC spectrum optimized for the detection of the aliphatic and aromatic (inset) CH signals of the ^{13}C , ^{15}N -labeled SAH bound to unlabeled RNA

the uniformly ^{13}C , ^{15}N -labeled sample and the selectively A/C-, G-, U- $^{13}\text{C}^{15}\text{N}$ -labeled samples to obtain sequential connectivities.

All remaining experiments were recorded in 100% (v/v) D_2O -containing buffers. ^{13}C -NOESY-HSQC experiments optimized for aliphatic CH moieties were recorded on the selectively A/C-, G- and U- $^{13}\text{C}^{15}\text{N}$ -labeled samples. They were used to delineate the H5/H6 correlations from strong intrabase H5-H6 NOE cross-peaks as well as base-to-ribose connectivities. To correlate the adenine H2 with the H8 resonances of the same spin system a 3D-TROSY relayed HCCH-COSY spectrum was recorded, which also yields the assignments of adenine C2, C4, C5, C6 and C8 spins (Simon et al. 2001). By comparing the ^1H , ^{13}C -HSQC spectra of the chemically synthesized site-specifically labeled samples to those of uniformly ^{13}C , ^{15}N -labeled samples, the assignments for the aromatic CH moieties were verified. The nucleobase spin systems were connected to their corresponding ribose spin systems by recording 2D-H(C)N spectra (Sklenár et al. 1993) optimized for either nucleobase or for ribose H1'C1' moieties, respectively, using the selectively A/C-, G- and U- $^{13}\text{C}^{15}\text{N}$ -labeled samples. The ribose spin systems were assigned using 3D-HC(C)H-COSY and -TOCSY experiments (Nikonowicz and Pardi 1992).

The signals of the bound ligand SAH were assigned using a ^1H , ^{13}C -HSQC for the nucleobase H2C2 and H8C8 assignment, a ^{13}C -NOESY-HMQC to assign the signals of the aminocarboxypropyl moiety (CH_α , CH_β , CH_γ) and an HC(C)H-COSY for the ribose spin system assignment. All experiments for the ligand assignment were performed on a sample with unlabeled RNA and ^{13}C , ^{15}N -labeled SAH at 25 °C in D_2O . The ligand is in slow exchange between the RNA-bound and the free form.

Assignment and data deposition

The ^1H , ^{15}N -HSQC spectrum at 10 °C showed imino group resonances for ten out of 12 guanine residues and for four out of the eleven uridine residues in the sequence which were all assigned (Fig. 1b). At lower temperatures (below 5 °C), an additional uridine imino group resonance was observed. It was assigned to U36 by correlating its N3 chemical shift to its H5 resonance in a Ir- ^1H , ^{15}N -HSQC. All nine cytidine, three adenine and five guanine amino group resonances appeared in an ^1H , ^{15}N -HSQC spectrum optimized for the detection of amino group signals and were assigned. Here, the amino group protons of one adenine and of four guanine amino groups have degenerate proton chemical shifts.

All aromatic H2C2, H5C5, H6C6 and H8C8 aromatic CH group resonances were assigned (Fig. 1d). The TROSY-relayed HCCH-COSY experiment also yielded all adenine C4 and ten out of eleven adenine C6 chemical

shifts. For all guanine residues with an observable imino proton, C5 chemical shifts were assigned from the H1/H8-C5-HMQC experiment. Additionally, C2 and C6 were assigned for these residues using the H(N)C experiment. Furthermore, using the H(N)C experiment the uridine C2 and C4 chemical shift were assigned for the four uridine residues with an observable imino proton at 10 °C. The combination of H(C)N and 2-bond ^1H , ^{15}N -HSQC experiments yielded assignments for all purine N9 and N7 resonances and for 91% of the adenine N1 and all adenine N3 resonances. Furthermore, the H(C)N experiments yielded assignments for 73% of the uridine N1 resonances and all of the cytidine N1 cytidine resonances. Complete cytidine N3 assignments were achieved using an intra-residual HNN-COSY experiment starting from the amino groups.

Base-to-ribose connectivities for all residues except U2, U16, U36 were obtained with the H(C)N experiments on nucleotide-selectively labeled samples. Intra-nucleotide NOE correlations from the ^{13}C -NOESY-HSQC spectra yielded complete assignments for all H1'C1' resonances. Also, complete H2'C2' assignments were achieved. These assignments were verified by recording a 2D-NOESY spectrum in D_2O for the 5- D_1 /ribose-3',4',5',5''- D_4 -labeled sample. There, the signal overlap is drastically reduced and H1'-H2' intra-residual NOE cross peaks were easily distinguishable from other inter-residual cross-peaks due to distinct differences in signal intensities. Additionally, 98% of the H3'C3'-CH moieties, 67% of the H4'C4'-moieties and 13 out of 43 H5'/H5''C5'-moieties were assigned.

The ligand SAH aromatic H2C2 and H8C8, ribose H1'C1', H2'C2'; H3'C3', H4'C4', H5'C5' and aminocarboxypropyl moiety (CH_α , CH_β , CH_γ) resonances were assigned using ^{13}C , ^{15}N -labeled SAH (Fig. 1e). Here, CH_β , CH_γ and $\text{CH}_5'/\text{H}_5''$ showed distinguishable resonances for both protons in the RNA-bound state. Furthermore, the N7 resonance of SAH was assigned in a 2-bond ^1H , ^{15}N -HSQC (Sklenár et al. 1994) on a sample containing ^{13}C , ^{15}N -labeled SAH and selectively uridine ^{15}N -labeled RNA.

Since in the 1D- ^{31}P spectrum the signals were broad and strongly overlapped we did not attempt to assign ^{31}P resonances for this RNA.

In the ^1H , ^{13}C -HSQC of the selectively uridine ^{13}C , ^{15}N -labeled sample, linker residues U28, U29, U33, U34, but also U13 showed very high signal intensities suggesting that these residues are flexible and not involved in stable tertiary interactions. This notion is borne out in a $\{^1\text{H}\}$, ^{13}C -hetNOE-experiment (Fig. 2) where the H6C6 groups of these nucleotides show elevated hetNOE values.

The chemical shifts for the SAH/RNA-complex have been deposited in the Biological Magnetic Resonance Bank (<http://www.bmrb.wisc.edu>) with the Accession Number 27452.

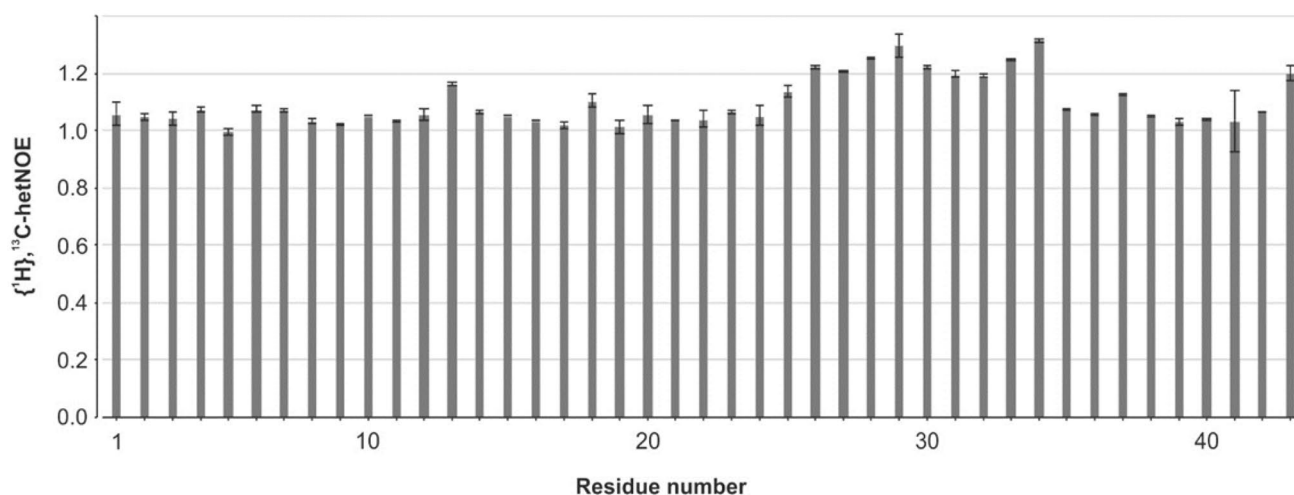


Fig. 2 $\{^1\text{H}\}, ^{13}\text{C}$ -hetNOE values for the aromatic H6C6 or H8C8 moieties of env9b_P1, respectively. HetNOE values > 1.2 indicate flexible residues

Acknowledgements We are grateful to Christian Richter and Manfred Strupf for maintenance of the NMR facility and to Kerstin Yacoub for help with sample preparation. This work was supported by the Center for Biomolecular Magnetic Resonance (BMRZ) of the Goethe University Frankfurt and the Deutsche Forschungsgemeinschaft (DFG) through the CRC (SFB) 902 “Molecular principles of RNA-based regulation”, Project B10.

References

- Chatterjee D, Kudlinzki D, Linhard V, Saxena K, Schieborr U, Gande SL, Wurm JP, Wöhnert J, Abele R, Rogov VV, Dötsch V, Osiewacz HD, Sreeramulu S, Schwalbe H (2015) Structure and biophysical characterization of the S-adenosylmethionine-dependent O-methyltransferase PaMTH1, a putative enzyme accumulating during senescence of *Podospira anserina*. *J Biol Chem* 290:16415–16430
- Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{\text{NN}}$ couplings. *J Am Chem Soc* 120:8293–8297
- Duchardt-Ferner E, Weigand JE, Ohlenschläger O, Schmidtke SR, Suess B, Wöhnert J (2010) Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angew Chem Int Ed Engl* 49:6216–6219
- Ferré-D’Amaré AR, Doudna JA (1996) Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res* 24:977–978
- Grundy FJ, Henkin TM (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* 41:329–338
- Juen MA, Wunderlich CH, Nußbaumer F, Tollinger M, Kontaxis G, Konrat R, Hansen DF, Kreutz C (2016) Excited states of nucleic acids probed by proton relaxation dispersion NMR spectroscopy. *Angew Chem Int Ed Engl* 55:12008–12012
- Keller R (2004) The computer aided resonance tutorial. CANTINA Verlag, Goldau
- Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. *Meth Enzymol* 180:51–62
- Nikonowicz EP, Pardi A (1992) Three-dimensional heteronuclear NMR studies of RNA. *Nature* 355:184–186
- Ohlenschläger O, Wöhnert J, Bucci E, Seitz S, Häfner S, Ramachandran R, Zell R, Görlach M (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure* 12:237–248
- Ottink OM, Nelissen FHT, Derks Y, Wijnenga SS, Heus HA (2010) Enzymatic stereospecific preparation of fluorescent S-adenosyl-L-methionine analogs. *Anal Biochem* 396:280–283
- Phan AT (2000) Long-range imino proton- ^{13}C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J Biomol NMR* 16:175–178
- Piotto M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665
- Roth A, Breaker RR (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 78:305–334
- Simon B, Zanier K, Sattler M (2001) A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly ^{13}C -labeled RNA molecules. *J Biomol NMR* 20:173–176
- Sklenár V, Peterson RD, Rejante MR, Feigon J (1993) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in $^{15}\text{N}, ^{13}\text{C}$ -labeled RNA oligonucleotides. *J Biomol NMR* 3:721–727
- Sklenár V, Peterson RD, Rejante MR, Feigon J (1994) Correlation of nucleotide base and sugar protons in a ^{15}N -labeled HIV-1 RNA oligonucleotide by $^1\text{H}-^{15}\text{N}$ HSQC experiments. *J Biomol NMR* 4:117–122
- Wang JX, Breaker RR (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem Cell Biol* 86:157–168
- Wang JX, Lee ER, Morales DR, Lim J, Breaker RR (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell* 29:691–702
- Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* 11:R31
- Wöhnert J, Dingley AJ, Stoldt M, Gölach M, Grzesiek S, Brown LR (1999a) Direct identification of NH... N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res* 27:3104–3110

Wöhnert J, Ramachandran R, Görlach M, Brown LR (1999b) Triple-resonance experiments for correlation of H5 and exchangeable pyrimidine base hydrogens in (13)C,(15)N-labeled RNA. *J Magn Reson* 139:430–433

Wunderlich CH, Spitzer R, Santner T, Fauster K, Tollinger M, Kreutz C (2012) Synthesis of (6-(13)C)pyrimidine nucleotides as spin-labels for RNA dynamics. *J Am Chem Soc* 134:7558–7569

4. WEICKHMANN ET AL – STRUCTURE OF THE SAM/SAH BINDING RIBOSWITCH

Title: The structure of the SAM/SAH-binding riboswitch

Authors: Weickhmann AK, Keller H, Wurm JP, Strebiter E, Juen MA, Kremser J, Weinberg Z, Kreutz C, Duchardt-Ferner E, Wöhnert J

Published in: Nucleic acids research, 47, 2654-2665

Contributions

(1) Entwicklung und Planung	Weickhmann AK (30%), Keller H (15%), Duchardt-Ferner E (25%), Wöhnert J (30%)
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Weickhmann AK: Klonierung, RNA-Präparation, NMR-Messungen, ITC-Messungen, Herstellung und Aufreinigung von markiertem SAM, Strukturrechnung; Keller H: initiale Klonierung, initiale RNA-Präparation, initiale NMR-Messungen; Wurm JP: Herstellung und Aufreinigung von markiertem SAM; Strebiter E: Herstellung selektiv markierter Proben; Juen MA: Herstellung selektiv markierter Proben; Kremser J: Herstellung selektiv markierter Proben; Weinberg Z: Bioinformatische Suche; Duchardt-Ferner E: NMR-Messungen, Strukturrechnung
(3) Erstellung der Datensammlung und Abbildungen	Weickhmann AK: Erstellung von Abbildungen, Datensammlung für PDB
(4) Analyse/ Interpretation der Daten	Weickhmann AK: Auswertung der NMR-Messungen, Auswertung der ITC-Messungen; Keller H: initiale NMR-Auswertungen und initiale ITC-Auswertungen; Duchardt-Ferner E: Auswertung der NMR-Messungen
(5) Übergeordnete Einleitung/ Ergebnisse/Diskussion	Weickhmann AK (40%), Weinberg Z (5%), Wöhnert J (55%)

The structure of the SAM/SAH-binding riboswitch

A. Katharina Weickhmann¹, Heiko Keller¹, Jan P. Wurm^{1,2}, Elisabeth Strebitzer³, Michael A. Juen³, Johannes Kremser³, Zasha Weinberg⁴, Christoph Kreutz³, Elke Duchardt-Ferner¹ and Jens Wöhnert^{1,*}

¹Institute for Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe-University Frankfurt, Max-von-Laue-Strasse 9, 60438 Frankfurt/M., Germany, ²Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31, 93053 Regensburg, Bavaria, Germany, ³Institute of Organic Chemistry, Centre for Molecular Biosciences (CMBI), University of Innsbruck, Innrain 80/82, 6020 Innsbruck, Austria and ⁴Bioinformatics Group, Department of Computer Science and Interdisciplinary Centre for Bioinformatics, Institute of Informatics, University of Leipzig, Härtelstrasse 16-18, 04107 Leipzig, Germany

Received October 09, 2018; Revised December 12, 2018; Editorial Decision December 13, 2018; Accepted December 26, 2018

ABSTRACT

S-adenosylmethionine (SAM) is a central metabolite since it is used as a methyl group donor in many different biochemical reactions. Many bacteria control intracellular SAM concentrations using riboswitch-based mechanisms. A number of structurally different riboswitch families specifically bind to SAM and mainly regulate the transcription or the translation of SAM-biosynthetic enzymes. In addition, a highly specific riboswitch class recognizes S-adenosylhomocysteine (SAH)—the product of SAM-dependent methyl group transfer reactions—and regulates enzymes responsible for SAH hydrolysis. High-resolution structures are available for many of these riboswitch classes and illustrate how they discriminate between the two structurally similar ligands SAM and SAH. The so-called SAM/SAH riboswitch class binds both ligands with similar affinities and is structurally not yet characterized. Here, we present a high-resolution nuclear magnetic resonance structure of a member of the SAM/SAH-riboswitch class in complex with SAH. Ligand binding induces pseudoknot formation and sequestration of the ribosome binding site. Thus, the SAM/SAH-riboswitches are translational ‘OFF’-switches. Our results establish a structural basis for the unusual bispecificity of this riboswitch class. In conjunction with genomic data our structure suggests that the SAM/SAH-riboswitches might be an evolutionary late invention and not a remnant of a primordial RNA-world as suggested for other riboswitches.

INTRODUCTION

S-adenosylmethionine (SAM, Figure 1A, top) is a central metabolite and a widely used co-substrate in many different enzymatic reactions. SAM is produced enzymatically from adenosine triphosphate (ATP) and methionine by SAM-synthetase. In the reaction catalyzed by this enzyme, the sulfur atom of methionine attacks the C5' of ATP and cleaves off its entire triphosphate chain while forming a new carbon-sulfur bond. SAM contains a positively charged sulfonium center substituted by a methyl, an aminocarboxypropyl and a 5'-deoxyadenosyl group. Each of these three functional groups can now be transferred from SAM to other substrates in SN₂ reactions (1). Most common is the transfer of the methyl group to substrates ranging from small molecules to protein sidechains, RNAs and DNAs catalyzed by a myriad of specialized methyltransferase enzymes (2). In these reactions, SAM is converted into the neutral compound S-adenosylhomocysteine (SAH, Figure 1A, bottom). SAH is toxic for the cell and rapidly broken down by SAH-nucleosidases into adenine and S-ribosylhomocysteine (3,4) or by SAH-hydrolases into homocysteine and adenosine (5,6). Less common is the use of SAM as a cosubstrate in reactions transferring the aminocarboxypropyl group to another substrate. Aminocarboxypropyl transfer reactions occur for instance during post-transcriptional modifications of rRNAs and tRNAs leaving 5'-deoxy-5'-methylthioadenosine as the second reaction product (7,8). Furthermore, in some metabolic pathways SAM is modified prior to any transfer reactions by SAM-decarboxylase (9,10). The aminopropyl group of the decarboxy-SAM produced by this enzyme is then used in the biosynthesis of spermine, spermidine and other polyamines (11). The transfer of the 5'-deoxyadenosyl group from SAM to a substrate is for instance observed in enzymatic fluorination reactions (12,13).

*To whom correspondence should be addressed. Tel: +49 69 79829785; Fax: +49 69 798 29527; Email: woehnert@bio.uni-frankfurt.de

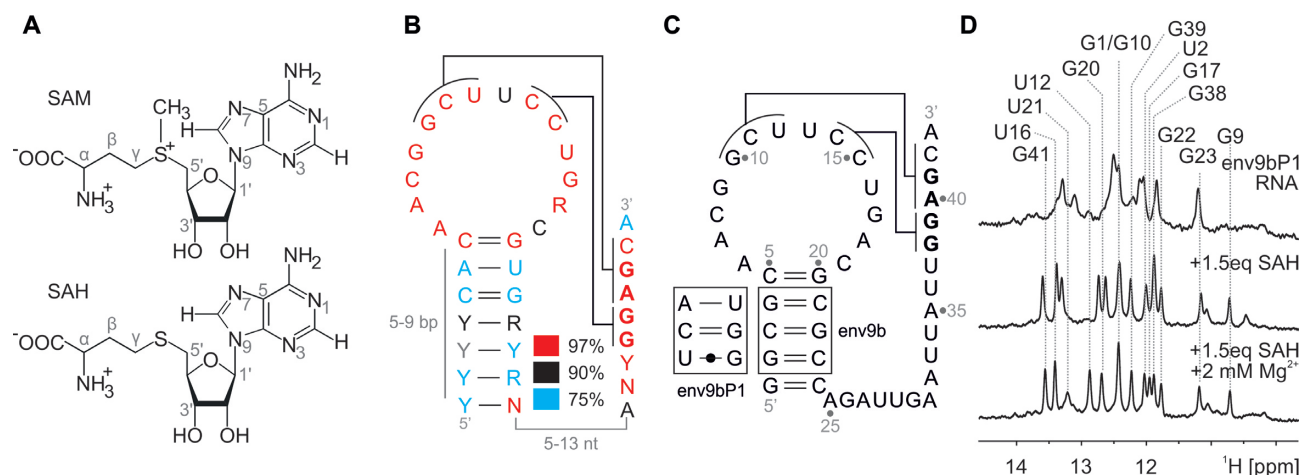


Figure 1. The SAM/SAH riboswitch and its ligands. (A) Chemical structures of SAM (top) and SAH (bottom). (B) Consensus sequence and secondary structure of the SAM/SAH-binding riboswitches. The ribosomal binding site is highlighted in bold letters. (C) Secondary structure of the SAM/SAH-binding riboswitch variant env9b and the P1 stem mutant env9bP1. The ribosomal binding site is highlighted in bold letters. The pseudoknot interactions indicated by connecting lines in (B) and (C) conform to the structure determined in this work and differ slightly from the original prediction. (D) Imino proton spectra of the env9bP1 RNA in the absence (top) and the presence of an 1.5-fold excess of SAH (middle) and 2 mM magnesium acetate (bottom). Resonance assignments in the SAH and Mg^{2+} bound state are indicated.

Due to the central importance of SAM-dependent reactions in metabolism, intracellular SAM concentrations and SAM biosynthesis are tightly regulated in all organisms. Transcription regulation mechanisms using a SAM-binding repressor protein (MetJ) are known for instance in *Escherichia coli* (14). However, many other bacteria use riboswitch-based transcriptional or translational control mechanisms to regulate SAM-biosynthesis (15,16). Riboswitches are structured domains in the 5'-UTRs of bacterial mRNAs that bind specifically to small molecules and change their structure upon ligand binding. These structural changes usually modulate either the stability of transcription termination signals or the accessibility of ribosome binding sites in these mRNAs. Depending on the structural properties of each riboswitch ligand binding can either activate ('ON'-switch) or inhibit ('OFF'-switch) transcription or translation (17).

SAM-sensing riboswitches are not only widespread in bacteria but also structurally highly diverse. Based on difference in sequence, secondary and tertiary structure they can be divided into three large superfamilies covering seven subclasses called SAM-I to SAM-VI and SAM-I/IV (15,16,18,19). Many SAM-riboswitches act as 'OFF'-switches for the expression of SAM-synthetase. While they differ in their tertiary structures and the architecture of their SAM-binding pockets, a common feature is the extensive amount of interactions between the RNA and the functional groups of the ligand SAM. The adenine moiety of SAM is involved in intermolecular base pairing or base-triplet interactions. The positively charged sulfonium group interacts electrostatically with polar carbonyl or hydroxyl groups of the RNA. Furthermore, in the SAM-I and SAM-II riboswitches, the carboxy and amino groups of the methionine side chain of SAM form hydrogen bonds to RNA nucleobases and/or the RNA backbone. Due to this extensive set of interactions between the ligand and the RNA, all SAM-riboswitches are highly specific for SAM.

In particular, these riboswitches discriminate strongly between SAM and SAH and bind SAM with a 100- to 1000-fold higher affinity compared to SAH. Interestingly, there is a specialized SAH-sensing riboswitch with the opposite pattern of specificity (20). It binds SAH with an ~1000-fold higher affinity compared to SAM and acts as an 'ON'-switch for the expression of SAH-hydrolase—an enzyme responsible for the breakdown of the toxic SAH. The high specificity of these riboswitches and their ability to discriminate strongly between SAM and SAH prevents unwanted regulatory cross talk between those two structurally very similar but functionally very different metabolites. This is an important feature for the regulatory logic in the associated metabolic pathways.

Defying these considerations, members of a novel class of riboswitches discovered in 2010 in the order Rhodobacterales of the α -proteobacteria bind SAM and SAH with similar affinities (21). Accordingly, this class of riboswitches was named the SAM/SAH-family of riboswitches. However, virtually all of its members were found in the 5'-UTRs of SAM-synthetase (MetK) mRNAs and were proposed to act as translational 'OFF'-switches. While it is immediately obvious, that translational repression of the SAM-synthetase upon SAM-binding to this riboswitch is a regulatory useful response, it is less obvious why these riboswitches should be able to bind to the breakdown product of SAM—SAH—and thereby shut off the expression of SAM synthetase. Sequence comparisons suggested a pseudoknot secondary structure for this riboswitch (Figure 1B). A 5'-hairpin loop structure contains a P1 stem with 5–9 bp and a formally single stranded loop region with 14 nt. A total of 12 of the 14 nt in the loop region are strictly conserved. The hairpin loop element is followed by a putatively single-stranded linker element of 5–13 nt of variable sequence and 8 conserved nucleotides that include the ribosome binding site. Some of these conserved nucleotides are complementary to nucleotides in the loop region of the 5'-hairpin al-

lowing the formation of the P2 stem of a putative H-type pseudoknot structure (Figure 1B). The three-dimensional structure of this riboswitch and its ligand recognition mode have not yet been determined.

Here, we present the high-resolution nuclear magnetic resonance (NMR) solution structure of the SAM/SAH-riboswitch bound to SAH. The structure rationalizes why this riboswitch in contrast to the classical SAM riboswitches binds SAM and SAH with similar affinity. We also show that ligand binding promotes the formation of an H-type pseudoknot structure that sequesters the ribosome-binding site and thereby induces ligand-dependent translational repression. Ligand binding also supports additional base pairing interactions in the loop region and between the hairpin loop and conserved nucleotides in the 3'-single-stranded region. SAM and SAH are bound in an intermolecular *trans* Hoogsteen/Watson-Crick ('reversed Hoogsteen') base pairing interaction between the adenosyl moiety of the cofactor and a strictly conserved U nucleotide of the riboswitch. Contacts between the sulfur atom of the cofactor and the RNA are limited to only a single polar interaction and therefore less extensive than in the other SAM-binding riboswitch families. Only a single electrostatic interaction is found between the amino group of the aminocarboxypropyl side chain of the ligand and the RNA. Overall, our structure and additional extensive thermodynamic data obtained for mutant RNAs and ligand analogs suggest that the SAM/SAH-riboswitch represents a structurally minimalistic solution for SAM recognition. Taken together with the poor ability to discriminate against SAH, its gene associations and its restricted occurrence in the Rhodobacterales suggest that the SAM/SAH riboswitch evolved in organisms that already contained an SAH-degrading enzyme or other mechanisms that keep SAH concentrations low enough to prevent interference with SAM-dependent riboswitch signaling.

MATERIALS AND METHODS

RNA preparation

RNAs were synthesized by *in vitro* transcription with T7 RNA polymerase and a linearized plasmid template DNA. The *in vitro* transcription mix contained either unlabeled (Sigma-Aldrich) or appropriately isotope-labeled nucleoside triphosphates (NTPs) (Silantes GmbH and Cambridge Isotope Laboratories, Inc.). To obtain RNA with a homogeneous 3'-end, the primary RNA-transcript contained a hepatitis delta virus ribozyme fused to the 3'-end of the target RNA. The processed RNA was purified as described (22). The purified RNAs were folded by heating to 80°C for 10 min followed by a rapid 5-fold dilution with ice cold water. The samples were exchanged into 25 mM potassium phosphate buffer (pH 6.2) containing 50 mM potassium chloride in multiple cycles of concentration and dilution in buffer using centrifugation devices (VivaSpin 2, MWC 3 kDa). RNA concentrations in the NMR samples ranged between 0.04 and 1.2 mM containing either 5–10% or 100% (v/v) D₂O, 2 mM magnesium acetate and up to four equivalents of ligand.

For assignment and structure calculation experiments samples containing unlabeled, uniformly ¹⁵N-, uniformly

¹³C,¹⁵N-, uniformly 5-D₁/ribose-3',4',5',5''-D₄-labeled and three different nucleotide-type selectively labeled samples (¹³C,¹⁵N-A/C, ¹³C,¹⁵N-G or ¹³C,¹⁵N-U) of the env9bP1 riboswitch (Figure 1C) in complex with either unlabeled or ¹³C,¹⁵N-labeled SAM or SAH were used. Furthermore, site and atom-specifically labeled samples (**env9b**: ¹⁵N₂-U12, ¹⁵N₂-U16, ¹⁵N₂-U37, ¹³C-C8-A7/¹³C-C6-U12, ¹³C-C8-G10/¹³C-C6-U16, ¹³C-C8-A6/¹³C-C6-C15, ¹³C-C6-C11/¹³C-C8-G17, ¹³C-C8-G9/¹³C-C6-C14; **env9bP1**: ¹³C-C6-U13/¹³C-C6-C19, ¹³C-C6-C8/¹³C-C8-A18, ¹³C-C8-A35/¹³C-C6-U36, ¹³C-C2,C8-A35/¹³C-C6-U37, ¹³C-C6-U34/¹³C-C2,C8-A35) were chemically synthesized (23–25). Pde-1-1 RNA was purchased commercially (Thermo Scientific) and deprotected according to the manufacturer's instructions.

Unlabeled SAM, SAH, adenosine, S-(5'-adenosyl)-3-thiopropylamine (decarboxy-SAH, Sigma-Aldrich), 5'-deoxy-5'-methylthioadenosine (Sigma-Aldrich) and S-(5'-adenosyl)-3-methylthiopropylamine (decarboxy-SAM, TRC, Canada) were purchased commercially and dissolved in water or NMR buffer, respectively.

Synthesis and purification of ¹³C,¹⁵N-labeled SAM and SAH

¹³C,¹⁵N-labeled SAM was enzymatically synthesized as described earlier (26) starting from ¹³C,¹⁵N-labeled methionine and ¹³C,¹⁵N-labeled ATP and purified using cation exchange chromatography.

¹³C,¹⁵N-labeled SAM was then incubated with the methyltransferase PaMTH1 from *Pseudospira anserina* and its substrate myricetin yielding ¹³C,¹⁵N-labeled SAH (27) which was further purified using RP-HPLC. Product purity was confirmed by NMR and thin-layer chromatography using ethanol/H₂O/acetic acid (65/34/1) as running buffer. The synthesis pathway for SAM and SAH is shown in Supplementary Figure S1.

NMR spectroscopy

All NMR experiments were carried out using Bruker AVANCE 600, 700, 800, 900 and 950 MHz spectrometers equipped with cryogenic triple resonance HCN- or HCP-probes and z-axis gradients. NMR experiments for resonance assignments have been described in detail before (28). NMR spectra were recorded at 5°C, 10°C and 20°C in 5–10% (v/v) D₂O for the detection of exchangeable protons and at 20°C in 100% D₂O for the non-exchangeable protons.

The data were processed using TOPSPIN 3.2 or 3.5 (Bruker BioSpin, Germany) and analyzed using CARA (29). Proton chemical shifts were referenced directly to 4,4-dimethyl-4-silapentane-1-sulfonic acid. Carbon, nitrogen and phosphorous chemical shifts were referenced indirectly (30).

Input restraints for structure calculations

All intra-, as well as intermolecular ¹H,¹H distance restraints were obtained from NOE intensities of 2D ¹H,¹H-NOESY, 2D ¹⁵N-CPMG-NOESY (31,32) and 2D or 3D ¹H,¹³C-NOESY-HSQC (optimized for either aromatic or

aliphatic CH moieties) experiments. Experiments optimized for aromatic CH moieties were carried out in 5–10% (v/v) D₂O, while those optimized for aliphatic CH moieties were recorded in 100% (v/v) D₂O. All NOESY experiments were performed with a mixing time of 120 ms. Hydrogen bonds between imino or amino hydrogen groups as donors and nitrogen acceptors were identified experimentally using HNN-COSY, amino HNN-COSY, long-range HNN-COSY and sellr-HNN-COSY experiments using appropriately labeled samples (33–37).

For the determination of the ribose sugar pucker, a TOCSY experiment with a mixing time of 20 ms was recorded on a uniformly 5-D₁/ribose-3',4',5',5''-D₄-labeled sample in 100% (v/v) D₂O. Nucleotides that showed a significant H1'-H2' TOCSY cross peaks were restraint to a C2'-endo conformation.

Structure calculation

All extracted NOE cross-peak intensities were referenced to the averaged intranucleotide H5/H6 NOEs in pyrimidine residues (2.4 Å). For spectra without H5/H6 NOE cross peaks, referencing was performed by matching the peak intensities to the corresponding intensities of identical NOEs in reference spectra. 1.9 Å were added to all upper limit distance restraints. They were classified in three distinct groups corresponding to weak, intermediate and strong NOEs. The geometry of canonical base pairs identified in the HNN-COSY experiments was constrained by introducing four upper and four lower distance restraints for A:U and G:U base pairs as well as for the U16:SAH intermolecular base pair and six upper and six lower distance restraints for G-C base pairs. All additional experimentally detected hydrogen bonds were introduced using two upper and two lower distance restraints. For residues in canonical base pairs (nucleotides 1–5 and 20–24 in P1 as well as 10–14 and 38–42 in the pseudoknot) the backbone torsion angles α , β , γ , δ , ϵ and ζ were set to standard A-form helical values ($\pm 20^\circ$). The ribose conformation of nucleotides 2–4 and 21–23 in P1 and 11–12 and 40–41 in P2 was constrained to C3'-endo. Residues G9, A18 and U12 were constrained to a C2'-endo ribose conformation. The ribose conformations of the flexible linker residues as well as U13 and U37 were left unconstrained. The glycosidic torsion angle χ for the bases in the A-form helical parts was set to restrict the nucleobases in an *anti*-conformation as was the torsion angle χ for A6 due to the characteristic NOE pattern in this region.

The structures were calculated using CYANA version 3.97. PyMol (DeLano Scientific LLC, Schrödinger Inc.) or MOLMOL (38) was used for structure visualization.

Isothermal titration calorimetry (ITC)

All isothermal titration calorimetry (ITC) measurements were performed using a MicroCal iTC₂₀₀ instrument (Malvern Pananalytical, UK) at 25°C. Samples of unlabeled RNAs and all ligands were prepared in 25 mM potassium phosphate buffer (pH 6.2) containing 50 mM potassium chloride and 2 mM magnesium acetate. The ligand (400–4000 μ M) was injected into a solution of 40–100 μ M RNA. After an initial delay of 120 s, the first injection of 0.2 μ l

was followed by 19 serial injections of 2 μ l, separated by intervals of 180 s using a reference power of 11 μ cal⁻¹ and a stirring speed of 750 rpm with high feedback mode. The titration of the ligand into the buffer solution resulted in negligible evolution of heat. Three independent titrations for env9b-derived RNAs and two independent titrations for Pde-1-1 or SK209-52 RNAs were performed for each ligand and the reported K_D values are the average from these titrations. The thermograms were processed using Origin 7.0 (OriginLab) assuming a one-site binding model. The first injection was excluded from the analysis. An overview of all binding parameters derived from the ITC measurements is given in Supplementary Table S1.

Bioinformatics

Additional SAM/SAH riboswitches were found using Infernal version 1.1 (39) to search the RefSeq nucleotide database version 72 (40) and environmental sequences essentially as previously described (39). We analyzed the occurrence of enzymes in different species using KEGG (41), using nine completely sequenced organisms with SAM/SAH riboswitches that are also annotated by KEGG (see Supplementary Table S2). We did not consider organisms with incomplete genomes (i.e. whose RefSeq accessions did not start with 'NC'), since apparently absent genes could be located in the missing parts of the genome.

RESULTS AND DISCUSSION

NMR-spectroscopy of the SAM/SAH-riboswitch

In order to find SAM/SAH riboswitch candidates suitable for high-resolution structure determination we screened sequences from different organisms in NMR-based titration experiments with SAM and/or SAH at different temperatures and Mg²⁺ concentrations. Changes in imino proton 1D ¹H-NMR spectra upon ligand addition as well as signal dispersion and signal line widths served as selection criteria. The spectra of two 43-nt RNA variants (env9b, env9bP1, Figure 1C) derived from a metagenomic sequence in an environmental sample called env9 (21) showed large spectral changes as well as well-dispersed imino proton signals with narrow line widths upon SAH addition (Figure 1D and Supplementary Figure S2). Spectral quality increased even further upon addition of Mg²⁺ (Figure 1D). Thus, this RNA is capable of binding SAH already in the absence of Mg²⁺. The imino proton spectrum of the free RNA showed only a small number of broadened signals suggesting that in the absence of the ligand the RNA is only partially structured and interconverting between different conformations (Figure 1D). This is in agreement with the previous proposal (21) that the ligand-free state of this riboswitch corresponds to the 'ON'-state where the ribosome-binding site is accessible to the ribosome for ligand binding. Addition of SAM lead to spectral changes in the RNA imino proton 1D-¹H spectra very similar to those observed upon SAH binding. Thus, this riboswitch binds indeed to both ligands. Furthermore, SAM and SAH induce the folding of this RNA into very similar structures. This was verified by the comparison of the ¹⁵N-HSQC-spectra of both complexes (Figure 2A). ITC showed that the env9bP1 ri-

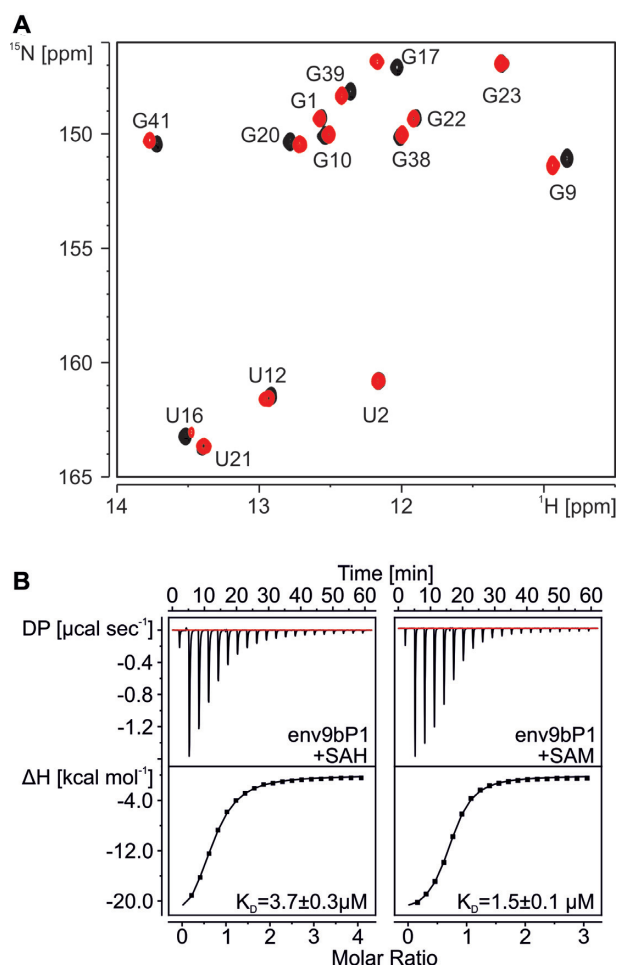


Figure 2. The env9 SAM/SAH-riboswitch binds SAM and SAH with a similar binding mode and with similar affinities. (A) Overlay of the ^{15}N -HSQC spectra of the imino region of uniformly ^{15}N -labeled env9bP1 RNA in complex with SAH (black) and SAM (red) recorded at 20°C in the presence of 2 mM magnesium acetate. NMR signal assignments are indicated. (B) Representative ITC thermograms for the env9bP1 riboswitch titrated with SAH (left) or SAM (right) in the presence of 2 mM magnesium acetate. The resulting K_D values are indicated.

riboswitch bound SAM ($K_D = 1.5 \mu\text{M}$) with a slightly higher affinity than SAH ($K_D = 3.7 \mu\text{M}$) in the presence of 2 mM MgCl_2 (Figure 2B). In the absence of Mg^{2+} the affinity of the RNA for both ligands is significantly diminished (Supplementary Figure S2). Despite its slightly lower affinity, we chose to determine the NMR-solution structure of the SAH-RNA complex in the presence of 2 mM magnesium acetate since NMR-experiments for the collection of structural constraints can require multiple days of measurement time and SAM is unstable at room temperature on this time scale. For NMR-signal assignment purposes and to minimize NMR-signal overlap we used both the env9b and the env9bP1 variant of the riboswitch. Since the slight sequence difference in the P1 stem region (Figure 1C) had no consequences for ligand binding and RNA tertiary structure (Figure 2 and Supplementary Figure S2), spectra recorded for both constructs could be used to derive structural constraints (Supplementary Table S1).

Table 1. Structural statistics for the SAM/SAH-binding riboswitch in complex with SAH (PDB entry: 6HAG)

NMR restraints	
Total distance restraints	743
RNA	
Intra-residue	234
Sequential	184
Long-range	166
Hydrogen-bond	76
RNA-SAH	
NOE	79
Hydrogen-bond	4
Total dihedral restraints	168
Ribose pucker	57
Backbone	111
Structural statistics	
Heavy atom RMSD to mean structure (\AA) (U2-U12, C14-G23, A35-U36, G38-C42, SAH)	0.58 ± 0.15

Importantly, the quality of the fingerprint ^{15}N -HSQC NMR-spectrum for the SAH-RNA complex was similar to that recorded for the SAM-RNA complex (Figure 2A). Due to significant structural deviations from A-form double-helical RNA, NMR signal assignments had to rely heavily on through-bond magnetization transfer experiments for the identification of base spin systems (42-46) and for connecting base and sugar spin systems (47,48) as well as on the use of a multitude of chemically synthesized site-specifically labeled RNAs (23,25) as described in detail previously (28). In order to collect a large number of NOE-based distance restraints as input for NMR-structure calculations we used base-type selectively labeled samples as well as a sample prepared from partially deuterated nucleotides (5-D₁/ribose-3',4',5',5''-D₄-labeled (49)). Furthermore, we employed variants of the HNN-COSY experiments optimized for different types of hydrogen bond donor and acceptor groups on both uniformly and base-type selectively ^{15}N -labeled samples in order to delineate the hydrogen bond network in the RNA-SAH complex (36,37). The final NMR-structure was calculated from 663 NOE-based distance restraints, 168 torsion angle restraints and 80 hydrogen bond restraints. The overall heavy atom RMSD for the 10 structures with the lowest CYANA-target function is $0.58 \pm 0.15 \text{\AA}$ when including all structurally well-defined nucleotides (Table 1).

Overall structure of the SAM/SAH-riboswitch in the ligand bound state

The structure of the SAM/SAH-riboswitch bound to SAH is shown in Figure 3A. As predicted from sequence comparisons, the RNA in the complex forms an H-type pseudoknot with the P2 stem consisting of base pairs between nucleotides of the apical loop of the 5' hairpin and nucleotides stemming from the conserved region in the 3'-single stranded region (Figures 1C, 3A and B). However, the base pairing interactions in P2 differ from the earlier prediction (21). The previously proposed secondary structure model predicted a continuous P2 helix consisting of nucleotides C11 to C14 (env9b numbering) in the 5'-hairpin loop and nucleotides 38 and 41 from the 3'-single stranded region. In our structure, we observe base pairing between

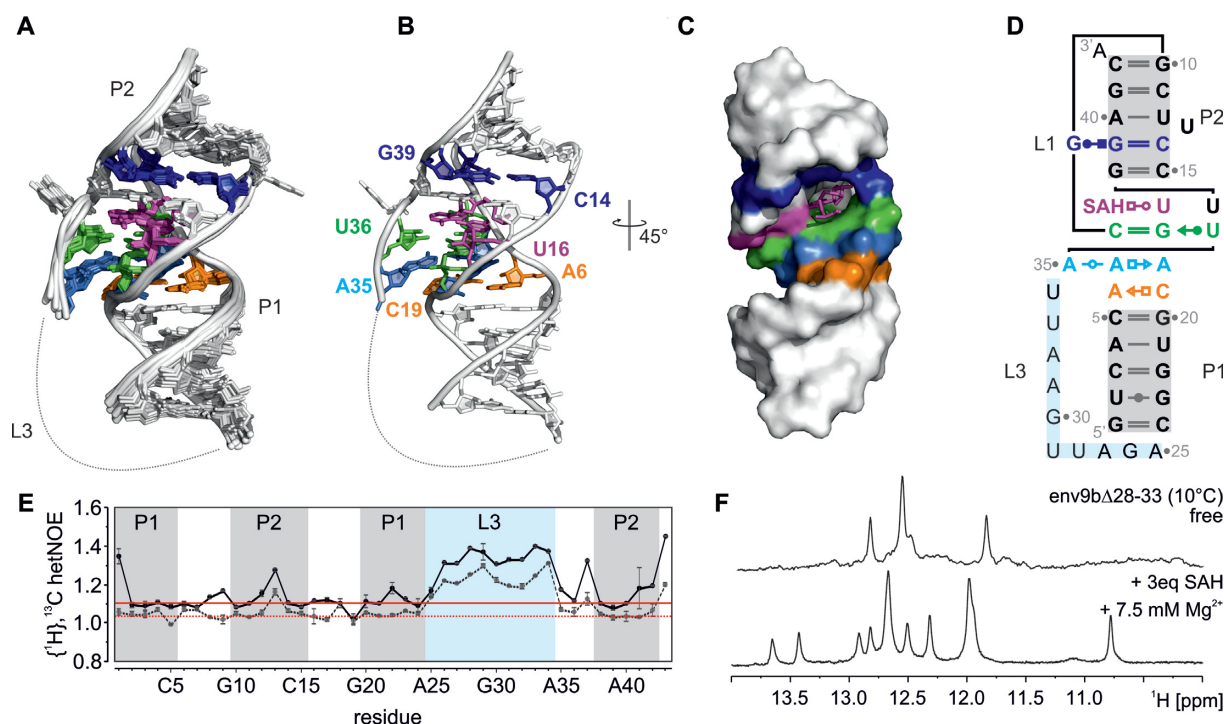


Figure 3. Overall structure of the SAM/SAH-riboswitch in complex with SAH. (A) Superimposition of the ten riboswitch/SAH complex structures with the lowest CYANA target function. The base triplet between the bases G9, C14 and G39 is shown in dark blue, U16 and SAH in magenta, the base triplet C8:G17:U36 in green, the base triplet of A6:A18:A35 in light blue and A6 and C19 in orange. (B) Average structure of the riboswitch/SAH complex with the same coloring as used in (A). (C) Surface representation of the riboswitch RNA and SAH in stick representation with the same coloring as in (A). (D) Structure diagram of the complex structure using the Leontis-Westhof notation. (E) $\{^1\text{H}\}, ^{13}\text{C}$ hetNOE values for the aliphatic H1'/C1' (solid black lines) and aromatic H8/C8- and H6/C6- (dashed gray lines) moieties along the sequence of the riboswitch in the SAH-bound state. The average hetNOE-values of the helical residues for aliphatic H1'/C1' aromatic H6/C6 and H8/C8 groups are marked by a solid and a dashed red line, respectively. The different structural regions of the riboswitch are indicated by different background colors. (F) Imino proton spectra of the mutant *env9bΔ28-33* in the ligand-free state (top) and in the presence of 3-fold excess of SAH and 7.5 mM magnesium acetate (bottom) at 10°C.

nucleotides G10, C11, U12, C14 and C15 from the hairpin loop and the G38-G39-A40-G41-C42 sequence of the 3'-single stranded region (Figures 1C and 3D). All nucleotides involved in pseudoknot-helix base pairing are strongly conserved (Figure 1B). Nucleotide U13 is looped-out from the pseudoknot helix and is not conserved (Figure 1B). It is conformationally flexible as seen from elevated $\{^1\text{H}\}, ^{13}\text{C}$ -hetNOEs for its H6/C6, H8/C8 and H1'/C1' moieties (Figure 3E) and induces a slight kink in the P2 helix (Figure 3A and C). Replacement of this nucleotide with A or C preserved the ligand binding capabilities of the RNA while its deletion lead to a complete loss of ligand binding (Supplementary Figure S3). Compared to the previously predicted secondary structure the P1 helix of the RNA is extended by two non-Watson-Crick base pairs—A6:C19 and A7:A18—and a Watson-Crick base pair between C8 and G17 (Figure 3). G9 formally corresponds to loop 1 (L1) of the pseudoknot and spans the major groove of the entire pseudoknot P2 helix (Figure 3). U16 on the other side of the hairpin loop represents loop 2 (L2) of the pseudoknot. It is not base paired with an RNA-nucleotide but instead binds the adenosyl moiety of SAH (see below). Formally, loop 3 (L3) of the pseudoknot consists of nucleotides 25–37 and crosses the minor groove of the P1 stem (Figure 3). $\{^1\text{H}\}, ^{13}\text{C}$ -hetNOE-measurements (Figure 3E) show that nucleotides 26–34 of L3 are highly flexible and not involved

in stabilizing interactions with the remainder of the RNA-structure. An RNA with a significantly shortened linker (*env9bΔ28-33*) where nucleotides 28–33 were deleted is still able to bind the ligand (Supplementary Figure S4) and folds into a wild-type (WT)-like structure demonstrating that these nucleotides are not important for the overall structure (Figure 3F). Nucleotide 35 of the linker is strictly conserved as an A and is conformationally rigid (Figure 3E). It stabilizes the non-canonical P1 base pair A7:A18 by forming a minor groove base triple (see below). Nucleotide U36 stabilizes the closing base pair of the P1 stem C8:G17 by forming another minor groove base triple (see below). Nucleotide U37 is flexible according to the $\{^1\text{H}\}, ^{13}\text{C}$ -HetNOE-data (Figure 3E) and does not interact with other parts of the RNA. It serves as a spacer for creating the SAM-binding pocket. Overall, the structure of the SAM/SAH riboswitch in complex with its ligand is significantly less compact compared to the structures of other riboswitches adopting similar H-type pseudoknot structures upon ligand binding. In the SAM-II riboswitch, the tertiary interactions between linker nucleotides in L1 (7 nt) and L3 (11 nt) and the remainder of the RNA are much more pronounced (50). Nucleotides in L1 are an integral part of the ligand binding site whereas almost all nucleotides in L3 form stabilizing tertiary interactions with the major groove of the P1 stem. Only one nucleotide in L1 and three nucleotides in L3 are

flexible in this structure. Similarly, in the H-type pseudoknot structure adopted by the preQ1 riboswitch, all nucleotides of L1 (2 nt) and L3 (6nt) are involved in tertiary interactions (51).

Ligand recognition and the ligand binding site

The ligand binding site of the SAM/SAH riboswitch is sandwiched between the C8:G17 closing base pair of the P1 stem and the C15:G38 closing base pair of the pseudoknot P2 stem (Figure 3). The adenosyl moiety of the ligand forms an intermolecular base pair with the strictly conserved U16 nucleotide of the riboswitch. An HNN-COSY-spectrum recorded for a selectively ^{15}N -U-labeled RNA bound to $^{13}\text{C}^{15}\text{N}$ -labeled SAH showed directly that the U16 imino group is hydrogen bonded to the N7 of the ligand (Figure 4A) and together with intermolecular NOEs established a *trans* Hoogsteen/Watson–Crick (‘reversed Hoogsteen’) geometry for the SAH:U16 intermolecular base pair (Figure 4B). A U16A mutant of the riboswitch is not capable of binding SAH anymore (Supplementary Figure S5). Unambiguous NMR resonance assignments of the U16 spin system were further supported by the use of a chemically synthesized U16 N3- ^{15}N and a U16 C6- ^{13}C selectively labeled sample (Supplementary Figure S5). A *trans* Hoogsteen/Watson–Crick base pairing interaction between the adenosyl moiety of the ligand and an U residue of the riboswitch structurally similar to the one observed here is also found in the SAM-II riboswitch (50). However, in this case the U residue binding the ligand base moiety is further stabilized by additional hydrogen bonding interactions with another U residue of the riboswitch. In contrast, in the SAM-I riboswitch the adenine base of SAM forms a *cis* Hoogsteen/Watson–Crick base pair with an U of the riboswitch (16).

The formation of an intermolecular base pair between U16 and SAH establishes continuous stacking between the P1 and the P2 stems of the pseudoknot (see Figure 3). The adenosyl moiety of SAH stacks with the bases of G38 from the C15:G38 base pair closing P2 on one side and of G17 from the C8:G17 closing base pair of P1 on the other side (Figure 3). The sulfur atom from SAH is in an appropriate distance ($3.11 \pm 0.03 \text{ \AA}$) and orientation to form a chalcogen bond (52) with the O6 carbonyl oxygen of G38 (Figure 4C). Notably, in a SAM-bound complex the sulfur atom would carry an additional positive charge strengthening this interaction due to the negative partial charge of G38 O6. The ribose hydroxyl groups of SAH also interact with the RNA. The 2'-OH group of the ligand is in hydrogen bonding distance to the 5'-phosphate group of G38 (average O–O distance for 10 best structures $3.1 \pm 0.4 \text{ \AA}$) (Figure 4C). The 3'-OH group of SAH is in hydrogen bonding distance to the O4' of the ribose ring in G9 as well as to the 2'-OH-group of C8 in P1 ($3.0 \pm 0.2 \text{ \AA}$ and $3.2 \pm 0.1 \text{ \AA}$).

The C8:G17 base pair at the lower base of the ligand binding site is further stabilized through the formation of a minor groove base triple with U36 from L1 (Figure 4D). The U36 C2 carbonyl group is a hydrogen bond acceptor for the G17 amino group (Figure 4D). The 2'-OH group of U36 acts as a hydrogen bond acceptor for the amino group of SAH (Figure 4E). A U36C mutant RNA that keeps these

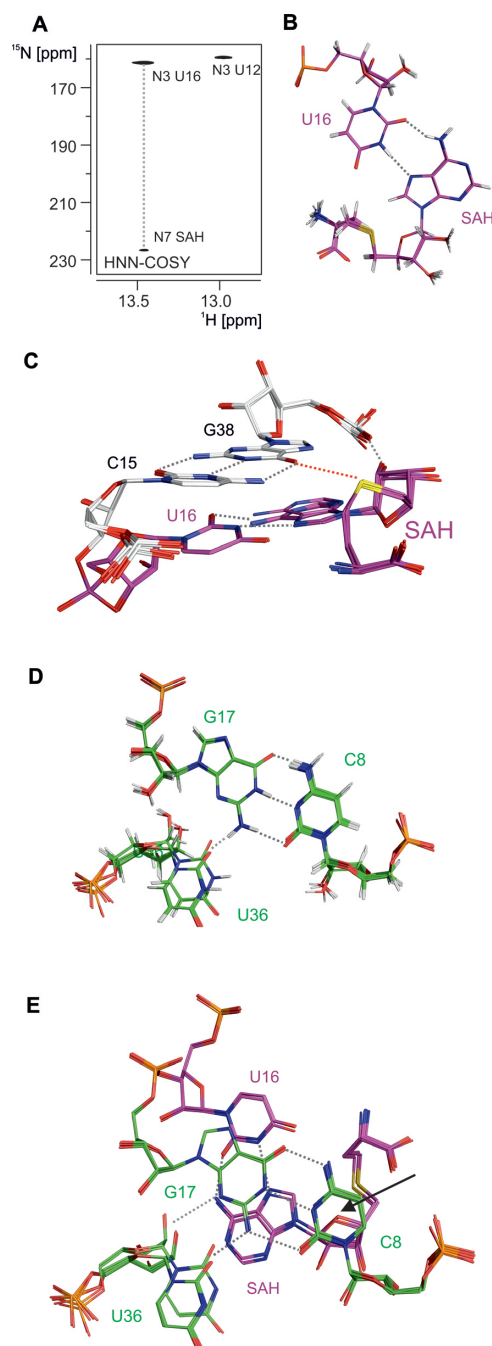


Figure 4. The SAH binding site. (A) HNN-COSY experiment on a ^{15}N -U-labeled RNA in complex with $^{13}\text{C}^{15}\text{N}$ -labeled SAH showing a correlation between the U16 imino group and the N7 nitrogen of SAH. (B) SAH recognition in a ‘reversed Hoogsteen’ (*trans* Hoogsteen/Watson–Crick) base pair between SAH and U16 in a bundle representation. Hydrogen bonds are indicated by dashed lines. (C) Interactions between the sulfur atom and the ribose moiety of SAH with the C15:G38 base pair forming the ‘roof’ of the SAH-binding site. Dashed gray lines indicate hydrogen bonds and a dashed red line indicates the putative chalcogen bond between U16 and SAH. (D) The C8:G17:U36 minor groove base triplet (carbon atoms in green) forming the floor of the SAH binding pocket in a bundle representation. (E) Stacking interactions between the U16:SAH intermolecular base pair (carbon atoms in magenta) and the C8:G17:U36 (carbon atoms in green) minor groove base triplet. Hydrogen bonds are indicated by dashed lines. A black arrow points toward the O4' oxygen of the SAH ribose group located atop the base of C8 in an orientation suitable for a lone pair– π interaction.

hydrogen bonding interactions intact binds SAM and SAH with the same affinity as the WT (Supplementary Figure S6 and Table S1). Furthermore, the base of C8 is packed against the ribose moiety of SAH (Figure 4E). The SAH ribose O4' oxygen is in an orientation suitable for a lone pair- π interaction with the C8 base (3.34 ± 0.08 Å) (53,54). At last, the positively charged amino group of the aminocarboxypropyl group of SAH is in a distance range (4.2 ± 0.1 Å) for an electrostatic interaction with O4 of U16 (Figure 4E).

A major groove base triple between L1 and helix P2 of the pseudoknot

In the immediate vicinity of the binding pocket, the structure of the SAH:RNA complex is stabilized by additional tertiary interactions (Figure 5A). Directly above the 'roof' of the SAH binding pocket formed by C15:G38, the closing base pair of P2, there is a major groove base triple consisting of the C14:G39 Watson-Crick base pair and G9, which formally represents L1 of the pseudoknot (Figure 5B). The amino group of G9 is hydrogen bonded to the N7 of G39 and the G9 imino group is hydrogen bonded to the O6 of G39 establishing a *cis* Watson-Crick/Hoogsteen base pairing interaction between G9 and G39. In addition, the G9 O6 serves as a hydrogen bond acceptor for the C14 amino group (Figure 5B). The geometry of this base triple was established based on a large number of NOE contacts involving the G9 imino group (Figure 5C). Its proton chemical shift of 10.8 ppm is in agreement with an oxygen containing group as a hydrogen bond acceptor. An RNA with a G39C/C14G mutation that should be able to form a stable P2 Watson-Crick base pair but no hydrogen bonds with G9 is not capable of ligand binding (Figure 5D).

Stabilizing tertiary interactions between residues of L3 and the extended P1 helix

Below the 'floor' of the SAH-binding pocket, which consists of the minor groove base triple C8:G17:U36 (see above), lies another base triple (Figure 6A). It involves base pairing between A7 and A18 as well as between A7 and A35. A7 and A18 are part of the extended stem P1 whereas A35 stems from the L3 loop. A7 and A18 interact with each other in a *trans* Hoogsteen/Sugar-edge base pairing interaction stabilized via a single hydrogen bond between the A7 amino group and the N3 nitrogen of A18 (Figure 6B). Two hydrogen bonds are present between the amino groups of A7 and A35 and their respective N1 nitrogen atoms establishing a *trans* Watson-Crick/Watson-Crick base pair (Figure 6B). The geometry of this base triplet is supported by a dense network of NOE-constraints (Figure 6C). The *trans* Hoogsteen/Sugar-edge base pair A7:A18 is isosteric to the classical 'sheared' G:A base pair (a *trans* Sugar-edge/Hoogsteen G:A base pair in the Westhof-Leontis nomenclature, Figure 6D). Interestingly, the consensus sequence of the SAM/SAH-riboswitch shows that position 18 can be either A or G whereas in position 7 A is strictly conserved (Figure 1B). A 'sheared' A7:G18 in this position would be stabilized by two hydrogen bonds—one between the A N6 amino group and the G N3 nitrogen and the other

between the G N2 amino group and the A N7 nitrogen (Figure 6D). In line with this idea, an RNA containing an A18G mutation not only bound SAM and SAH with a structure very similar to the WT (Figure 6D) but its affinity for both SAM and SAH was increased compared to the WT (Figure 6E). The ^{15}N -HSQC spectrum of this mutant reveals an additional G imino group signal with a chemical shift of 9.3 ppm (Figure 6D) which is in line with the presence of a 'sheared' A:G base pair where the G imino group has no direct hydrogen bond acceptor (55).

Right underneath this base triple A6 and C19 form a base pair with a very unusual geometry (Figure 7A). Here, the amino group of C19 is involved in two hydrogen bonds with the N3 nitrogen and the 2'-OH group of A6 as acceptors (Figure 7B). The peculiar geometry of this base pair is supported by a cross-hydrogen bond scalar coupling between the C19 amino group resonances and the A6 N3 nitrogen observed in an amino group optimized HNN-COSY experiment (Figure 7C). Furthermore, a very strong NOE cross peak between the H5 resonance of C15 and the H2 resonance of A6 unequivocally confirms the conformation of this base pair (Figure 7D). Additional NMR data illustrating the signal assignment process for the A6 and C19 spin systems as well as additional evidence for the A6:C19 base pairing interaction are shown in the Supplementary material (Supplementary Figure S7). Both A6 and C19 are strictly conserved in the consensus sequence of the SAM/SAH riboswitches suggesting an important functional role of this unique A:C mismatch. However, the A6:C19 base pair is not directly involved in ligand binding or additional tertiary interactions (Figure 3). We tested the functional importance of its unusual geometry by replacing A6 with G thereby enforcing the formation of a normal stable G6:C19 Watson-Crick base pair at this position. The mutant RNA binds SAM with a significantly reduced affinity as reflected by the 27-fold higher K_D compared to the WT (Figure 7E). Apparently, one important structural consequence of the presence of the adjacent A6:C19 and A7:A18 base pairs in the extension of the P1 stem is a significant deviation from the typical A-form helical intrastrand base stacking interactions. In particular, the P1 helix is strongly underwound at the A6/A7 helical step. This departure from standard A-form geometry, however, enables the placement of C8 at the floor of the ligand binding pocket.

Binding of SAM/SAH analogs to the riboswitch

Overall, our structure of the SAM/SAH riboswitch RNA bound to SAH establishes a mode of ligand recognition, which relies on a more limited number of riboswitch-ligand interactions compared to the highly specific SAM riboswitches, and explains why both SAM and SAH bind with rather similar affinities. To gain further insights into ligand recognition and the importance of different functional groups in the ligand we tested different ligand analogs for binding. To assess the contribution of the aminocarboxypropyl side chain of SAH to riboswitch binding, we measured the affinity of 5'-deoxy-5'-methylthioadenosine (Supplementary Figure S8 and Table S1). 5'-deoxy-5'-methylthioadenosine bound the riboswitch with a 6-fold reduced affinity compared to SAH in line with the loss of

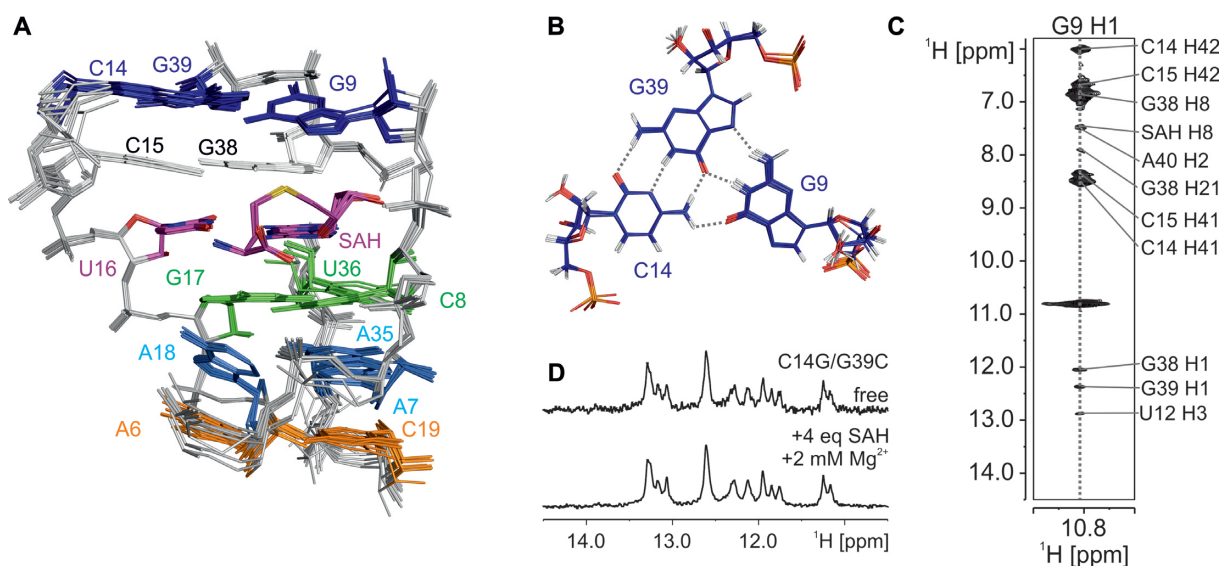


Figure 5. Additional tertiary interactions in the vicinity of the ligand binding pocket. (A) Tertiary structure elements surrounding the intermolecular U16:SAH base pair (magenta). The Watson-Crick base pair C15:G38 at the base of stem P2 is shown in white and the base triplet C8:G17:U36 closing P1 is shown in green. Directly underneath this base triplet is the A7:A18:A35 base triplet (light blue) and the A6:C29 base pair (orange). A major groove base triplet G9:C14:G39 (dark blue) is located directly above the P2 closing base pair C15:G38 (white). (B) The G9:C14:G39 base triplet (dark blue) in a bundle representation. (C) Strip from the $^1\text{H}, ^1\text{H}$ -NOESY at the H1 G9 resonance showing NOE contacts to neighboring nucleotides. (D) Imino proton spectra of the mutant env9bPIC14G/G39C without (top) and with 1.5-fold excess of SAH and 2 mM magnesium acetate (bottom). In the mutant with the inverted G:C base pairing orientation the base triplet with G9 cannot form and ligand binding is abrogated.

the electrostatic interaction of the aminocarboxypropyl side chain amino group with U16 O4. Adenosine, where the sulfur atom at the 5'-position involved in a chalcogen bond with O6 of G38 is replaced by oxygen, has a 4-fold weaker affinity compared to 5'-deoxy-5'-methylthioadenosine and a 22 times weaker affinity compared to SAH. Importantly, adenosine monophosphate (AMP), which is abundant in the cell, does not bind the riboswitch (Supplementary Figure S8 and Table S1). Surprisingly, the analogs of SAH and SAM that only lack the carboxyl group in the amino acid side chain—decarboxy-SAH and decarboxy-SAM—bind with better affinity compared to SAH and SAM, respectively (Supplementary Figure S8 and Table S1). This is particularly pronounced for decarboxy-SAM with a K_D of 0.06 μM for the env9b riboswitch. This K_D is 25-fold lower compared to the K_D for SAM. In contrast, other SAM riboswitches strongly prefer binding to SAM over decarboxy-SAM binding (56,57). Structurally, the preference of the SAM/SAH riboswitch for the decarboxy-analogs of SAH and SAM can be easily rationalized. In the structure of the SAH-RNA complex, the carboxy group of the ligand points into the direction of the negatively charged phosphate backbone, which is energetically unfavorable. The 1D imino proton spectrum of the RNA bound to decarboxy-SAM resembles that of the SAH- and SAM-bound RNA demonstrating that the overall structure of the complexes is highly similar (Supplementary Figure S8). In order to investigate if this preference for decarboxy-SAM over SAM and SAH is a more general feature of the SAM/SAH-riboswitches we measured the affinities for the three ligands for SAM/SAH-riboswitches from *Paracoccus denitrificans* (Pde-1-1) and *Roseobacter* sp. SK209-2-6 (SK209-52-II) (21). The secondary structures for these two SAM/SAH

riboswitch variants are shown in Supplementary Figure S9. Both riboswitches bound to the three ligands with the same order of affinities as the env9b riboswitch demonstrating that the preference for decarboxy-SAM over SAM is maybe a general feature of the SAM/SAH-riboswitches (Supplementary Figure S9). However, their affinity for decarboxy-SAM is only ~ 3 -fold (Pde-1-1) or ~ 2 -fold higher (SK209-52-II) than their affinity for SAM and not 25-fold higher as in the env9b riboswitch (Supplementary Table S1).

The biologically relevant ligand of the SAM/SAH-riboswitch

Interestingly, decarboxy-SAM is a naturally occurring metabolite in bacterial cells used in the biosynthesis of spermine, spermidine and other polyamines. A general assumption in the riboswitch field is that the cognate ligand of the riboswitch has the highest affinity for the RNA. Accordingly, the higher affinity of decarboxy-SAM compared to SAM for the riboswitch would suggest that decarboxy-SAM is the 'real' ligand of the SAM/SAH-riboswitch and not SAM or SAH. A natural target gene for regulation by a decarboxy-SAM-binding 'OFF'-riboswitch would be SAM-decarboxylase. Thus, we conducted an extended bioinformatics search for additional occurrences of the SAM/SAH-riboswitch that might also reveal novel gene associations. While we found 552 novel sequences for this riboswitch compared to version 14.0 of the Rfam database (58), they were almost exclusively associated with the SAM-synthetase gene *metK* and never with SAM-decarboxylase (E.C. 4.1.1.50). Six of nine of the organisms with fully sequenced genomes containing the SAM/SAH riboswitches (Supplementary Table S2) do not even contain genes for SAM-decarboxylase and might therefore not

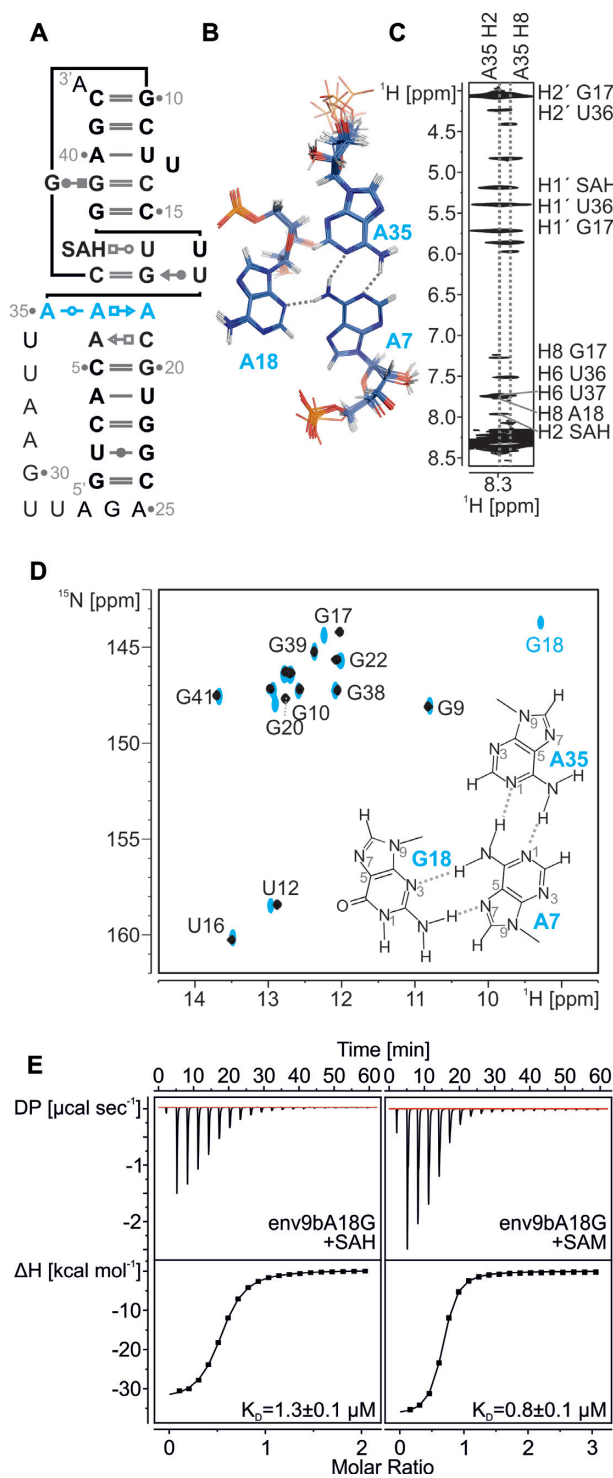


Figure 6. The adenine base triplet A7:A18:A35. (A) Structure diagram of the complex structure using the Leontis–Westhof notation. The base triplet A7:A18:A35 is colored in light blue. (B) The A7:A18:A35 base triplet in a bundle representation. (C) Strip from the $^1\text{H}, ^1\text{H}$ -NOESY-spectrum at the H2/H8 A35 resonances showing diagnostic NOE contacts between the A35 base protons and neighboring nucleotides. Signal assignments are indicated. (D) Overlay of the ^{15}N -HSQC spectra of the imino region of uniformly ^{15}N -labeled env9b RNA (black) and env9bA18G (light blue) in complex with SAH recorded at 20°C. The inset shows the geometry of the base triplet upon A18 replacement with G. (E) Representative ITC thermograms for env9bA18G titrated with SAH (left) or SAM (right). The resulting K_D values are indicated.

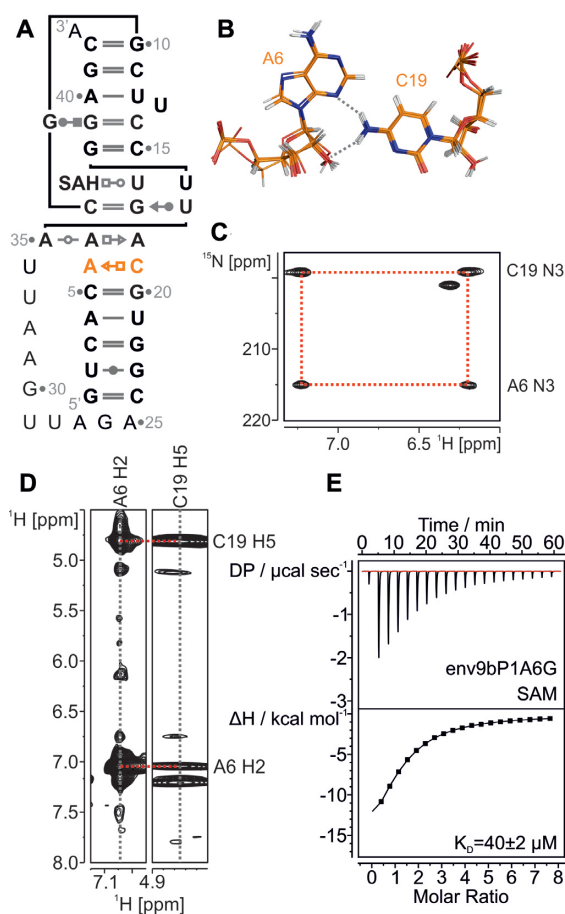


Figure 7. The non-canonical A6:C19 base pair. (A) Structure diagram of the complex structure using the Leontis–Westhof notation. The non-canonical base pair A6:C19 is colored in orange. (B) The A6:C19 base pair in a bundle representation. (C) HNN-COSY experiment for the identification of the hydrogen bond between the N4 amino group of C19 and the N3 nitrogen of A6. In this experiment the amino group protons of C19 simultaneously show cross peaks to the N3 nitrogen of C19 due to the intranucleotide through-bond $^2J_{\text{N4(C)N3(C)}}$ scalar coupling as well as to the A6 N3 nitrogen due to the internucleotide cross-hydrogen-bond $^2hJ_{\text{N4(C)N3(A)}}$ scalar coupling. (D) Strips from the $^1\text{H}, ^{13}\text{C}$ -NOESY-HSQC at the H2 A6 and H5 C19 resonance frequencies showing the strong NOEs between the H5 proton of C19 and the H2 proton of A6 supporting the unusual geometry of this A:C base pair. (E) Representative ITC thermogram for the titration of the env9bP1A6G riboswitch mutant with SAM. In this mutant a Watson–Crick G6:C19 base pair would replace the A:C base pair of the WT. The resulting K_D value is indicated.

even produce decarboxy-SAM. Overall, this argues against a role of decarboxy-SAM as the native cognate ligand for the SAM/SAH-riboswitch. Thus, the enhanced affinity for decarboxy-SAM compared to SAM and SAH might just be a side effect of the low level of structural complexity in the ligand binding site of the SAM/SAH riboswitch and the lack of ‘sophistication’ in the ligand recognition mode compared to the SAM-specific riboswitches. A similar argument can be made with regard to the question if SAM or SAH is the biologically relevant ligand for the SAM/SAH riboswitch. Our structure of the SAM/SAH riboswitch shows that there are fewer recognition elements that are able to distinguish a positively charged sulfonium

group in SAM from a neutral sulfur atom in SAH compared to the SAM-specific riboswitches. Furthermore, our structure clearly demonstrates that the SAM/SAH riboswitch is a translational 'OFF' switch since ligand binding helps to sequester the ribosome binding site. The dominating gene association for the SAM/SAH riboswitch is the SAM synthase gene *metK*. However, a shutdown in SAM-synthase expression induced by high concentrations of SAH signaling conditions where SAM consumption is apparently high seems to defy the logic of metabolic regulation. On the other hand, of the nine genome sequences from organisms with a SAM/SAH riboswitch that were practical to analyze (see 'Materials and Methods' section) only one apparently has no easily recognizable gene encoding an SAH hydrolase (Supplementary Table S2). Moreover, this organism, *Ketogulonicigenium vulgare* WSH-001, is closely related to *K. vulgare* Y25, which has a SAM/SAH-riboswitch as well as a SAH hydrolase. This situation suggests the possibility of a recent loss of SAH hydrolase in *K. vulgare* WSH-001. Thus, one might assume that in almost all organisms carrying the SAM/SAH riboswitch, the intracellular SAM levels are significantly higher than the SAH concentrations. Therefore, we tend to favor a scenario where SAM is the cognate, biologically important ligand for the SAM/SAH riboswitch. Its reduced structural complexity compared to the other SAM-specific riboswitches might thus represent a minimal solution for a SAM-based regulatory mechanism that would only work properly in a background where an SAH degradation system is already established. Moreover, all known examples for SAM/SAH-riboswitches are restricted to bacteria from the order Rhodobacterales. Taken together, this could imply that in contrast to other riboswitches the SAM/SAH riboswitch appeared late during bacterial evolution and is not a remnant of the RNA-world.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported solution NMR structures were deposited with the Protein Data bank under accession number 6HAG. The chemical shift assignments for the SAH/SAM-riboswitch bound to SAH were submitted to the BioMagResBank (BMRB) under accession number 27452.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Christian Richter and Manfred Strupf for maintenance of the NMR facility and to Kerstin Yacoub for help with sample preparation. Riboswitch searches used computer time provided by the Center for Information Services and High Performance Computing (ZIH) at TU Dresden.

FUNDING

Center for Biomolecular Magnetic Resonance (BMRZ) of the Goethe University Frankfurt; Deutsche Forschungsgemeinschaft (DFG) [CRC 902 'Molecular principles of

RNA-based regulation' B10]. Funding for open access charge: DFG.

Conflict of interest statement. None declared.

REFERENCES

- Fontecave, M., Atta, M. and Mulliez, E. (2004) S-adenosylmethionine. Nothing goes to waste. *Trends Biochem. Sci.*, **29**, 243–249.
- Struck, A.-W., Thompson, M.L., Wong, L.S. and Micklefield, J. (2012) S-adenosyl-methionine-dependent methyltransferases. Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBiochem.*, **13**, 2642–2655.
- Lee, J.E., Cornell, K.A., Riscoe, M.K. and Howell, P.L. (2003) Structure of *Escherichia coli* 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase inhibitor complexes provide insight into the conformational changes required for substrate binding and catalysis. *J. Biol. Chem.*, **278**, 8761–8770.
- Parveen, N. and Cornell, K.A. (2011) Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a crucial enzyme for bacterial metabolism. *Mol. Microbiol.*, **79**, 7–20.
- Kusakabe, Y., Ishihara, M., Umeda, T., Kuroda, D., Nakanishi, M., Kitade, Y., Gouda, H., Nakamura, K.T. and Tanaka, N. (2015) Structural insights into the reaction mechanism of S-adenosyl-L-homocysteine hydrolase. *Sci. Rep.*, **5**, 16641.
- Turner, M.A., Yang, X., Yin, D., Kuczera, K., Borchardt, R.T. and Howell, P.L. (2000) Structure and function of S-adenosylhomocysteine hydrolase. *Cell. Biochem. Biophys.*, **33**, 101–125.
- Meyer, B., Wurm, J.P., Sharma, S., Immer, C., Pogoryelov, D., Kötter, P., Lafontaine, D.L.J., Wöhnert, J. and Entian, K.-D. (2016) Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Res.*, **44**, 4304–4316.
- Umitsu, M., Nishimasu, H., Noma, A., Suzuki, T., Ishitani, R. and Nureki, O. (2009) Structural basis of AdoMet-dependent aminocarboxypropyl transfer reaction catalyzed by tRNA-wybutosine synthesizing enzyme, TYW2. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 15616–15621.
- Pegg, A.E., Xiong, H., Feith, D.J. and Shantz, L.M. (1998) S-adenosylmethionine decarboxylase. Structure, function and regulation by polyamines. *Biochem. Soc. Trans.*, **26**, 580–586.
- Bale, S., Baba, K., McCloskey, D.E., Pegg, A.E. and Ealick, S.E. (2010) Complexes of *Thermotoga maritima* S-adenosylmethionine decarboxylase provide insights into substrate specificity. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 181–189.
- Tabor, C.W. and Tabor, H. (1984) Polyamines. *Ann. Rev. Biochem.*, **53**, 749–790.
- O'Hagan, D., Schaffrath, C., Cobb, S.L., Hamilton, J.T.G. and Murphy, C.D. (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature*, **416**, 279.
- Schaffrath, C., Deng, H. and O'Hagan, D. (2003) Isolation and characterisation of 5'-fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya*. *FEBS Lett.*, **547**, 111–114.
- Holloway, C.T., Greene, R.C. and Su, C.H. (1970) Regulation of S-adenosylmethionine synthetase in *Escherichia coli*. *J. Bacteriol.*, **104**, 734–747.
- Wang, J.X. and Breaker, R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. Cell Biol.*, **86**, 157–168.
- Batey, R.T. (2011) Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdiscip. Rev. RNA*, **2**, 299–311.
- Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, **15**, 342–348.
- Gilbert, S.D., Montange, R.K., Stoddard, C.D. and Batey, R.T. (2006) Structural studies of the purine and SAM binding riboswitches. *Cold Spring Harb. Symp. Quant. Biol.*, **71**, 259–268.
- Gayan, M.A., Sherlock, M.E., Weinberg, Z. and Breaker, R.R. (2018) SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches. *RNA Biol.*, **4**, 371–378.
- Wang, J.X., Lee, E.R., Morales, D.R., Lim, J. and Breaker, R.R. (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol. Cell*, **29**, 691–702.

21. Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K., Moy, R.H. and Breaker, R.R. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.*, **11**, R31.
22. Duchardt-Ferner, E., Weigand, J.E., Ohlenschläger, O., Schmidtke, S.R., Suess, B. and Wöhnert, J. (2010) Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angew. Chem. Int. Ed. Engl.*, **49**, 6216–6219.
23. Wunderlich, C.H., Spitzer, R., Santner, T., Fauster, K., Tollinger, M. and Kreutz, C. (2012) Synthesis of (6-(13C)pyrimidine nucleotides as spin-labels for RNA dynamics. *J. Am. Chem. Soc.*, **134**, 7558–7569.
24. Alvarado, L.J., Longhini, A.P., LeBlanc, R.M., Chen, B., Kreutz, C. and Dayie, T.K. (2014) Chemo-enzymatic synthesis of selectively ¹³C/¹⁵N-labeled RNA for NMR structural and dynamics studies. *Methods Enzymol.*, **549**, 133–162.
25. Juen, M.A., Wunderlich, C.H., Nußbaumer, F., Tollinger, M., Kontaxis, G., Konrat, R., Hansen, D.F. and Kreutz, C. (2016) Excited states of nucleic acids probed by proton relaxation dispersion NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **55**, 12008–12012.
26. Ottink, O.M., Nelissen, F.H.T., Derks, Y., Wijmenga, S.S. and Heus, H.A. (2010) Enzymatic stereospecific preparation of fluorescent S-adenosyl-L-methionine analogs. *Anal. Biochem.*, **396**, 280–283.
27. Chatterjee, D., Kudlinzki, D., Linhard, V., Saxena, K., Schieborr, U., Gande, S.L., Wurm, J.P., Wöhnert, J., Abele, R., Rogov, V.V. *et al.* (2015) Structure and biophysical characterization of the S-adenosylmethionine-dependent O-methyltransferase PaMTH1, a putative enzyme accumulating during senescence of *Podospira anserina*. *J. Biol. Chem.*, **290**, 16415–16430.
28. Weickhmann, A.K., Keller, H., Duchardt-Ferner, E., Strebiter, E., Juen, M.A., Kremser, J., Wurm, J.P., Kreutz, C. and Wöhnert, J. (2018) NMR resonance assignments for the SAM/SAH-binding riboswitch RNA bound to S-adenosylhomocysteine. *Biomol. NMR Assign.*, **12**, 329–334.
29. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*. CANTINA, Goldau.
30. Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
31. Mueller, L., Legault, P. and Pardi, A. (1995) Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino protons. *J. Am. Chem. Soc.*, **117**, 11043–11048.
32. Grzesiek, S. and Bax, A. (1993) The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.*, **115**, 12593–12594.
33. Dingley, A.J. and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide ²J_{NN} couplings. *J. Am. Chem. Soc.*, **120**, 8293–8297.
34. Wöhnert, J., Dingley, A.J., Stoldt, M., Gorchach, M., Grzesiek, S. and Brown, L.R. (1999) Direct identification of NH...N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res.*, **27**, 3104–3110.
35. Majumdar, A., Kettani, A. and Skripkin, E. (1999) Observation and measurement of internucleotide ²J_{NN} coupling constants between ¹⁵N nuclei with widely separated chemical shifts. *J. Biomol. NMR*, **14**, 67–70.
36. Hennig, M. and Williamson, J.R. (2000) Detection of N-H...N hydrogen bonding in RNA via scalar couplings in the absence of observable imino proton resonances. *Nucleic Acids Res.*, **28**, 1585–1593.
37. Dallmann, A., Simon, B., Duszczyc, M.M., Kooshapur, H., Pardi, A., Bermel, W. and Sattler, M. (2013) Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences. *Angew. Chem. Int. Ed. Engl.*, **52**, 10487–10490.
38. Konradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 29–32.
39. Weinberg, Z., Lünse, C.E., Corbino, K.A., Ames, T.D., Nelson, J.W., Roth, A., Perkins, K.R., Sherlock, M.E. and Breaker, R.R. (2017) Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res.*, **45**, 10811–10823.
40. O’Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI. Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
41. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG. New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
42. Sklenár, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1994) Correlation of nucleotide base and sugar protons in a ¹⁵N-labeled HIV-1 RNA oligonucleotide by ¹H-¹⁵N HSQC experiments. *J. Biomol. NMR*, **4**, 117–122.
43. Simorre, J.P., Zimmermann, G.R., Pardi, A., Farmer, B.T. and Mueller, L. (1995) Triple resonance HNCCCH experiments for correlating exchangeable and nonexchangeable cytidine and uridine base protons in RNA. *J. Biomol. NMR*, **6**, 427–432.
44. Sklenár, V., Dieckmann, T., Butcher, S.E. and Feigon, J. (1996) Through-bond correlation of imino and aromatic resonances in ¹³C-, ¹⁵N-labeled RNA via heteronuclear TOCSY. *J. Biomol. NMR*, **7**, 83–87.
45. Wöhnert, J., Ramachandran, R., Görlach, M. and Brown, L.R. (1999) Triple-resonance experiments for correlation of H5 and exchangeable pyrimidine base hydrogens in (¹³C), (¹⁵N)-labeled RNA. *J. Magn. Reson.*, **139**, 430–433.
46. Phan, A.T. (2000) Long-range imino proton-¹³C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175–178.
47. Nikonowicz, E.P. and Pardi, A. (1992) Three-dimensional heteronuclear NMR studies of RNA. *Nature*, **355**, 184–186.
48. Sklenár, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in ¹⁵N, ¹³C-labeled RNA oligonucleotides. *J. Biomol. NMR*, **3**, 721–727.
49. Tolbert, T.J. and Williamson, J.R. (1997) Preparation of specifically deuterated and ¹³C-labeled RNA for NMR studies using enzymatic synthesis. *J. Am. Chem. Soc.*, **119**, 12100–12108.
50. Gilbert, S.D., Rambo, R.P., van Tyne, D. and Batey, R.T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.*, **15**, 177–182.
51. Klein, D.J., Edwards, T.E. and Ferré-D’Amaré, A.R. (2009) Cocystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.*, **16**, 343–344.
52. Koebel, M.R., Cooper, A., Schmadeke, G., Jeon, S., Narayan, M. and Sirimulla, S. (2016) S...O and S...N sulfur bonding interactions in protein-ligand complexes. Empirical considerations and scoring function. *J. Chem. Inf. Model.*, **56**, 2298–2309.
53. Egli, M. and Sarkhel, S. (2007) Lone pair-aromatic interactions. To stabilize or not to stabilize. *Acc. Chem. Res.*, **40**, 197–205.
54. Chawla, M., Chermak, E., Zhang, Q., Bujnicki, J.M., Oliva, R. and Cavallo, L. (2017) Occurrence and stability of lone pair-π stacking interactions between ribose and nucleobases in functional RNAs. *Nucleic Acids Res.*, **45**, 11019–11032.
55. Chen, G., Znosko, B.M., Kennedy, S.D., Krugh, T.R. and Turner, D.H. (2005) Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry*, **44**, 2845–2856.
56. Corbino, K.A., Barrick, J.E., Lim, J., Welz, R., Tucker, B.J., Puskarczyk, I., Mandal, M., Rudnick, N.D. and Breaker, R.R. (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.*, **6**, R70.
57. Lim, J., Winkler, W.C., Nakamura, S., Scott, V. and Breaker, R.R. (2006) Molecular-recognition characteristics of SAM-binding riboswitches. *Angew. Chem. Int. Ed. Engl.*, **45**, 964–968.
58. Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D. and Petrov, A.I. (2018) Rfam 13.0. Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, **46**, D335–D342.

IV DISCUSSION

1. STRUCTURAL FEATURES FOR STABILITY AND LIGAND BINDING

1.1. PROTONATION AS STRUCTURAL FEATURES

Protonation is a commonly used feature in proteins since e.g., amino acid histidine has a favorable pK_a of 6.04 and the intrinsic property is used e.g., as a sensor. In contrast, the pK_a value of nucleobases seems unfavorable for protonation since it is far from neutrality ($pK_{a,AMP} = 3.5$, $pK_{a,CMP} = 4.2$). Furthermore, the detection of protonated residues remains challenging because in crystal structure determination only gives a static picture of the investigated complex and is insensitive to the position of hydrogen atoms. Hydrogen bonds are only modeled into the structures due to the orientation of the nucleobase and are assumed due to proximity parameters. Since the pK_a values of the free or single stranded nucleobases are far from neutrality, protonation seems almost impossible and therefore are rarely assumed. In contrast, NMR experiments can directly proof presence of hydrogen bonds. Furthermore, ITC experiments give early and fast insights into the stability of the protonation. The emerging research on aptamers and riboswitches show that protonation is used, even if rarely. RNA structures can build lasting environment to stabilize the protonated state to either maintain a structure (structural protonation) or to realize a functional need (functional protonation).

The investigated GTP class II aptamer was shown to use a structural protonation. Here, the protonation is needed to stabilize a base quartet C15:G24:A11⁺:G9 that lies underneath a base triplet building the floor of ligand binding site. On the one hand, the increased number of potential hydrogen bonds stabilizes the environment to enable the protonated state. On the other hand, the base triplet builds a stable foundation for ligand binding. The protonation is stable even at temperatures up to 40°C. ITC experiments showed that the protonated A11 has a lower limit pK_a of about 8.9 which is drastically shifted from the values determined for single stranded RNA.

The CDN riboswitches with the G20A mutation investigated showed that the usage of protonated bases can lead to bispecific ligand binding. Therefore, these riboswitches use protonation as a functional feature. In the case for the investigated riboswitches the protonation of A20 leads to a promiscuous binding and therefore disrupts the selectivity. In

the investigated case the protonation and the pK_a of the protonated A20 are influenced by the overall stability of the riboswitch, especially of the stem regions. The potassium ion concentration seems to have an impact of the stability as well, but further investigations are needed for a profound explanation. Additionally, the biological significance for the promiscuity needs further bioinformatical and functional investigations.

Both, the structural and functional protonated states investigated are sensible to the ligand concentration and the pH. The latter binds different ligands due to the protonation leading to the same regulatory result with or without protonation while the former is stabilized by the protonation and the regulatory result is different for the protonated as for the unprotonated state. Therefore, the structural protonation follows an AND logic. Especially, the pH dependent switching of protonation states gives hints that riboswitches could not only be regulated by the presence of the ligand but also but the environmental pH or other conditions. The potential usages are in a wide range. To date, it is known that protonated states influence frame shifting and ligand specificity, but also potential thermometer might be possible since protonated states are often temperature sensible. Even though, the example for structural protonation investigated here is from an aptamer and therefore artificial, it gives good insights into the potential that RNA has - it might be possible that protonation will be found more frequently in riboswitches as well.

1.2. PROMISCUITY

Riboswitches are known for their ability to highly discriminate against even the slightest ligand differences. The fluoride Riboswitch is one prominent example. It binds fluoride ions and discriminates against chloride ions which only differ in the atom radii (Ren et al. 2012). It is an impressive example for the potential of riboswitches that can even sense e.g., the atom radii, functional groups, or charge. In some very few known cases, the riboswitch was shown to bind several ligands with high affinities in experimental set-ups. The *yjdF* motif RNA were found of being azaaromatic riboswitches. In *in vivo* essays they bind five different azaaromatics and even more in *in vitro* experiments (Li et al. 2016). The binding behavior is assumed to influences by non-specific interactions. In cases when the *yjdF* gene is not riboswitch regulated, the gene sequence for PadR regulatory proteins is present which are known to bind the same class of ligands as found for the riboswitch (Silva et al. 2005). Further investigations are needed since it is even possible that the native ligand just was not found yet.

Another example is the cobalamin riboswitch which is grouped into two classes: Cbl-I and Cbl-II, according to the binding pocket architecture. Cobalamin derivatives differ in the functional group which is axially coordinated to the corrin ring and the riboswitches from the different classes or sub-classes are specific to either of the derivatives e.g., Cbl-I and Cbl-IIb are selective for adenosylcobalamin, while Cbl-IIa binds hydroxocobalamin and methylcobalamin. The cobalamin riboswitch from *B. subtilis* has structural features from the Cbl-I and Cbl-IIa classes. *In vivo*, the riboswitch reacts to cobalamin but *in vitro* it is found to also bind methyl cobalamin or hydroxocobalamin, respectively. The data indicates that the riboswitch might react to a spectrum of cobalamin derivatives rather than to a distinct native ligand (Chan and Mondragón 2020).

The promiscuous behavior of the SAM/SAH binding riboswitch was shown to result from a minimal recognition mode. In the case, the high selectivity normally found for riboswitches is not necessarily needed due to low SAH intracellular concentration. Bioinformatic investigations suggest that the riboswitch was introduced late in the evolutionary path of riboswitches and that most organisms that encode the SAM/SAH binding riboswitch do also encode a SAH degrading gene. Therefore, it is most likely that the SAH concentrations and thereby SAH binding to the riboswitch are negligible (Publication III.4 and further discussion Chapter IV.2).

Also, protonation *in vivo* can help riboswitches induce promiscuity as was shown for the Vc2 G20A mutant and the CDN binding riboswitches. In this case, promiscuity can be found either because false positive regulation is not severe in the cellular regulation process or the ligands are bound in different circumstances due to different external stimuli, respectively. Nevertheless, the CDN riboswitch promiscuity is not fully understood in cellular function so far and further investigation in the cellular interplay are needed to get the full picture (Publication III.2)

Promiscuity seems to have different reasons as can be derived from the limited data available so far. Firstly, similar structures are willingly recognized because the same output is needed due to potential cell toxicity as found for the azaromatics riboswitches or potentially for the cobalamin riboswitches. Secondly, a minimal recognition mode is used since the concentration of the other analogues is too low to be significant as was shown for the SAM/SAH binding riboswitch. And lastly, the binding pattern and reasons for promiscuity are

just poorly understood until now, due to a lack of global insights and further investigations are needed as was shown for the CDN binding riboswitches.

2. SAM/SAH BINDING RIBOSWITCH

2.1. LIGAND RECOGNITION COMPARED TO OTHER SAM BINDING RIBOSWITCHES

The SAM binding riboswitches are of the longest known riboswitches and until now many different classes were identified. The SAM binding riboswitch classes show several different strategies to build stable structures. Furthermore, they show different strategies to selectively recognize SAM and discriminate against SAH. While the size and global architecture of the riboswitch classes are diverse, all SAM binding riboswitches have the coordination of the sulfonium ion by two or more uracil carboxyl groups in common (Figure 9). Furthermore, the SAM selective riboswitches coordinate the acp group specifically to enable discrimination against analogs without this group. In contrast, the SAM/SAH binding riboswitch shows less interactions to methionine moiety compared to the other riboswitches exclusively binding SAM. Furthermore, it shows a cavity with no interactions to the acp group.

Both the SAM-I riboswitch and the SAM/SAH riboswitch are pseudoknots. The SAM-I riboswitch shows that the loops residues are integral parts of the structure, and no flexible residues are found. Here, the loop residues interact with the helices to further stabilize the overall structure. In contrast, the SAM/SAH riboswitch consensus sequence shows a stretch of 5-13 residues that are not integrated into the structure and remain flexible. While residues are needed to span the P1 helix, the number of residues and the lack of specific interactions increase the degree of freedom and flexibility.

The recognition and discrimination pattern exhibited by the SAM/SAH binding riboswitch is not very selective but sufficient since SAH level are very low. The SAH levels are most likely stringently controlled by other mechanisms, due to its cell toxicity. Furthermore, the genes associated with the SAM/SAH binding riboswitch sequence are almost exclusively SAM methylases genes. A false positive regulation by SAH can be assumed to be not very severe due to the gene association and low SAH levels.

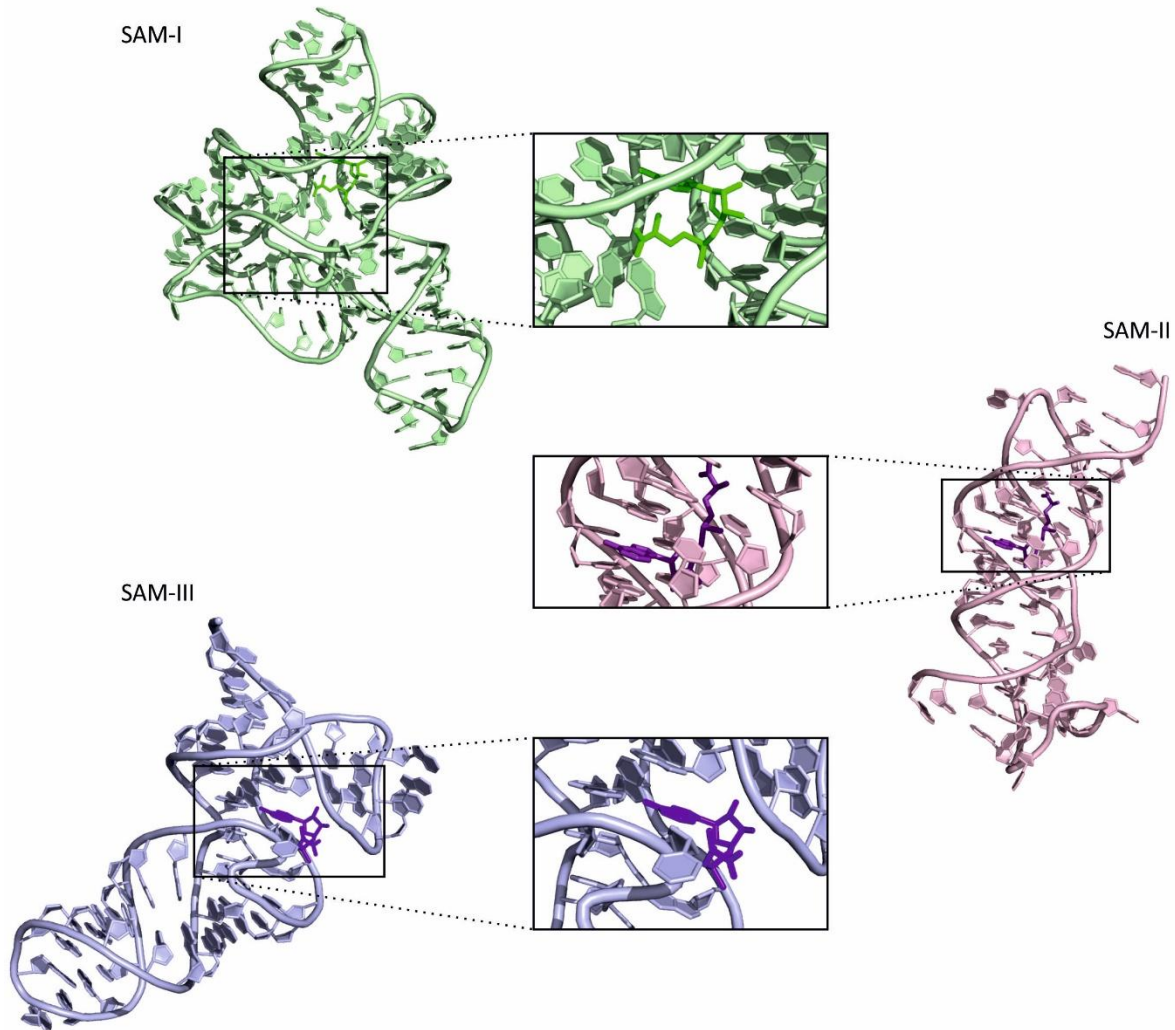


Figure 9: Representatives of the three super-families of SAM binding riboswitches which are using different ligand orientation. The SAM-I riboswitch (PDB 2GIS) is shown in green, the structure of the SAM-II riboswitch (PDB 2QWY) in pink and the SAM-III (PDB 3E5C) in blue.

2.2. COMPARISON NMR AND CRYSTAL STRUCTURE

Recently, Lilley and coworkers (Huang et al. 2020) solved the crystal structure of the SAM/SAH binding riboswitch in complex with SAH, SAM, and other related analogues, respectively. In the following, the numbering according to Publications 3 and 4 will be used. For the crystallization they used an RNA that composed of two strands: the first containing the P1 stem-loop and the second containing the conserved nucleotides of the pseudoknot building part. Hence, they were able to leave out the flexible linker. Due to the high content of conservation in the used segments (Weinberg et al. 2010), the sequences used for the crystal and NMR structure were mainly identical, except for some minor differences which are all in agreement with the conservation pattern. The P1 stem is one base pair shorter in the NMR structure than in the crystal structure and contains at position 23 a G-to-A alteration in the

crystal structure. For phasing information, 5-bromocytidine was incorporated at position 3. additionally, the crystal structure contains an A-to-G alteration at position 18.

In the obtained crystal lattice, six riboswitch macromolecules that form an equilateral triangle. Two riboswitches coaxially stack with their P1 helices to build the sides of the triangle, respectively. At the angles, A35 from one riboswitch interacts with the P2 minor groove of the second riboswitch. An additional SAH molecule is bound at the vertices to each riboswitch. The adenosine moiety is hydrogen bonded to A9, G17 and U36 of one riboswitch, while its sugar moiety contacts U16' of the second riboswitch. All crystal structures show that the ligands used are bound in very similar fashion and the overlay has an RMSD of only 0.197-0.722 Å, respectively, referenced to the SAH bound riboswitch (AB chains).

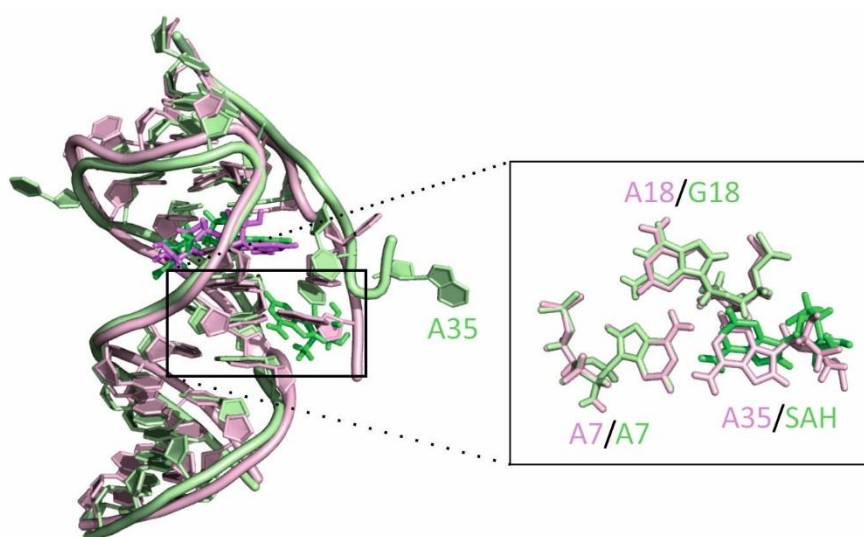


Figure 10: Overlay of the NMR and crystal SAM/SAH binding riboswitch structures. NMR structure (6HAG) is shown in pink and the crystal structure is shown in green. Ligands are shown in stick representations in respective color. Inlet shows overlay of A base triplet A7:A18:A35. The crystal structure contains a A-to-G mutation at position 18 and SAH displaces A35.

Overall, the crystal structure of the riboswitch in complex with SAH is in good agreement with NMR structure (Chapter III, Publication 4). Nevertheless, in the crystal structure A35 loops out of the structure and interacts with an adjacent riboswitch in the crystal

lattice, while in the NMR structure A35 is part of the intramolecular base triplet A9-A18-A35. As mentioned, at position 18 a G is found in the crystal structure which alone does not explain the difference in the base pairing pattern. For the NMR structure it was shown that an A-to-G mutation at position 18 improves the binding affinity, while the spectra look similar. Therefore, it can be assumed that the G-to-A mutation does not interfere with the formation of the base triplet. Furthermore, in the crystal lattice a second SAH displaces A35 which in turn is free to interact with the minor groove of the adjacent P2 helix. An overlay of the two structures shows that SAH indeed displaces A35 (Figure 10, inlet). A35 is conserved (Weinberg

et al. 2010) and it can be assumed that A35 is not looped out of the structure under physiological conditions.

Lilley and coworkers also showed in FRET experiments that the riboswitch exists in three distinct forms: an unbound and/or flexible population, a compact structure, and a dynamic population (Huang et al. 2020). The equilibrium is shifted to the compact structure in presence of ligand. Nevertheless, the percentage of unbound population remained unchanged and can be assumed as a misfolded population. Despite the presence or lack of ligand, about half of the riboswitches remained dynamic. In contrast, for the SAM-II riboswitch it was shown that the dynamic population was reduced to 1-5% in presence of ligand. Taken together, the high portion of dynamic riboswitches underlie the assumption that the SAM/SAH riboswitch is regulated by SAM which has higher intracellular concentrations than SAH.

3. CONCLUSION

Generally, riboswitches are found to show a wide range of ligand binding and ligand selection strategies. They are understood as being very selective for one native ligand that in most cases has a direct connection to the gene regulated by the riboswitch. Recent studies on RNA structures deepened the knowledge we have to date about riboswitches and the building blocks that are used to form the intriguing structures. Nevertheless, studies on riboswitch classes and aptamer families that bind the same ligand using a different global architecture show that the strategies to bind the same ligands is highly diverse. Furthermore, the overall stability of riboswitches seems to be a feature that is used to influence various outcomes. All structural features have a complex interplay with each other. Therefore, rational design studies remain difficult. Furthermore, protonation seems to be a feature that has a wide potential usage, even though it is rarely found in functional riboswitches so far. Also, ligand selection processes in most cases can be easily explained by structural insights while the rare promiscuous binding behavior of some riboswitches remain poorly understood. Therefore, every structural investigation is important to increase the insights into RNA structural potential.

V REFERENCES

- Atilho RM, Mirihana Arachchilage G, Greenlee EB, Knecht KM, Breaker RR (2019) *A bacterial riboswitch class for the thiamin precursor HMP-PP employs a terminator-embedded aptamer*. *Elife* 8
- Ballin JD, Prevas JP, Ross CR, Toth EA, Wilson GM, Record MT (2010) *Contributions of the histidine side chain and the N-terminal alpha-amino group to the binding thermodynamics of oligopeptides to nucleic acids as a function of pH*. *Biochemistry* 49:2018–2030
- Barnwal RP, Yang F, Varani G (2017) *Applications of NMR to structure determination of RNAs large and small*. *Arch Biochem Biophys* 628:42–56
- Barrick JE, Breaker RR (2007) *The power of riboswitches*. *Sci Am* 296:50–57
- Boehm A, Kaiser M, Li H, Spangler C, Kasper CA, Ackermann M, Kaefer V, Sourjik V, Roth V, Jenal U (2010) *Second messenger-mediated adjustment of bacterial swimming velocity*. *Cell* 141:107–116
- Brown A, Shao S (2018) *Ribosomes and cryo-EM: a duet*. *Curr Opin Struct Biol* 52:1–7
- Brown RS, Clark BF, Coulson RR, Finch JT, Klug A, Rhodes D (1972) *Crystallization of pure species of bacterial tRNA for x-ray diffraction studies*. *Eur J Biochem* 31:130–134
- Burgstaller P, Famulok M (1994) *Isolation of RNA Aptamers for Biological Cofactors by In Vitro Selection*. *Angew Chem Int Ed Engl*:1084–1087
- Burgstaller P, Kochoyan M, Famulok M (1995) *Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding*. *Nucleic Acids Res* 23:4769–4776
- Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, Disdier F, Baltzinger M, 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
- Cameron ADS, Volar M, Bannister LA, Redfield RJ (2008) *RNA secondary structure regulates the translation of *sxy* and competence development in *Haemophilus influenzae**. *Nucleic Acids Res* 36:10–20
- Cantoni GL (1975) *Biological methylation: selected aspects*. *Annu Rev Biochem* 44:435–451
- Carothers JM, Oestreich SC, Davis JH, Szostak JW (2004) *Informational complexity and functional activity of RNA structures*. *J Am Chem Soc* 126:5130–5137
- Carothers JM, Oestreich SC, Szostak JW (2006) *Aptamers selected for higher-affinity binding are not more specific for the target ligand*. *J Am Chem Soc* 128:7929–7937
- Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T (2004) *Structural basis of activity and allosteric control of diguanylate cyclase*. *Proc Natl Acad Sci U S A* 101:17084–17089
- Chan CW, Mondragón A (2020) *Crystal structure of an atypical cobalamin riboswitch reveals RNA structural adaptability as basis for promiscuous ligand binding*. *Nucleic Acids Res* 48:7569–7583
- Chang KY, Tinoco I (1994) *Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome*. *Proc Natl Acad Sci U S A* 91:8705–8709
- Chen X, Kang H, Shen LX, Chamorro M, Varmus HE, Tinoco I (1996) *A characteristic bent conformation of RNA pseudoknots promotes -1 frameshifting during translation of retroviral RNA*. *J Mol Biol* 260:479–483

- Chen J-H, Yajima R, Chadalavada DM, Chase E, Bevilacqua PC, Golden BL (2010) A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* 49:6508–6518
- Chen H, Egger M, Xu X, Flemmich L, Krasheninina O, Sun A, Micura R, Ren A (2020) Structural distinctions between NAD⁺ riboswitch domains 1 and 2 determine differential folding and ligand binding. *Nucleic Acids Res* 48:12394–12406
- Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP (1996) S-Adenosylmethionine and methylation. *FASEB J* 10:471–480
- Christen M, Christen B, Folcher M, Schauerte A, Jenal U (2005) Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem* 280:30829–30837
- Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, Breaker RR (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol* 6:R70
- Crick F (1970) Central dogma of molecular biology. *Nature* 227:561–563
- Dahlberg AE (1989) The functional role of ribosomal RNA in protein synthesis. *Cell* 57:525–529
- Dann CE, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130:878–892
- Davies BW, Bogard RW, Young TS, Mekalanos JJ (2012) Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell* 149:358–370
- Davis JH, Szostak JW (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc Natl Acad Sci U S A* 99:11616–11621
- Della Ragione F, Porcelli M, Carteni-Farina M, Zappia V, Pegg AE (1985) *Escherichia coli* S-adenosylhomocysteine/5'-methylthioadenosine nucleosidase. Purification, substrate specificity and mechanism of action. *Biochem J* 232:335–341
- Dieckmann T, Suzuki E, Nakamura GK, Feigon J (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* 2:628–640
- Doherty EA, Doudna JA (2000) Ribozyme structures and mechanisms. *Annu Rev Biochem* 69:597–615
- Duerig A, Abel S, Folcher M, Nicollier M, Schwede T, Amiot N, Giese B, Jenal U (2009) Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev* 23:93–104
- Eaton BE, Gold L, Zichi DA (1995) Let's get specific: the relationship between specificity and affinity. *Chem Biol* 2:633–638
- Ebrahimi M, Rossi P, Rogers C, Harbison GS (2001) Dependence of ¹³C NMR chemical shifts on conformations of rna nucleosides and nucleotides. *J Magn Reson* 150:1–9
- Edwards AL, Reyes FE, Héroux A, Batey RT (2010) Structural basis for recognition of S-adenosylhomocysteine by riboswitches. *RNA* 16:2144–2155
- Egli M, Minasov G, Su L, Rich A (2002) Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc Natl Acad Sci U S A* 99:4302–4307
- Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822
- Fan P, Suri AK, Fiala R, Live D, Patel DJ (1996) Molecular recognition in the FMN-RNA aptamer complex. *J Mol Biol* 258:480–500
- Farabaugh PJ (1996) Programmed translational frameshifting. *Microbiol Rev* 60:103–134

- Feig AL (2007) *Applications of isothermal titration calorimetry in RNA biochemistry and biophysics*. *Biopolymers* 87:293–301
- Fuchs RT, Grundy FJ, Henkin TM (2006) *The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase*. *Nat Struct Mol Biol* 13:226–233
- Fuchs RT, Grundy FJ, Henkin TM (2007) *S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the SMK box translational riboswitch RNA*. *Proc Natl Acad Sci U S A* 104:4876–4880
- Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) *NMR spectroscopy of RNA*. *Chembiochem* 4:936–962
- Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC, Breaker RR (2015) *Bacterial riboswitches cooperatively bind Ni(2+) or Co(2+) ions and control expression of heavy metal transporters*. *Mol Cell* 57:1088–1098
- Gao A, Serganov A (2014) *Structural insights into recognition of c-di-AMP by the ydaO riboswitch*. *Nat Chem Biol* 10:787–792
- Garriga G, Lambowitz AM (1986) *Protein-dependent splicing of a group I intron in ribonucleoprotein particles and soluble fractions*. *Cell* 46:669–680
- Garriga G, Lambowitz AM, Inoue T, Cech TR (1986) *Mechanism of recognition of the 5' splice site in self-splicing group I introns*. *Nature* 322:86–89
- Garst AD, Héroux A, Rambo RP, Batey RT (2008) *Crystal structure of the lysine riboswitch regulatory mRNA element*. *J Biol Chem* 283:22347–22351
- Gilbert SD, Rambo RP, van Tyne D, Batey RT (2008) *Structure of the SAM-II riboswitch bound to S-adenosylmethionine*. *Nat Struct Mol Biol* 15:177–182
- Gottstein-Schmidtke SR, Duchardt-Ferner E, Groher F, Weigand JE, Gottstein D, Suess B, Wöhnert J (2014) *Building a stable RNA U-turn with a protonated cytidine*. *RNA* 20:1163–1172
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme*. *Cell* 35:849–857
- Haebel PW, Gutmann S, Ban N (2004) *Dial tm for rescue: tmRNA engages ribosomes stalled on defective mRNAs*. *Curr Opin Struct Biol* 14:58–65
- Hengge R (2010) *Cyclic-di-GMP reaches out into the bacterial RNA world*. *Sci Signal* 3:pe44
- Hennelly SP, Sanbonmatsu KY (2011) *Tertiary contacts control switching of the SAM-I riboswitch*. *Nucleic Acids Res* 39:2416–2431
- Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A (1965) *Structure of a ribonucleic acid*. *Science* 147:1462–1465
- Holloway CT, Greene RC, Su CH (1970) *Regulation of S-adenosylmethionine synthetase in Escherichia coli*. *J Bacteriol* 104:734–747
- Howard MT, Aggarwal G, Anderson CB, Khatri S, Flanigan KM, Atkins JF (2005) *Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons*. *EMBO J* 24:1596–1607
- Huang L, Serganov A, Patel DJ (2010) *Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch*. *Mol Cell* 40:774–786
- Huang L, Ishibe-Murakami S, Patel DJ, Serganov A (2011) *Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch*. *Proc Natl Acad Sci U S A* 108:14801–14806
- Huang L, Wang J, Wilson TJ, Lilley DMJ (2017) *Structure of the Guanidine III Riboswitch*. *Cell Chem Biol* 24:1407–1415.e2

- Huang L, Lilley DMJ (2018) *Structure and ligand binding of the SAM-V riboswitch*. *Nucleic Acids Res* 46:6869–6879
- Huang L, Liao T-W, Wang J, Ha T, Lilley DMJ (2020) *Crystal structure and ligand-induced folding of the SAM/SAH riboswitch*. *Nucleic Acids Res* 48:7545–7556
- Iwasaki A, Medzhitov R (2010) *Regulation of adaptive immunity by the innate immune system*. *Science* 327:291–295
- Iwata-Reuyl D (2003) *Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA*. *Bioorg Chem* 31:24–43
- Johnson JE, Reyes FE, Polaski JT, Batey RT (2012) *B12 cofactors directly stabilize an mRNA regulatory switch*. *Nature* 492:133–137
- Jones CP, Ferré-D'Amaré AR (2014) *Crystal structure of a c-di-AMP riboswitch reveals an internally pseudo-dimeric RNA*. *EMBO J* 33:2692–2703
- Jucker FM, Pardi A (1995) *Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif*. *Biochemistry* 34:14416–14427
- Kalani MR, Moradi A, Moradi M, Tajkhorshid E (2013) *Characterizing a histidine switch controlling pH-dependent conformational changes of the influenza virus hemagglutinin*. *Biophys J* 105:993–1003
- Kang M, Peterson R, Feigon J (2009) *Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA*. *Mol Cell* 33:784–790
- Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh Y-F, Hammond MC (2015) *GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP*. *Proc Natl Acad Sci U S A* 112:5383–5388
- Kim SH, Quigley G, Suddath FL, Rich A (1971) *High-resolution x-ray diffraction patterns of crystalline transfer RNA that show helical regions*. *Proc Natl Acad Sci U S A* 68:841–845
- Kim SH, Quigley GJ, Suddath FL, McPherson A, Sneden D, Kim JJ, Weinzierl J, Rich A (1973) *Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain*. *Science* 179:285–288
- Klein DJ, Ferré-D'Amaré AR (2006) *Structural basis of glmS ribozyme activation by glucosamine-6-phosphate*. *Science* 313:1752–1756
- Knappenberger AJ, Reiss CW, Strobel SA (2018) *Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA*. *Elife* 7
- Koebel MR, Cooper A, Schmadeke G, Jeon S, Narayan M, Sirimulla S (2016) *S···O and S···N Sulfur Bonding Interactions in Protein-Ligand Complexes: Empirical Considerations and Scoring Function*. *J Chem Inf Model* 56:2298–2309
- Krasteva PV, Fong JCN, Shikuma NJ, Beyhan S, Navarro MVAS, Yildiz FH, Sondermann H (2010) *Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP*. *Science* 327:866–868
- Kulshina N, Baird NJ, Ferré-D'Amaré AR (2009) *Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch*. *Nat Struct Mol Biol* 16:1212–1217
- Kusakabe Y, Ishihara M, Umeda T, Kuroda D, Nakanishi M, Kitade Y, Gouda H, Nakamura KT, Tanaka N (2015) *Structural insights into the reaction mechanism of S-adenosyl-L-homocysteine hydrolase*. *Sci Rep* 5:16641
- Layer G, Moser J, Heinz DW, Jahn D, Schubert W-D (2003) *Crystal structure of coproporphyrinogen III oxidase reveals cofactor geometry of Radical SAM enzymes*. *EMBO J* 22:6214–6224

- Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR (2010) *An allosteric self-splicing ribozyme triggered by a bacterial second messenger*. *Science* 329:845–848
- Li S, Hwang XY, Stav S, Breaker RR (2016) *The yjdB riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds*. *RNA* 22:530–541
- Lieberman JA, Salim M, Krucinska J, Wedekind JE (2013) *Structure of a class II preQ1 riboswitch reveals ligand recognition by a new fold*. *Nat Chem Biol* 9:353–355
- Lieberman JA, Bogue JT, Jenkins JL, Salim M, Wedekind JE (2014) *ITC analysis of ligand binding to preQ₁ riboswitches*. *Meth Enzymol* 549:435–450
- Lieberman JA, Suddala KC, Aytenfisu A, Chan D, Belashov IA, Salim M, Mathews DH, Spitale RC, Walter NG, Wedekind JE (2015) *Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics*. *Proc Natl Acad Sci U S A* 112:E3485-94
- Lim J, Winkler WC, Nakamura S, Scott V, Breaker RR (2006) *Molecular-recognition characteristics of SAM-binding riboswitches*. *Angew Chem Int Ed Engl* 45:964–968
- Liu Y, Huynh DT, Yeates TO (2019) *A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold*. *Nat Commun* 10:1864
- Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A (2008) *Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism*. *Nat Struct Mol Biol* 15:1076–1083
- Lu K, Miyazaki Y, Summers MF (2010a) *Isotope labeling strategies for NMR studies of RNA*. *J Biomol NMR* 46:113–125
- Lu C, Ding F, Chowdhury A, Pradhan V, Tomsic J, Holmes WM, Henkin TM, Ke A (2010b) *SAM recognition and conformational switching mechanism in the Bacillus subtilis yitJ S box/SAM-I riboswitch*. *J Mol Biol* 404:803–818
- Lukavsky PJ, Kim I, Otto GA, Puglisi JD (2003) *Structure of HCV IRES domain II determined by NMR*. *Nat. Struct. Biol.* 10:1033–1038
- Malevanets A, Chong PA, Hansen DF, Rizk P, Sun Y, Lin H, Muhandiram R, Chakrabartty A, Kay LE, Forman-Kay JD, Wodak SJ (2017) *Interplay of buried histidine protonation and protein stability in prion misfolding*. *Sci Rep* 7:882
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR (2004) *A glycine-dependent riboswitch that uses cooperative binding to control gene expression*. *Science* 306:275–279
- Matyjasik MM, Batey RT (2019) *Structural basis for 2'-deoxyguanosine recognition by the 2'-dG-II class of riboswitches*. *Nucleic Acids Res* 47:10931–10941
- Meyer MM, Ames TD, Smith DP, Weinberg Z, Schwalbach MS, Giovannoni SJ, Breaker RR (2009) *Identification of candidate structured RNAs in the marine organism 'Candidatus Pelagibacter ubique'*. *BMC Genomics* 10:268
- Mirihana Arachchilage G, Sherlock ME, Weinberg Z, Breaker RR (2018) *SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches*. *RNA Biol* 15:371–378
- Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E (2002) *Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria*. *Cell* 111:747–756
- Moore SD, Sauer RT (2005) *Ribosome rescue: tmRNA tagging activity and capacity in Escherichia coli*. *Mol Microbiol* 58:456–466

- Moore SD, Sauer RT (2007) *The tmRNA system for translational surveillance and ribosome rescue*. *Annu Rev Biochem* 76:101–124
- Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002) *Genetic control by a metabolite binding mRNA*. *Chem Biol* 9:1043
- Nakano S, Chadalavada DM, Bevilacqua PC (2000) *General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme*. *Science* 287:1493–1497
- Nasiri AH, Wurm JP, Immer C, Weickmann AK, Wöhnert J (2016) *An intermolecular G-quadruplex as the basis for GTP recognition in the class V-GTP aptamer*. *RNA* 22:1750–1759
- Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR (2013) *Riboswitches in eubacteria sense the second messenger c-di-AMP*. *Nat Chem Biol* 9:834–839
- Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR (2015) *Control of bacterial exoelectrogenesis by c-AMP-GMP*. *Proc Natl Acad Sci U S A* 112:5389–5394
- Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA (2001) *RNA tertiary interactions in the large ribosomal subunit: the A-minor motif*. *Proc Natl Acad Sci U S A* 98:4899–4903
- Nixon PL, Giedroc DP (2000) *Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot*. *J Mol Biol* 296:659–671
- Nixon PL, Rangan A, Kim Y-G, Rich A, Hoffman DW, Hennig M, Giedroc DP (2002) *Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot*. *J Mol Biol* 322:621–633
- Noma A, Kirino Y, Ikeuchi Y, Suzuki T (2006) *Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA*. *EMBO J* 25:2142–2154
- Nonin-Lecomte S, Felden B, Dardel F (2006) *NMR structure of the Aquifex aeolicus tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation*. *Nucleic Acids Res* 34:1847–1853
- Nou X, Kadner RJ (1998) *Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli*. *J Bacteriol* 180:6719–6728
- Novichkov PS, Li X, Kuehl JV, Deutschbauer AM, Arkin AP, Price MN, Rodionov DA (2014) *Control of methionine metabolism by the SahR transcriptional regulator in Proteobacteria*. *Environ Microbiol* 16:1–8
- Ohlenschläger O, Haumann S, Ramachandran R, Görlach M (2008) *Conformational signatures of ¹³C chemical shifts in RNA ribose*. *J Biomol NMR* 42:139–142
- Old IG, Hunter MG, Wilson DT, Knight SM, Weatherston CA, Glass RE (1988) *Cloning and characterization of the genes for the two homocysteine transmethylases of Escherichia coli*. *Mol Gen Genet* 211:78–87
- Otto GA, Puglisi JD (2004) *The pathway of HCV IRES-mediated translation initiation*. *Cell* 119:369–380
- Parveen N, Cornell KA (2011) *Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism*. *Mol Microbiol* 79:7–20
- Pesavento C, Hengge R (2009) *Bacterial nucleotide-based second messengers*. *Curr Opin Microbiol* 12:170–176
- Peselis A, Serganov A (2012) *Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch*. *Nat Struct Mol Biol* 19:1182–1184
- Peselis A, Serganov A (2018) *ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands*. *Nat Chem Biol* 14:887–894
- Pfingsten JS, Costantino DA, Kieft JS (2006) *Structural basis for ribosome recruitment and manipulation by a viral IRES RNA*. *Science* 314:1450–1454

- Pfingsten JS, Costantino DA, Kieft JS (2007) Conservation and diversity among the three-dimensional folds of the *Dicistroviridae* intergenic region IRESes. *J Mol Biol* 370:856–869
- Pikovskaya O, Polonskaia A, Patel DJ, Serganov A (2011) Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol* 7:748–755
- Poiata E, Meyer MM, Ames TD, Breaker RR (2009) A variant riboswitch aptamer class for *S*-adenosylmethionine common in marine bacteria. *RNA* 15:2046–2056
- Price IR, Gaballa A, Ding F, Helmann JD, Ke A (2015) Mn(2+)-sensing mechanisms of *yybP-ykoY* orphan riboswitches. *Mol Cell* 57:1110–1123
- Quigley GJ, Rich A (1976) Structural domains of transfer RNA molecules. *Science* 194:796–806
- Quirk DJ, Raines RT (1999) His ... Asp catalytic dyad of ribonuclease A: histidine pKa values in the wild-type, D121N, and D121A enzymes. *Biophys J* 76:1571–1579
- Quirk DJ, Park C, Thompson JE, Raines RT (1998) His...Asp catalytic dyad of ribonuclease A: conformational stability of the wild-type, D121N, D121A, and H119A enzymes. *Biochemistry* 37:17958–17964
- Ravnum S, Andersson DI (2001) An adenosyl-cobalamin (coenzyme-B12)-repressed translational enhancer in the *cob* mRNA of *Salmonella typhimurium*. *Mol Microbiol* 39:1585–1594
- Redfield RJ (1991) *sxy-1*, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. *J Bacteriol* 173:5612–5618
- Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL, Breaker RR (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol Microbiol* 68:918–932
- Reiss CW, Strobel SA (2017) Structural basis for ligand binding to the guanidine-II riboswitch. *RNA* 23:1338–1343
- Reiss CW, Xiong Y, Strobel SA (2017) Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure* 25:195–202
- Ren A, Patel DJ (2014) *c-di*-AMP binds the *ydaO* riboswitch in two pseudo-symmetry-related pockets. *Nat Chem Biol* 10:780–786
- Ren A, Rajashankar KR, Patel DJ (2012) Fluoride ion encapsulation by Mg²⁺ ions and phosphates in a fluoride riboswitch. *Nature* 486:85–89
- Ren A, Rajashankar KR, Patel DJ (2015a) Global RNA Fold and Molecular Recognition for a *pfl* Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. *Structure* 23:1375–1381
- Ren A, Xue Y, Peselis A, Serganov A, Al-Hashimi HM, Patel DJ (2015b) Structural and Dynamic Basis for Low-Affinity, High-Selectivity Binding of L-Glutamine by the Glutamine Riboswitch. *Cell Rep* 13:1800–1813
- Ren A, Wang XC, Kellenberger CA, Rajashankar KR, Jones RA, Hammond MC, Patel DJ (2015c) Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep* 11:1–12
- Rey DA, Nentwich SS, Koch DJ, Rückert C, Pühler A, Tauch A, Kalinowski J (2005) The *McbR* repressor modulated by the effector substance *S*-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032. *Mol Microbiol* 56:871–887
- Rey DA, Pühler A, Kalinowski J (2003) The putative transcriptional repressor *McbR*, member of the *TetR*-family, is involved in the regulation of the metabolic network directing the synthesis of sulfur containing amino acids in *Corynebacterium glutamicum*. *J Biotechnol* 103:51–65

- Rich A, Kasha M, Pullman B (1962) *Horizons in biochemistry*. Academic Press, New York:103
- Robertson DL, Joyce GF (1990) *Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA*. *Nature* 344:467–468
- Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A (1974) *Structure of yeast phenylalanine tRNA at 3 Å resolution*. *Nature* 250:546–551
- Römling U (2008) *Great times for small molecules: c-di-AMP, a second messenger candidate in Bacteria and Archaea*. *Sci Signal* 1:pe39
- Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, Vroom E de, van der Marel GA, van Boom JH, Benziman M (1987) *Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid*. *Nature* 325:279–281
- Ryan RP, Fouhy Y, Lucey JF, Crossman LC, Spiro S, He Y-W, Zhang L-H, Heeb S, Cámara M, Williams P, Dow JM (2006) *Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover*. *Proc Natl Acad Sci U S A* 103:6712–6717
- Saenger W (1984) *Principles of Nucleic Acid Structure*, 1st edn. Springer-Verlag New York
- Salim NN, Feig AL (2009) *Isothermal titration calorimetry of RNA*. *Methods* 47:198–205
- Sassanfar M, Szostak JW (1993) *An RNA motif that binds ATP*. *Nature* 364:550–553
- Scarabino D, Crisari A, Lorenzini S, Williams K, Tocchini-Valentini GP (1999) *tRNA prefers to kiss*. *EMBO J* 18:4571–4578
- Schäfer A, Tauch A, Droste N, Pühler A, Kalinowski J (1997) *The Corynebacterium glutamicum cglIM gene encoding a 5-cytosine methyltransferase enzyme confers a specific DNA methylation pattern in an McrBC-deficient Escherichia coli strain*. *Gene* 203:95–101
- Schauder S, Shokat K, Surette MG, Bassler BL (2001) *The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule*. *Mol Microbiol* 41:463–476
- Serganov A, Yuan Y-R, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004) *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs*. *Chem Biol* 11:1729–1741
- Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ (2006) *Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch*. *Nature* 441:1167–1171
- Serganov A, Huang L, Patel DJ (2008) *Structural insights into amino acid binding and gene control by a lysine riboswitch*. *Nature* 455:1263–1267
- Serganov A, Huang L, Patel DJ (2009) *Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch*. *Nature* 458:233–237
- Sherlock ME, Sadeeshkumar H, Breaker RR (2019) *Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates*. *Biochemistry* 58:401–410
- Shoeman R, Coleman T, Redfield B, Greene RC, Smith AA, Saint-Girons I, Brot N, Weissbach H (1985) *Regulation of methionine synthesis in Escherichia coli: effect of metJ gene product and S-adenosylmethionine on the in vitro expression of the metB, metL and metJ genes*. *Biochem Biophys Res Commun* 133:731–739
- Silva RS de, Kovacicova G, Lin W, Taylor RK, Skorupski K, Kull FJ (2005) *Crystal structure of the virulence gene activator AphA from Vibrio cholerae reveals it is a novel member of the winged helix transcription factor superfamily*. *J Biol Chem* 280:13779–13783
- Singh V, Fedeles BI, Essigmann JM (2015) *Role of tautomerism in RNA biochemistry*. *RNA* 21:1–13

- Sinha S, Mell J, Redfield R (2013) *The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy*. *Mol Microbiol* 88:1106–1119
- Smith AA, Greene RC, Kirby TW, Hindenach BR (1985) *Isolation and characterization of the product of the methionine-regulatory gene metJ of Escherichia coli K-12*. *Proc Natl Acad Sci U S A* 82:6104–6108
- Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA (2009) *Structural basis of ligand binding by a c-di-GMP riboswitch*. *Nat Struct Mol Biol* 16:1218–1223
- Smith KD, Lipchock SV, Livingston AL, Shanahan CA, Strobel SA (2010) *Structural and biochemical determinants of ligand binding by the c-di-GMP riboswitch*. *Biochemistry* 49:7351–7359
- Smith KD, Shanahan CA, Moore EL, Simon AC, Strobel SA (2011) *Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches*. *Proc Natl Acad Sci U S A* 108:7757–7762
- Stoner GL, Eisenberg MA (1975a) *Biosynthesis of 7, 8-diaminopelargonic acid from 7-keto-8-aminopelargonic acid and S-adenosyl-L-methionine. The kinetics of the reaction*. *J Biol Chem* 250:4037–4043
- Stoner GL, Eisenberg MA (1975b) *Purification and properties of 7, 8-diaminopelargonic acid aminotransferase*. *J Biol Chem* 250:4029–4036
- Su L, Chen L, Egli M, Berger JM, Rich A (1999) *Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot*. *Nat. Struct. Biol.* 6:285–292
- Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR (2006) *Tandem riboswitch architectures exhibit complex gene control functions*. *Science* 314:300–304
- Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR (2008) *Riboswitches in eubacteria sense the second messenger cyclic di-GMP*. *Science* 321:411–413
- Sun L, Wu J, Du F, Chen X, Chen ZJ (2013) *Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway*. *Science* 339:786–791
- Sun A, Gasser C, Li F, Chen H, Mair S, Krasheninina O, Micura R, Ren A (2019) *SAM-VI riboswitch structure and signature for ligand discrimination*. *Nat Commun* 10:5728
- Tabor CW, Tabor H (1984) *Methionine adenosyltransferase (S-adenosylmethionine synthetase) and S-adenosylmethionine decarboxylase*. *Adv Enzymol Relat Areas Mol Biol* 56:251–282
- Tok JB, Cho J, Rando RR (2000) *RNA aptamers that specifically bind to a 16S ribosomal RNA decoding region construct*. *Nucleic Acids Res* 28:2902–2910
- Touboul D, Bouchoux G, Zenobi R (2008) *Gas-phase protonation thermochemistry of adenosine*. *J Phys Chem B* 112:11716–11725
- Trausch JJ, Xu Z, Edwards AL, Reyes FE, Ross PE, Knight R, Batey RT (2014) *Structural basis for diversity in the SAM clan of riboswitches*. *Proc Natl Acad Sci U S A* 111:6624–6629
- Tuerk C, Gold L (1990) *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. *Science* 249:505–510
- Turner MA, Yang X, Yin D, Kuczera K, Borchardt RT, Howell PL (2000) *Structure and function of S-adenosylhomocysteine hydrolase*. *Cell Biochem Biophys* 33:101–125
- Ueland PM (1982) *Pharmacological and biochemical aspects of S-adenosylhomocysteine and S-adenosylhomocysteine hydrolase*. *Pharmacol Rev* 34:223–253

- Venditti V, Clos L, Niccolai N, Butcher SE (2009) *Minimum-energy path for a u6 RNA conformational change involving protonation, base-pair rearrangement and base flipping*. *J Mol Biol* 391:894–905
- Wang JX, Breaker RR (2008) *Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine*. *Biochem Cell Biol* 86:157–168
- Wang JX, Lee ER, Morales DR, Lim J, Breaker RR (2008) *Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling*. *Mol Cell* 29:691–702
- Watson PY, Fedor MJ (2012) *The ydaO motif is an ATP-sensing riboswitch in Bacillus subtilis*. *Nat Chem Biol* 8:963–965
- Weinberg Z, Wang JX, Bogue J, Yang J, Corbino KA, Moy RH, Breaker RR (2010) *Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes*. *Genome Biol* 11:R31
- Winkler W, Nahvi A, Breaker RR (2002) *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression*. *Nature* 419:952–956
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR (2003) *An mRNA structure that controls gene expression by binding S-adenosylmethionine*. *Nat. Struct. Biol.* 10:701–707
- Winzer K, Hardie KR, Burgess N, Doherty N, Kirke D, Holden MTG, Linforth R, Cornell KA, Taylor AJ, Hill PJ, Williams P (2002) *LuxS: its role in central metabolism and the in vitro synthesis of 4-hydroxy-5-methyl-3(2H)-furanone*. *Microbiology (Reading)* 148:909–922
- Woese CR (1967) *The genetic Code*. Harper & Row
- Woese CR, Winker S, Gutell RR (1990) *Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops"*. *Proc Natl Acad Sci U S A* 87:8467–8471
- Wood S, Ferré-D'Amaré AR, Rueda D (2012) *Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding*. *ACS Chem Biol* 7:920–927
- Wunderlich CH, Spitzer R, Santner T, Fauster K, Tollinger M, Kreutz C (2012) *Synthesis of (6-(13)C)pyrimidine nucleotides as spin-labels for RNA dynamics*. *J Am Chem Soc* 134:7558–7569
- Wunderlich CH, Juen MA, LeBlanc RM, Longhini AP, Dayie TK, Kreutz C (2015) *Stable isotope-labeled RNA phosphoramidites to facilitate dynamics by NMR*. *Meth Enzymol* 565:461–494
- Yan F, Fujimori DG (2011) *RNA methylation by radical SAM enzymes RlmN and Cfr proceeds via methylene transfer and hydride shift*. *Proc Natl Acad Sci U S A* 108:3930–3934
- Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M (1996) *Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy*. *Science* 272:1343–1347
- Yarnell WS, Roberts JW (1999) *Mechanism of intrinsic transcription termination and antitermination*. *Science* 284:611–615
- Zamay AS, Zamay GS, Kolovskaya OS, Zamay TN, Berezovski MV (2017) *Aptamer-Based Methods for Detection of Circulating Tumor Cells and Their Potential for Personalized Diagnostics*. *Adv Exp Med Biol* 994:67–81
- Zhang K, Li S, Kappel K, Pintilie G, Su Z, Mou T-C, Schmid MF, Das R, Chiu W (2019) *Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution*. *Nat Commun* 10:5511
- Zou X, Wu J, Gu J, Shen L, Mao L (2019) *Application of Aptamers in Virus Detection and Antiviral Therapy*. *Front Microbiol* 10:1462

VI APPENDIX

1. SUPPLEMENTARY INFORMATION WOLTER ET AL



Supporting Information

A Stably Protonated Adenine Nucleotide with a Highly Shifted pK_a Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer

*Antje C. Wolter, A. Katharina Weickhmann, Amir H. Nasiri, Katharina Hantke, Oliver Ohlenschläger, Christoph H. Wunderlich, Christoph Kreutz, Elke Duchardt-Ferner, and Jens Wöhnert**

anie_201609184_sm_miscellaneous_information.pdf

Supporting Information

1. Tables S1 – S3
2. Supplementary Figures S1 – S7
3. Materials & Methods
4. Supplementary References

1. Tables

Table S1. Overview of experimentally measured pK_a values for individual adenine nucleotides in structured RNAs.

system	method	pK_a	source
leadzyme	^{13}C -NMR	3.1-6.5	Legault and Pardi, 1994 ^[1] , 1997 ^[2]
hairpin ribozyme domain A	^{13}C -NMR	6.2	Cai and Tinoco, 1996 ^[3]
branch point helix spliceosome	^{13}C -NMR	5.0-6.1	Smith and Nikonowicz, 1998 ^[4]
hairpin ribozyme domain B	^{13}C -NMR	4.8-5.8	Ravindranathan et al., 2000 ^[5]
U6 RNA ISL spliceosome	^{13}C -NMR	6.0-6.5	Huppler et al., 2002 ^[6]
hairpin ribozyme - pre-catalytic state	Raman crystallography	5.4	Guo et al., 2009 ^[7] Libermann et al., 2012 ^[8]
hairpin ribozyme - transition state		6.7	
<i>Neurospora VS</i> ribozyme A730 loop	^{13}C -NMR	4.4-4.7	Desjardins et al., 2011 ^[9]
MLV pseudoknot	^{13}C -NMR	6.23	Houck-Loomis et al., 2011 ^[10]
model dsRNAs	^{31}P -NMR	6.5-8.1	Wilcox and Bevilacqua, 2013 ^[11]
group II intron domain 5	^{13}C -NMR	6.4, 6.7	Pechlaner et al., 2015 ^[12]
mini-twister ribozyme pre-catalytic state	^{13}C -NMR	5.1	Košutić et al., 2015 ^[13]
pistol ribozyme pre-catalytic state	^{13}C -NMR	4.7	Ren et al., 2016 ^[14]

A range of pK_a values is reported in cases where multiple adenine nucleotides have shifted pK_a 's in a given structure. For comparison, the pK_a value for adenine nucleotides in single stranded RNA's is ~ 3.7 ^[14,15]. Bevilacqua and coworkers have also developed a method relying on the fluorescence properties of a 2-aminopurine

substituted nucleotide inserted next to the protonated nucleotide ^[16]. However, since the 2-aminopurine (2-AP) substitution is not possible without destabilizing effects in all structural contexts this method has been applied so far only to dsDNAs where an A:U base pair next to an A⁺:C mismatch was replaced by an isoenergetic 2-AP:U base pair and to a protonated C⁺ found in the BWYV pseudoknot RNA. In the case of the pistol ribozyme the adenine protonation is discussed to possibly occur at the N3 nitrogen of adenine and not the N1 as in the other cases shown in Table S1.

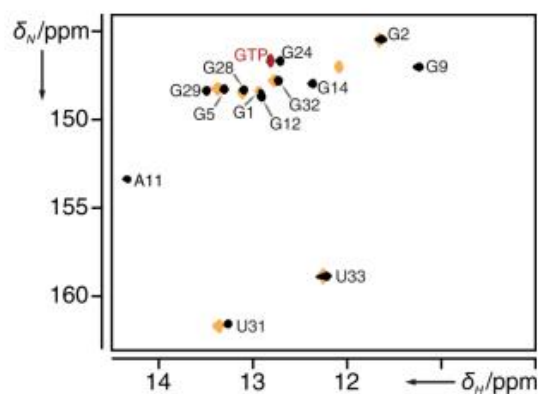
Table S2. Structural statistics for the GTP-class II RNA aptamer complex (pdb entry 5LWJ).

NMR restraints	
Total distance restraints	927
RNA	
Intra-residue	399
Sequential	275
Long-range	129
Hydrogen bond	74
RNA-GTP	
NOE	40
Hydrogen bond	10
Total dihedral restraints	194
Ribose pucker	60
Backbone	134
Structural statistics	
RMSD to mean structure (Å) (G2-G12, G14-A20, A22-U33)	0.52 ±0.07

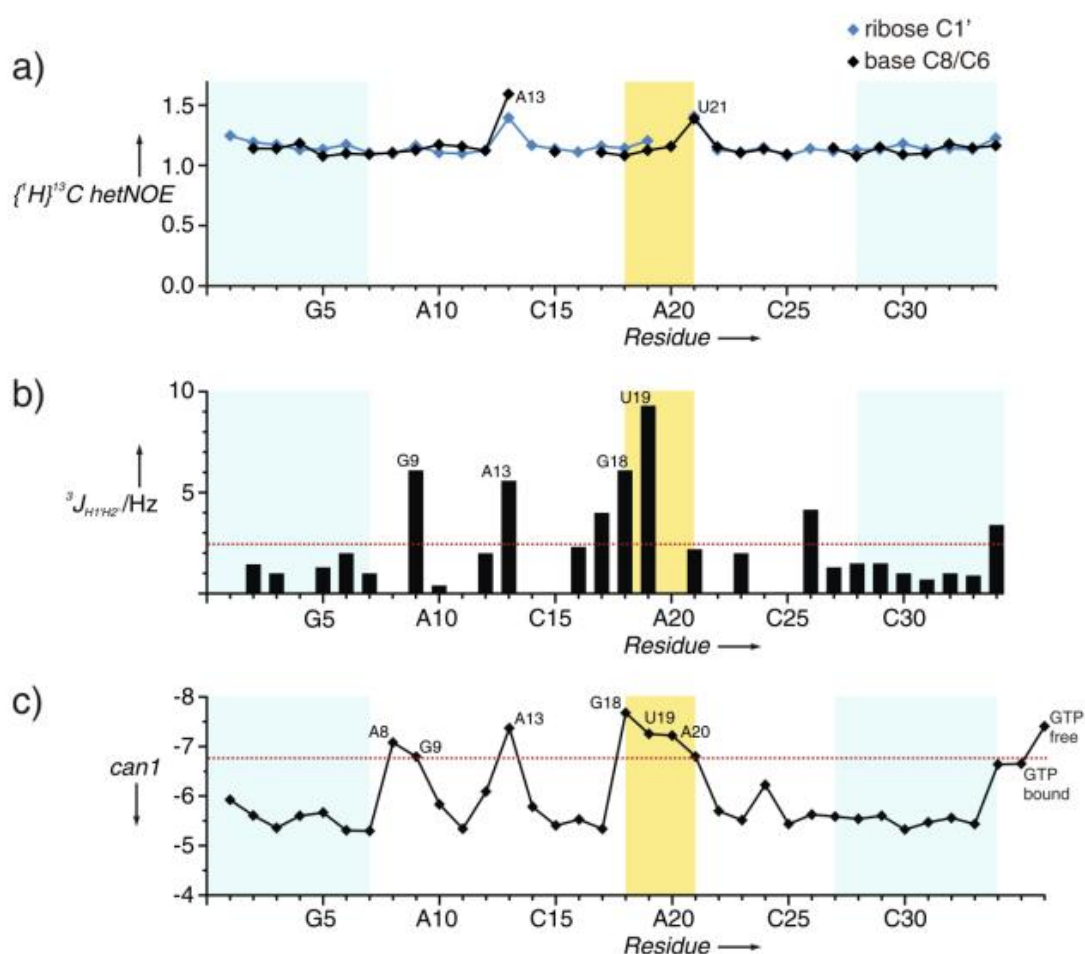
Table S3. pH-dependent K_D values and stoichiometry (n) for GTP-binding to the class II aptamer RNA as derived from isothermal titration calorimetry (ITC). All K_D and n values are averages from three independent titrations. RNA and ligand concentrations were optimized for each pH value in order to yield titration curves with clearly discernable plateaus at the beginning and the end of each titration (see Supplementary Fig. 6).

pH	K_D [nM]	n
5.8	721±72	0.99±0.03
6.0	812±33	0.94±0.02
6.3	1011±119	0.84±0.03
6.5	1015±87	0.95±0.09
6.8	1503±118	1.02±0.13
7.3	2863±114	1.00±0.06
7.5	4303±12	1.16±0.01
7.8	8230±351	1.16±0.01
8.0	17433±1305	0.84±0.02
8.3	19833±1464	1.26±0.03

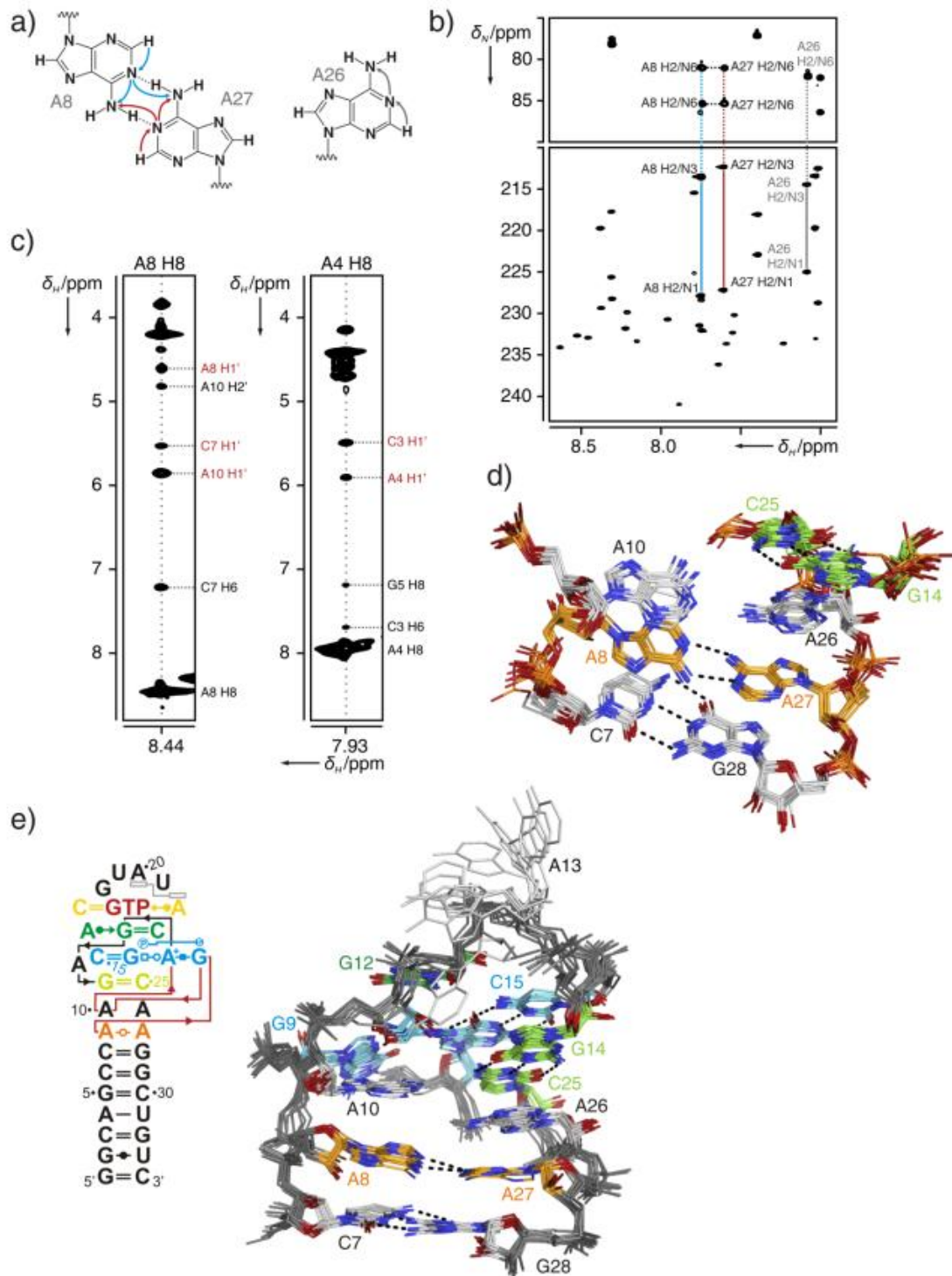
2. Supplementary Figures



Supplementary Fig. 1. GTP-binding induces tertiary structure folding of the class II aptamer RNA. Overlay of the imino group regions of $^1\text{H},^{15}\text{N}$ -HSQC spectra at 20°C , pH 6.3, 2 mM Mg^{2+} , of the free class II RNA aptamer (orange), the aptamer RNA in the presence of 2 equivalents of $^{13}\text{C}/^{15}\text{N}$ -labeled GTP and 2 mM Mg^{2+} (black) and $^{13}\text{C}/^{15}\text{N}$ -GTP in complex with unlabeled RNA in the presence of 2 mM Mg^{2+} showing only the imino group signal of the bound GTP (red). Resonance assignments for the GTP-bound state are indicated. Additional exchange-protected imino proton signals are observed upon GTP-binding in agreement with the formation of novel hydrogen bonding and base pairing interactions in the presence of ligand.

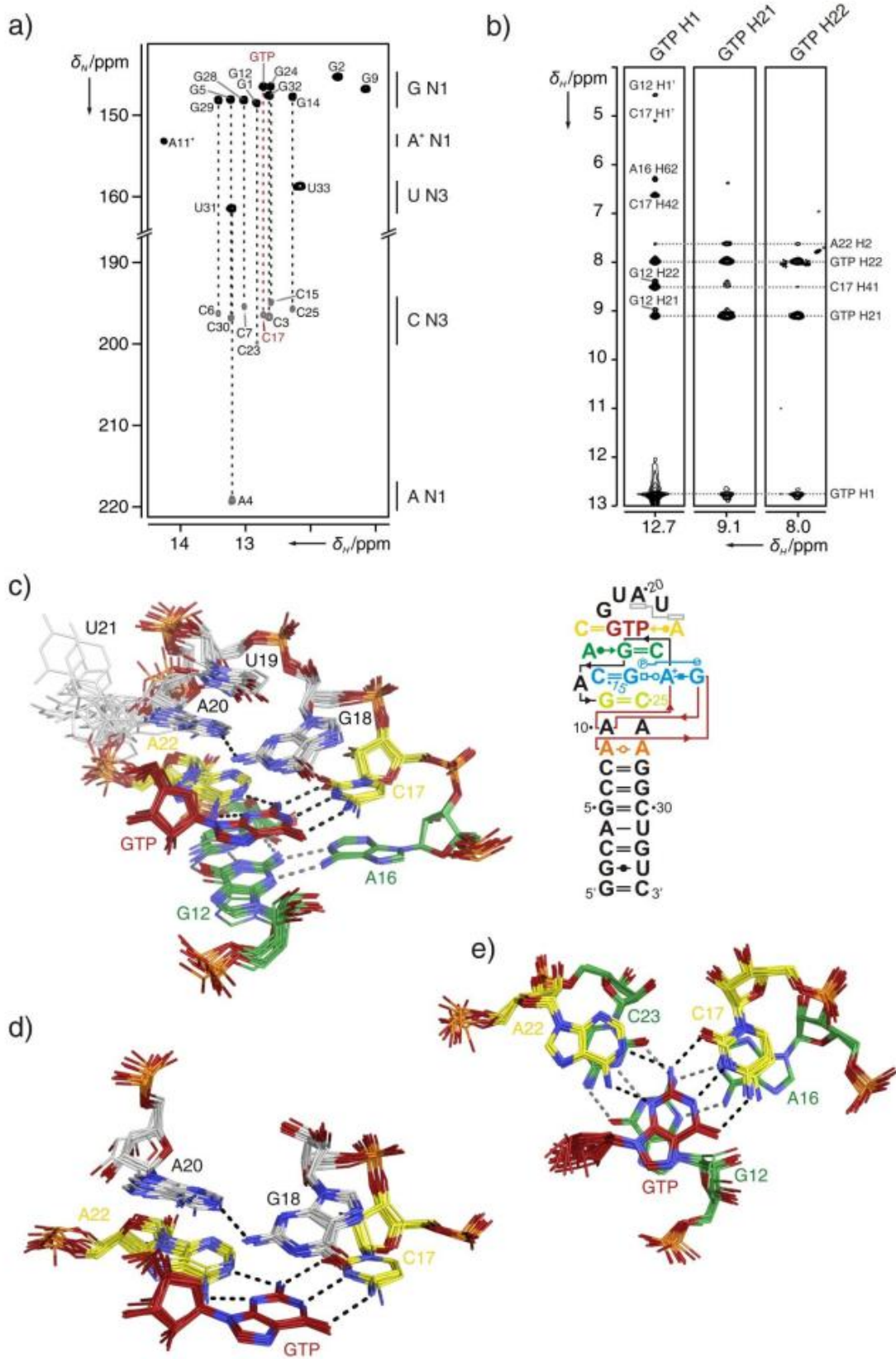


Supplementary Fig. 2. Dynamics of the GTP class II aptamer RNA complex. (a) $\{^1\text{H}\}^{13}\text{C}$ -hetNOE values measured at 20°C, pH 6.3, 2 mM Mg^{2+} , for the aromatic H8C8 and H6C6 moieties (black squared line) of the nucleobases and the anomeric H1'C1' groups (blue squared line) of the ribose moieties along the sequence of the aptamer. Flexible residues with elevated $\{^1\text{H}\}^{13}\text{C}$ -hetNOE values are assigned. The light blue and the light yellow backgrounds indicate nucleotides in the terminal helix and the apical loop, respectively. (b) $^3J_{\text{H1H2}}$ -scalar couplings as a function of sequence. Nucleotides with scalar coupling values < 2.5 Hz (red dotted line) have a C3'-endo ribose conformation, and were constrained accordingly in the structure calculations. (c) Canonical coordinate *can1* calculated from the ribose carbon C1', C2', C3' and C4' chemical shifts. Nucleotides with *can1* values > 6.8 (red dotted line) have a C2'-endo ribose conformation and were constrained accordingly in structure calculations.^[17]

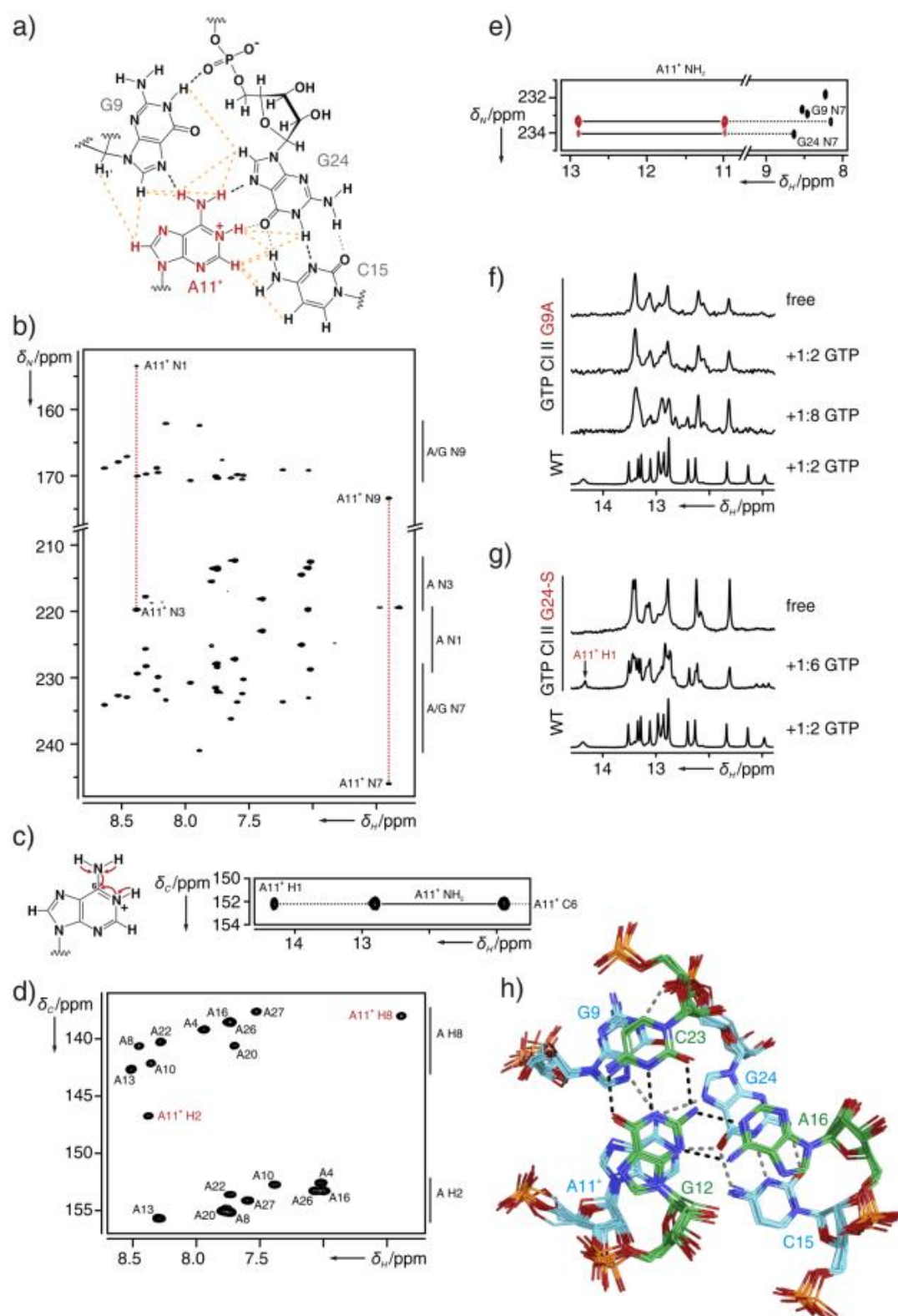


Supplementary Fig. 3. Direct evidence for the formation of a *trans* A8:A27 base pair accompanied by a reversion of the backbone direction between nucleotides G7 and

A8 and conformation of the internal bulge region of the GTP-aptamer complex. (a) Schematic representation of a *trans* A:A base pair stabilized by two hydrogen bonds. Hydrogen bonds are indicated by dashed lines. Arrows indicate the expected magnetization transfer pathways for the long-range HNN-COSY experiment in this type of base pair (left) as well as for a non-hydrogen bonded A (right).^[18] (b) Long-range HNN-COSY spectrum optimized for detection of hydrogen bonds involving amino groups (top) and two-bond ^1H , ^{15}N -HSQC (bottom) measured at 20°C, pH 6.3, 2 mM Mg^{2+} . Cross peaks are observed between adenine N1 nitrogen nuclei and the N6 amino nitrogens of the same nucleotide as well as the N6 amino nitrogen of the hydrogen bonding partner. For comparison the observed correlations for A26 is shown for which the amino group is not hydrogen bonded to a nitrogen acceptor. (c) Slice from a ^1H , ^{13}C -NOESY-HSQC measured at 20°C, pH 6.3, 2 mM Mg^{2+} , for the A8 H8C8 group showing NOE cross peaks (red) indicative of a chain reversal at this position. For comparison a slice for the H8C8 group of residue A4 is also shown which is located in the canonical A-form terminal helix of the class II aptamer RNA. (d) Local environment of the *trans* A8:A27 base pair shown with orange colored carbon atoms. The A8:A27 base pair directly stacks on top of the last Watson-Crick base pair C7:G28 of the terminal helix. A26 stacks on top of A27 and A10 stacks on top of A8. A10 and A26 do not form a base pair and efforts to enforce hydrogen bonding between these two bases artificially lead to the violation of many NOE restraints. However, A26 forms additional extensive stacking interactions with the G14:C25 Watson-Crick base pair (carbon atoms colored in light green) of the central helix. (e) Overall conformation of the bulge region. In the predicted secondary structure (Figure 1b) the central bulge is highly asymmetric with 6 nucleotides (A8, G9, A10, A11, G12 and A13) on its 5'-side and only two nucleotides (A26 and A27) on its 3'-side. In the structure of the GTP-complex G9, A11, G12 and A13 are extruded from the bulge and G9, A11 and G12 are involved in long-range tertiary interactions. G9 and A11 form a base quartet (see below) with the C15:G24 Watson-Crick base pair from the central helix. G9, A11, C15 and G24 carbon atoms are colored blue. A11 stacks on A10. G12 (carbon atoms colored dark green) forms a long-range Watson-Crick base pair with C23 from the apical loop (see Supplementary Fig. 4, not shown here) which is part of a base triple that also includes A16 and forms the floor of the GTP-binding site. A13 (all atoms colored gray) is a flexible spacer residue (see Supplementary Fig. 2).

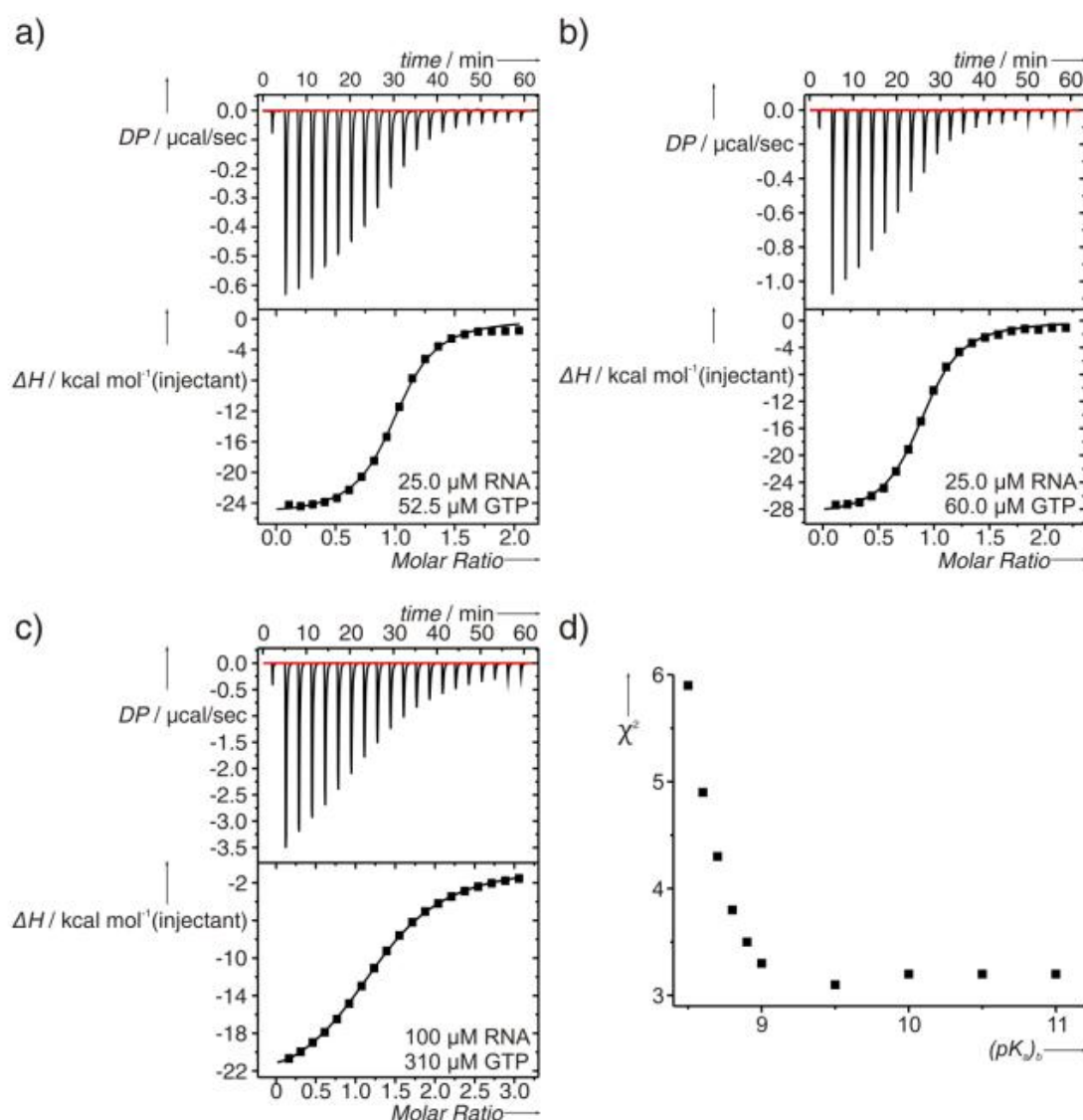


Supplementary Fig. 4. Binding of GTP to the class II aptamer RNA. (a) HNN-COSY spectrum recorded at 10°C, pH 6.3, 2 mM Mg²⁺ for a sample containing ¹⁵N-labeled RNA and ¹³C/¹⁵N-labeled GTP demonstrating directly the presence of an intermolecular hydrogen bond between the GTP H1N1 imino group and cytidine C17 N3 of the aptamer. Imino donor group signals are shown in black, signals corresponding to the acceptor nitrogens are colored gray. Correlations are indicated by dashed lines and the assignments are indicated. The correlation due to the intermolecular hydrogen bond between the imino group of GTP and the N3 nitrogen of C17 is highlighted by a red dashed line. (b) Slices of a ¹H,¹⁵N-NOESY-HSQC spectrum recorded for a sample containing ¹³C/¹⁵N-labeled GTP and unlabeled aptamer RNA at 10°C, pH 6.3, 2 mM Mg²⁺ showing intermolecular NOEs originating from the H1N1 imino and the H2N2 amino group protons of the ligand GTP. Assignments are indicated. (c) The GTP-binding site in the apical loop. A local superimposition of the 10 structures with the lowest target function is shown. GTP is colored dark red. The carbon atoms of C17 and A22 forming an intermolecular base triple with the bound GTP are colored yellow. The carbon atoms of the nucleotides G12, A16 and C23 forming an intramolecular base triple directly below are colored green. Hydrogen bonds are indicated by dashed lines. For better orientation the schematic representation of the structure using the Leontis-Westhof notation is shown on the right in the same color code. (d) The G18:A20 base pair stabilized by a single hydrogen bond forms the lid of the GTP-binding site and stacks on top of the intermolecular GTP:C17:A22 intermolecular base triple. (e) Stacking between the intramolecular G12:A16:C23 base triple forming the floor of the GTP-binding site and the intermolecular GTP:C17:A22 base triple.



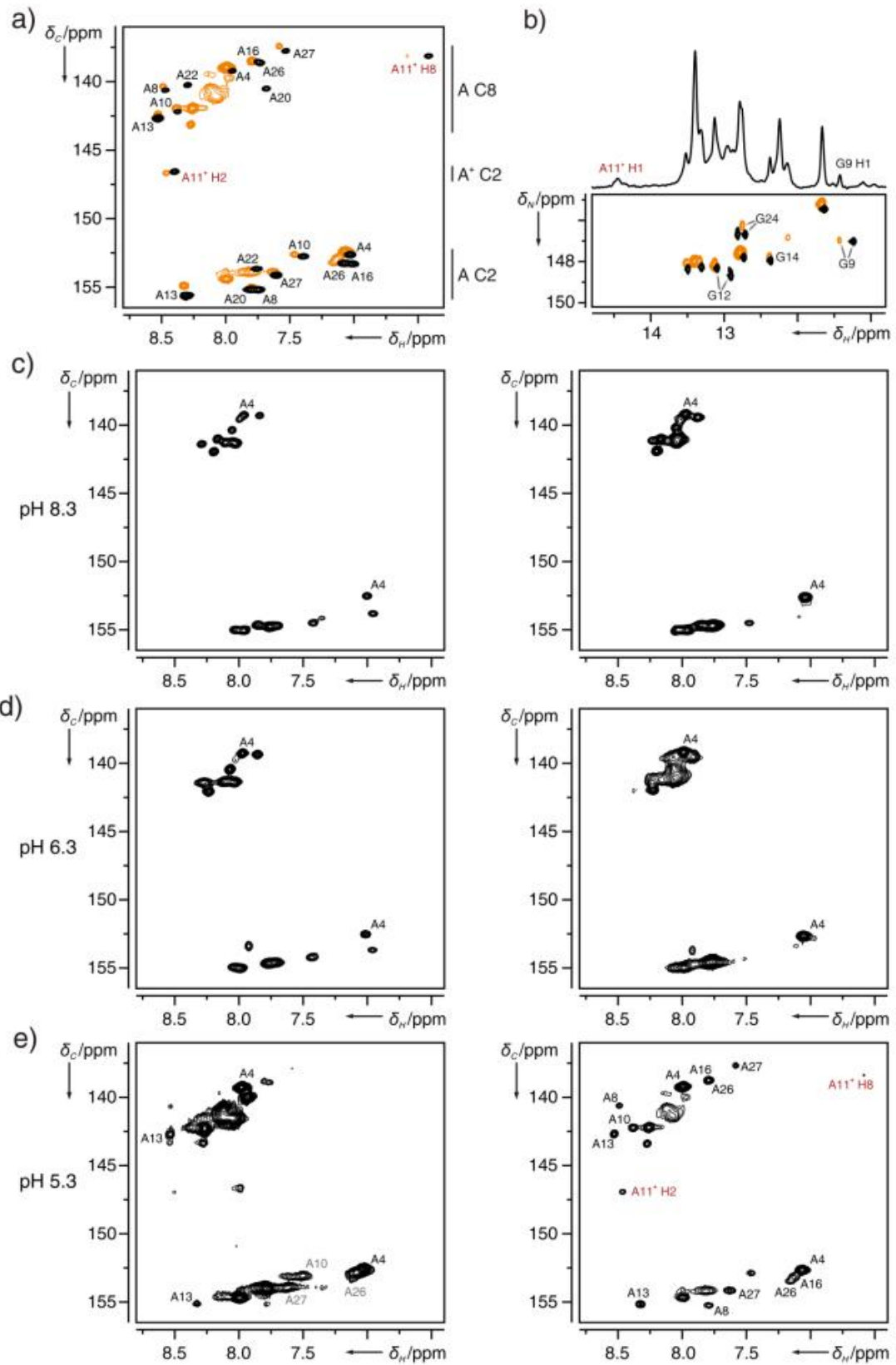
Supplementary Fig. 5. The G9:A11:C15:G24 base quartet. (a) Schematic representation of the structure and hydrogen bonding interactions in the base

quartet. Experimentally observed hydrogen bonds (HNN-COSY and ^1H , ^{31}P -long-range correlation experiments) are shown as thick dashed black lines. Key experimental NOEs are indicated by orange dashed lines. (b) Two-bond ^1H , ^{15}N -HSQC spectrum recorded for a ^{15}N -labeled RNA showing correlations between H2 and the N1 and N3 nitrogens as well as between H8 and the N7 and N9 nitrogens, respectively.^[19] The signal for the protonated N1 nitrogen of A11⁺ is shifted upfield by ~ 70 ppm compared to non-protonated adenine N1 nitrogen nuclei as expected for an imino nitrogen. The N7 and N9 nitrogen chemical shifts of the protonated A11 are also clearly outside the standard chemical shift ranges for these nitrogen nuclei. (c) 2D H(N)C-correlation experiment connecting the H1 imino proton and the H61/H62 amino protons to the C6 carbon of the adenine ring in A11 according to the magnetization transfer schema. This experiment explicitly rules out a N7 or a N3 protonation for A11. (d) ^1H , ^{13}C -HSQC spectrum recorded for a $^{13}\text{C}/^{15}\text{N}$ -A selectively labeled sample of the GTP-RNA complex. The signals for the H2C2 and H8C8 of A11 are indicated to highlight their unusual chemical shifts in comparison to the bulk of non-protonated A nucleotides in the complex. (e) Overlay of a HNN-COSY-spectrum showing the correlation signals between the amino group of the protonated A11 and the N7 nitrogens of both G9 and G24 (red) and the two-bond ^1H , ^{15}N -HSQC spectrum with the correlations between H8 and N7 of G9 and G24, respectively. (f) Titration of an RNA sample containing a G9A mutation abolishing the formation of the hydrogen bond between the G9 imino group and the backbone phosphate 5' of G24 with GTP. While small spectral changes indicate weak GTP binding, no ligand-induced folding of the RNA tertiary structure is observed. (g) Titration of an RNA sample containing a thiophosphate modification of the backbone phosphate 5' of G24 with GTP. The RNA is a ~ 50:50 racemic mixture containing the thiophosphate substitution either at the proR- or the proS-oxygen position. Even at high excess of GTP only approximately one half of the RNA is able to fold and bind GTP, suggesting that for one of the two stereoisomers the formation of the hydrogen bond to the G9 imino group and subsequent tertiary structure formation is impaired. (h) Stabilizing stacking interactions between the G9:A11:C15:G24 base quartet (carbon atoms colored light blue) and the G12:A16:C23 base triple (carbon atoms colored dark green) forming the floor of the GTP-binding site. All spectra shown in this figure were measured at 20°C, pH 6.3, 2 mM Mg^{2+} .



Supplementary Fig. 6. pH-dependent GTP-binding to the class II aptamer RNA. Shown are representative ITC-thermograms and the fitted binding curves using a one-site binding model for titrations at pH 5.8 (a), pH 6.3 (b) and 8.3 (c). The concentrations of RNA and GTP used for the individual titrations at each pH-value were optimized individually to obtain ITC-thermograms with plateaus at the beginning and the end of the titrations and are given as insets in the fitted binding curves. (d) χ^2 values for fits describing the pH-dependence of the K_D for GTP as a function of pK_a (free) and pK_a (bound) calculated with fixed values for pK_a (bound) (see Figure 4c in the main text). When both pK_a values are treated as free variables in the fit simultaneously, a pK_a (free) of 6.7 and a pK_a (bound) value of 8.9 is obtained (see Figure 4c in the main text). When the pK_a (bound) value is fixed iteratively to values

between 8.5 to 11 and the pK_a (free) is used as the only free variable, very similar pK_a (free) values (~ 6.7) are obtained whereas the χ^2 values describing the quality of the fit become smaller until a pK_a (bound) of 8.9 is reached and stay constantly low for pK_a (bound) values > 8.9 . Thus, we consider 8.9 only as a lower limit for the pK_a (bound) of the RNA in the GTP-bound form. The values obtained in this way from ITC are in qualitative agreement with the NMR-results. $^1H, ^{13}C$ -HSQC spectra recorded for the RNA-GTP complex at pH 8.3 do not show any evidence for either a second set of signals for a species with a deprotonated A11 or for line broadening for the H2C2 cross peak of A11⁺ (Figure 4b in the main text). This suggests that at pH 8.3 the population of the deprotonated A11 must be very low and thereby the pK_a for the protonation of A11 must be significantly higher than 8.3 in agreement with the ITC-based results. On the other hand, spectra of the free RNA show significant line broadening at pH 6.3 (Supplementary Fig. 7c, right) suggesting the simultaneous presence of a significant amount of protonated/folded and deprotonated/unfolded conformers at the same time. This is exactly what would be expected at pH values close to the pK_a (6.7) for the free RNA as determined from the ITC measurements at different pH values. In principle, NMR-spectra recorded at pH values > 8.3 or K_D -measurements by ITC at pH values > 8.3 would help to determine the pK_a for the GTP-bound RNA. However, the pK_a 's for G and U imino protons that are not involved in base pairing interactions and/or hydrogen bonds are ~ 9.2 . pH values > 8.3 during either the NMR or the ITC measurements would therefore lead to a beginning deprotonation of such G and U residues and the observed effects could no longer be attributed to A11⁺ deprotonation alone. Unfortunately, the structure of the complex shows that for instance G18 and U19 which have such imino groups are in the direct vicinity of the ligand (Supplementary Fig. 4c, 4d).



Supplementary Fig. 7. pH- and Mg^{2+} -dependent folding of the ligand-free class II aptamer RNA. (a) Overlay of the aromatic region of $^1H, ^{13}C$ -HSQC spectra for a $^{13}C/^{15}N$ -A selectively labeled sample at pH 6.3, 20°C, in the presence of 2 equivalents unlabeled GTP and 2 mM Mg^{2+} with complete assignments given (black) and a $^{13}C/^{15}N$ -A selectively labeled RNA in the presence of 2 mM Mg^{2+} at pH 5.3, 20°C, without GTP (orange). Chemical shifts of the GTP-complex and the free RNA at low pH are very similar for all A nucleotides with the exception of A20 and A22 (part of the intermolecular base triple with GTP) from the apical loop. (b) (top) The 1H -1D spectrum (top) of the unlabeled RNA in the absence of GTP at pH 5.3, 20°C, shows imino proton signals for the A11⁺ imino group as well as for the G9 imino proton involved in a tertiary hydrogen bonding interaction with the 5' backbone phosphate group of G24 and part of the G9:A11:C15:G24 base quartet (see Figure 3a). (bottom) An $^1H, ^{15}N$ -HSQC-spectrum of a G-only ^{15}N -labeled RNA in the absence of GTP at 20°C, pH 5.3, 2 mM Mg^{2+} , (dark orange) reveals the presence of imino group signals for G9 (see above), G14 and G24 from the short central helix as well as for G12 which is part of the G12:C23 long-range Watson-Crick base pair and the G12:A16:C23 base triple (see Figure 2c). The corresponding spectrum for the GTP-RNA complex at pH 6.3, 20°C, 2 mM Mg^{2+} (black) is shown for comparison. Taken together with the chemical shift similarities observed for the aromatic CH-groups (Supplementary Figure 7a,e) the presence of these imino group signals indicate that the central bulge, the G9:A11:C15:G24 base quartet and the G12:A16:C23 base quartet are formed under these conditions even in the absence of ligand. (c) Aromatic region of $^1H, ^{13}C$ -HSQC spectra for a $^{13}C/^{15}N$ -A selectively labeled sample recorded at pH 8.3, 20°C in the absence (left) and the presence (right) of 2 mM Mg^{2+} . (d) Aromatic region of $^1H, ^{13}C$ -HSQC spectra for a $^{13}C/^{15}N$ -A selectively labeled sample recorded at pH 6.3, 20°C in the absence (left) and the presence (right) of 2 mM Mg^{2+} . In the presence of Mg^{2+} , substantial line broadening is observable in agreement with conformational exchange. (e) Aromatic region of $^1H, ^{13}C$ -HSQC spectra for a $^{13}C/^{15}N$ -A selectively labeled sample recorded at pH 5.3, 20°C in the absence (left) and the presence (right) of 2 mM Mg^{2+} .

3. Materials & Methods

RNA preparation.

Unlabeled and isotopically labeled GTP class II aptamer RNA (5'-GGC AGC CAG AAG AGC ACG UAU ACG CAA GGC UGU C-3') as well as the unlabeled G9A mutant was synthesized by *in vitro* transcription with T7 RNA-polymerase and linearized plasmid-DNA as template using unlabeled (SigmaAldrich) and appropriately isotope labeled, commercially available nucleotide triphosphates (Silantes GmbH). In order to generate uniform 3'-ends, the primary RNA-transcripts contained a hammerhead ribozyme. Processed RNA transcripts were purified by preparative denaturing PAGE according to standard protocols. The purified RNA was folded by heating to 95°C for 10 minutes prior to injection into 2 equivalents of ice cold water. 10 mM EDTA (pH 8.0) was added to the solution to remove residual Mg²⁺. The samples were then exchanged into NMR buffer containing 25 mM potassium phosphate (pH 6.3) and 25 mM potassium chloride and concentrated using Vivaspin concentrators (MW cutoff 3 kDa). RNA concentrations in the NMR samples varied between 0.33 and 1.2 mM. Samples used for structure determination contained either 5%-10% or 100 % (v/v) D₂O and 2 mM magnesium acetate as well as ~ 2 equivalents of either ¹⁵N- or ¹³C/¹⁵N-labeled GTP.

For the assignment and structure calculation experiments of the GTP-RNA aptamer complex an unlabeled, a ¹⁵N-labeled, a ¹³C/¹⁵N-labeled and base type selectively ¹³C/¹⁵N-A, ¹³C/¹⁵N-C, ¹³C/¹⁵N-GU and ¹⁵N-AG labeled samples containing appropriately labeled ¹⁵N- or ¹³C/¹⁵N-GTP were used. Moreover, samples carrying site specific ¹³C-C8 or ¹³C-C6 labels of individual nucleotides (C15/A26, C25/A16, C17/A22, C23/A13, A10/U19, A20/U21, G9, G12, A11/G14, A8/G24) were chemically synthesized.^[20] The G24-S thiophosphate modified RNA was purchased from GE Dharmacon, deprotected according to the manufacturer's protocol and afterwards folded and transferred into NMR buffer as described.

NMR spectroscopy.

NMR experiments for obtaining resonance assignments have been described in detail before.^[21] Experiments for obtaining NMR distance, torsional and hydrogen bonding restraints for the structure calculations of the GTP-RNA aptamer complex were recorded at 10 °C and 20 °C in 5% D₂O/95% H₂O (v/v) for the exchangeable

protons and in 100% (v/v) D₂O at 20 °C for the non-exchangeable protons. All NMR spectra were collected on Bruker AVANCE 600, 700, 800, 900 and 950 MHz spectrometers equipped with either room temperature or cryogenic triple resonance HCN- or HCP-probes and z-axis gradients. ¹H-¹³C heteronuclear NOE data were collected at 20 °C in an interleaved manner with and without ¹H saturation periods of 5 sec. NMR data were processed with TOPSPIN2.1 to TOPSPIN3.5 software (Bruker Biospin). Spectra for the assignment of the RNA-ligand complex were analyzed in CARA.^[22] Proton chemical shifts were referenced directly to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Carbon, nitrogen and phosphorous chemical shifts were referenced indirectly.^[23]

Input restraints.

Intra- and intermolecular ¹H,¹H distance restraints involving exchangeable RNA and ligand protons were extracted from NOE intensities of 2D ¹H,¹H-NOESY and 2D ¹⁵N-CPMG-HSQC-NOESY experiments recorded for a ¹⁵N-labeled sample containing ¹³C/¹⁵N-GTP at 20°C using WATERGATE water suppression and GARP decoupling.^[24] Interproton distance restraints involving aromatic CH-groups of the RNA and the ligand were derived from NOE intensities measured in 2D and 3D ¹H,¹³C-NOESY-HSQC spectra in 5% D₂O/ 95% H₂O (v/v) optimized for aromatic CH ¹J_{CH} coupling constants recorded for samples containing base type selectively labeled RNA or unlabeled RNA and ¹³C,¹⁵N-GTP. 3D ¹H,¹³C-NOESY-HSQC spectra in 100% D₂O optimized for aliphatic CH moieties were recorded on base type selectively labeled samples to measure NOE intensities involving aliphatic CH-groups. A 2D ¹³C F2-edited ¹H,¹H-NOESY of a sample containing unlabeled RNA and ¹³C/¹⁵N-labeled GTP in H₂O using Echo-Antiecho water suppression allowed the observation of NOEs between the GTP ribose and exchangeable RNA protons.^[25] Intermolecular ¹H,¹H distance restraints involving the GTP amino and imino group protons to the RNA were extracted from a 2D ¹⁵N F2-edited ¹H,¹H-NOESY experiment of an unlabeled sample containing ¹³C/¹⁵N-labeled GTP. ¹³C F1-filtered, ¹³CF3-edited 2D and 3D NOESY-HSQC experiments using base type selectively labeled samples yielded additional internucleotide NOE intensities.^[26] All NOESY experiments were carried out with a mixing time of 120 ms.

For the determination of the ribose sugar pucker, ³J_{H1',H2'} scalar couplings were measured in 3D HCCH-TOCSY-CCH-E.COSY experiments recorded with base type

selectively labeled samples.^[27] Furthermore, the canonical coordinate *can1* derived from the ribose carbon C1', C4' and C5' ¹³C chemical shifts was calculated.^[17] All nucleotides with ³J_{H1',H2'} scalar couplings < 2.5 Hz were constrained to a C3'-*endo* ribose conformation. All nucleotides with a canonical coordinate > 6.8 were restrained to a C2'-*endo* conformation. All other nucleotides were left with an unrestrained ribose conformation (see Supplementary Fig. 2).

For the experimental identification of hydrogen bonds HNN-COSY experiments optimized for imino or amino hydrogen bond donor groups as well as long-range HNN-COSY and sellr-HNN-COSY experiments were recorded on appropriately labeled samples.^[18,28]

Structure calculation.

NOE cross-peak intensities were extracted from 2D NOESY and 2D or 3D NOESY-HSQC spectra and were translated into upper limit distance restraints by referencing to the intranucleotide 2.4 Å H5/H6 NOEs of fixed distance. NOE intensities of spectra without H5/H6 cross peaks of pyrimidines were calibrated by matching to corresponding intensities of identical NOEs of referenced spectra. 1.7 Å were added to all upper limit distance restraints. Distance constraints for hydrogen bonds were introduced for both canonical base pairs and amino group proton hydrogen bond donors exhibiting cross peaks in the HNN-COSY-spectra. For each A:U and G:U base pair, four upper and four lower and for each G:C base pair, including the C17:GTP intermolecular base pair, six upper and six lower distance restraints were used. Four lower and upper distance restraints were introduced for the A8:A27 base pair. For all other experimentally detected hydrogen bonds, e.g. A16-H62...G12-N3, two upper and lower distance restraints between the hydrogen bond donor and acceptor atoms were introduced. For conformationally rigid residues with a canonical ³¹P chemical shift (nucleotides 2-6, 15, 16, 25-33, see Supplementary Fig. 2) the backbone torsion angles α , β , γ , δ , ϵ and ζ were loosely restricted to standard A-form helical values ($\pm 30^\circ$). Residues A8 and G9 of the internal bulge and loop residues G18, U19 and A20 were set to a C2'-*endo* sugar pucker conformation, flexible residues A13 and U21 as well as the 3'-end residue C34 were left unconstrained. The remaining residues were set to C3'-*endo* conformation. The glycosidic torsion angle χ of residue G9 was deduced from the characteristic NOE patterns for a *syn* arrangement of the nucleobase. For A22, χ was set to "non-*syn*" to restrict the

nucleobase to a non-*syn* orientation consistent with NOE patterns in this region. Residues G1-C7, A11, G12, C14, G15, C17 and C23-C34 showed a clear NOE pattern indicative for *anti* orientation of the nucleobase, for the remaining residues A8, A10, 13, A16 and G18-U21 the χ angle remained unconstrained.

Structure calculations were performed using CYANA version 3.97. Energy minimization was performed with OPAL^[29] with the implemented AMBER94 force field^[30] for the 10 out of 100 conformers with the lowest target function. PyMol (DeLano Scientific LLC; Schrödinger, Inc) or MOLMOL^[31] were used for visualization of the structures.

Isothermal Titration Calorimetry (ITC).

Isothermal titration calorimetry (ITC) experiments were performed using a MicroCal iTC₂₀₀ instrument (Malvern Instruments, Malvern, UK). Samples of unlabeled RNA were prepared in buffer containing 25 mM potassium phosphate (pH 5.3 – 8.3), 25 mM potassium chloride and 2 mM magnesium acetate. The ligand GTP in the same buffer was injected into a solution of 25-100 μ M GTP class II aptamer RNA. All measurements were performed at 25°C. After an initial waiting time of 120 s, the first injection of 0.2 μ l was followed by 19 serial injections of 2 μ l, separated by an interval of 180 s. For each experiment, the reference power was set to 11 μ cal⁻¹, stirring speed to 750 rpm and the high feedback mode was selected. Three independent titrations were performed at each pH value. The thermograms were processed using Origin7.0 (OriginLab) assuming a one site binding model.

For determination of the pK_a values, the averaged K_D's were plotted against the corresponding pH value. The resulting data were fitted according to the equation given in Fig. 4c using Origin 2016G (OriginLab).

4. Supplementary References

- [1] P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1994**, *116*, 8390–8391.
- [2] P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1997**, *119*, 6621–6628.
- [3] Z. Cai, I. Tinoco, JR, *Biochemistry* **1996**, *35*, 6026–6036.
- [4] J. S. Smith, E. P. Nikonowicz, *Biochemistry* **1998**, *37*, 13486–13498.
- [5] S. Ravindranathan, S. E. Butcher, J. Feigon, *Biochemistry* **2000**, *39*, 16026–16032.
- [6] A. Huppler, L. J. Nikstad, A. M. Allmann, D. A. Brow, S. E. Butcher, *Nat. Struct. Biol.* **2002**, *9*, 431–435.
- [7] M. Guo, R. C. Spitale, R. Volpini, J. Krucinska, G. Cristalli, P. R. Carey, J. E. Wedekind, *J. Am. Chem. Soc.* **2009**, *131*, 12908–12909.
- [8] J. A. Liberman, M. Guo, J. L. Jenkins, J. Krucinska, Y. Chen, P. R. Carey, J. E. Wedekind, *J. Am. Chem. Soc.* **2012**, *134*, 16933–16936.
- [9] G. Desjardins, E. Bonneau, N. Girard, J. Boisbouvier, P. Legault, *Nucleic Acids Res.* **2011**, *39*, 4427–4437.
- [10] B. Houck-Loomis, M. A. Durney, C. Salguero, N. Shankar, J. M. Nagle, S. P. Goff, V. M. D'Souza, *Nature* **2011**, *480*, 561–564.
- [11] J. L. Wilcox, P. C. Bevilacqua, *Biochemistry* **2013**, *52*, 7470–7476.
- [12] M. Pechlaner, D. Donghi, V. Zelenay, R. K. O. Sigel, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 9687–9690.
- [13] M. Kosutic, S. Neuner, A. Ren, S. Flur, C. Wunderlich, E. Mairhofer, N. Vusurovic, J. Seikowski, K. Breuker, C. Hobartner et al., *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 15128–15133.
- [14] A. Ren, N. Vusurovic, J. Gebetsberger, P. Gao, M. Juen, C. Kreutz, R. Micura, D. J. Patel, *Nat. Chem. Biol.* **2016**, *12*, 702–708.
- [15] W. Saenger, *Principles of nucleic acid structure*, Springer, New York, **1988**, p. 108.
- [16] J. L. Wilcox, P. C. Bevilacqua, *J. Am. Chem. Soc.* **2013**, *135*, 7390–7393.
- [17] a) P. Rossi, G. S. Harbison, *J. Magn. Reson.* **2001**, *151*, 1–8; b) O. Ohlenschläger, S. Haumann, R. Ramachandran, M. Görlach, *J. Biomol. NMR* **2008**, *42*, 139–142.
- [18] a) B. Luy, J. P. Marino, *J. Am. Chem. Soc.* **2000**, *122*, 8095–8096; b) M. Hennig, J. R. Williamson, *Nucleic Acids Res.* **2000**, *27*, 1585–1593.
- [19] V. Sklenár, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1994**, *4*, 117–122.
- [20] C. H. Wunderlich, R. Spitzer, T. Santner, K. Fauster, M. Tollinger, C. Kreutz, *J. Am. Chem. Soc.* **2012**, *134*, 7558–7569.
- [21] A. C. Wolter, E. Duchardt-Ferner, A. H. Nasiri, K. Hantke, C. H. Wunderlich, C. Kreutz, J. Wöhnert, *Biomol. NMR Assign.* **2016**, *10*, 101–105.
- [22] R. Keller, *The computer aided resonance assignment tutorial.*, CANTINA Verlag, Goldau, **2004**.
- [23] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. Dyson, E. Oldfield, J. L. Markley, B. D. Sykes, *J. Biomol. NMR* **1995**, *6*, 135–140.
- [24] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 12593–12594.
- [25] T.-L. Hwang, A. J. Shaka, *J. Magn. Reson.* **1995**, *112*, 275–279.
- [26] C. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat, L. E. Kay, *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
- [27] H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1995**, *117*, 7251–7252.

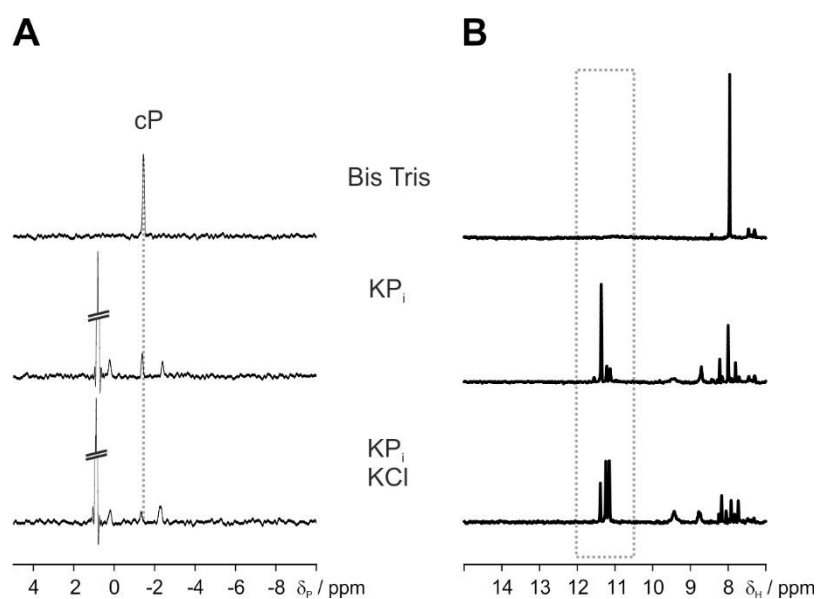
- [28] a) A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297; b) A. Dallmann, B. Simon, M. M. Duszczuk, H. Kooshapur, A. Pardi, W. Bermel, M. Sattler, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 10487–10490.
- [29] P. Luginbühl, P. Güntert, M. Billeter, K. Wüthrich, *J. Biomol. NMR* **1996**, *8*, 136–146.
- [30] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [31] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51–55.

Supplementary Material

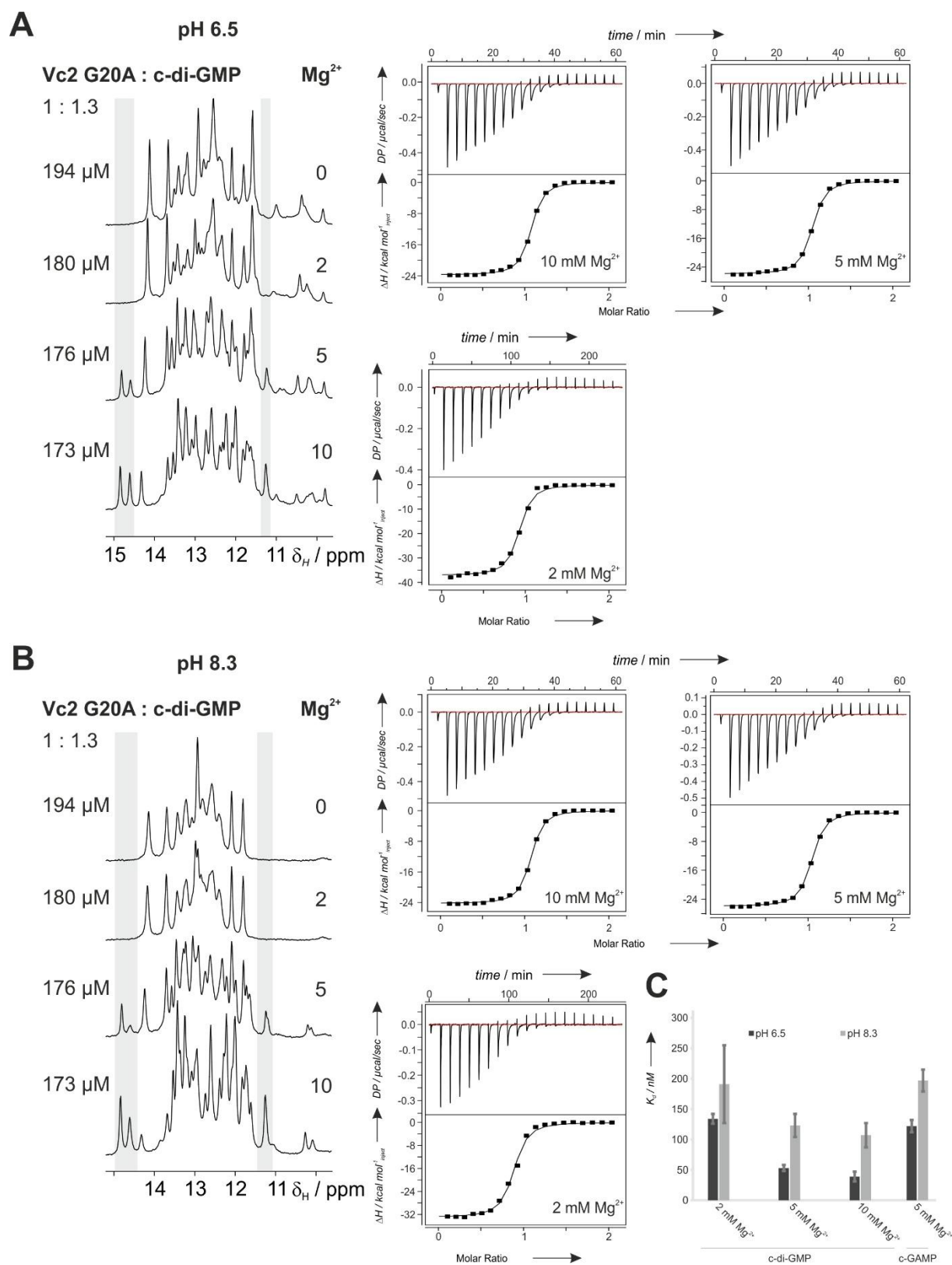
Adenine protonation enables c-di-GMP binding to cyclic-GAMP sensing riboswitches

Heiko Keller¹, A. Katharina Weickhmann¹, Thomas Bock¹, Jens Wöhnert^{1,*}

Supplementary Figures.



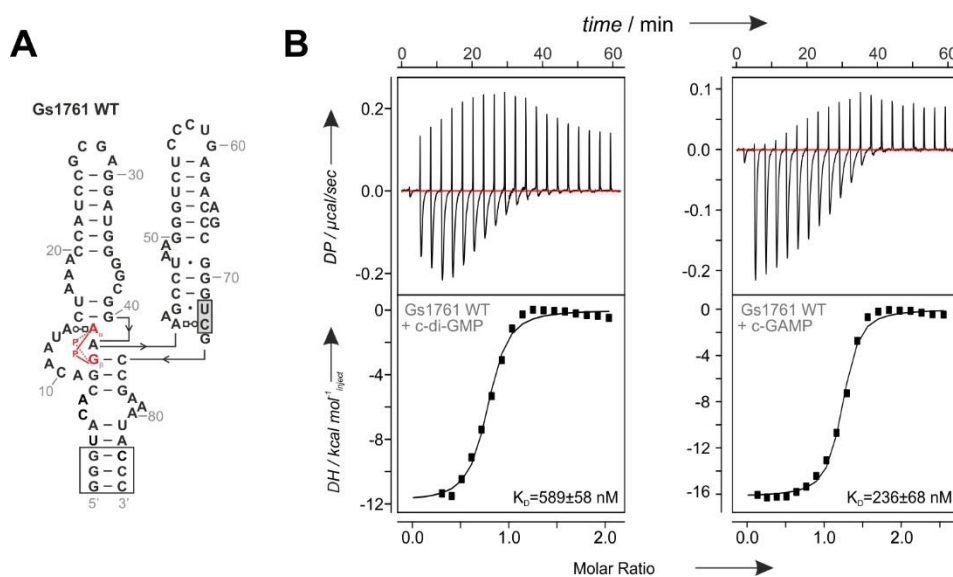
Supplementary Figure S1. Potassium ion-containing buffers induce G-quadruplex and oligomer formation by c-di-GMP at concentrations required in NMR- and ITC studies. (A) $1\text{D-}^{31}\text{P}$ spectra and (B) imino and aromatic region of $1\text{D-}^1\text{H}$ spectra of $450\ \mu\text{M}$ c-di-GMP at $25\ ^\circ\text{C}$ in $50\ \text{mM}$ BisTris buffer, pH 6.5 (top), $25\ \text{mM}$ KP_i buffer, pH 6.5 (middle) or $25\ \text{mM}$ KP_i buffer, pH 6.5, with $250\ \text{mM}$ KCl (bottom). The concentration of c-di-GMP in these samples ($450\ \mu\text{M}$) corresponds to concentrations used in the titrant solution of ITC measurements. The ${}^{31}\text{P}$ resonance frequency of the c-di-GMP cyclic phosphate group in monomeric c-di-GMP is highlighted with a dashed line in the ${}^{31}\text{P}$ spectra and labeled cP. The ${}^{31}\text{P}$ resonance of the phosphate buffer appears at a chemical shift of $\sim 1\ \text{ppm}$. The chemical shift region typical for imino resonances of c-di-GMP G-quadruplexes and oligomers as discussed in Gentner *et al.* (2012) is boxed in (B).



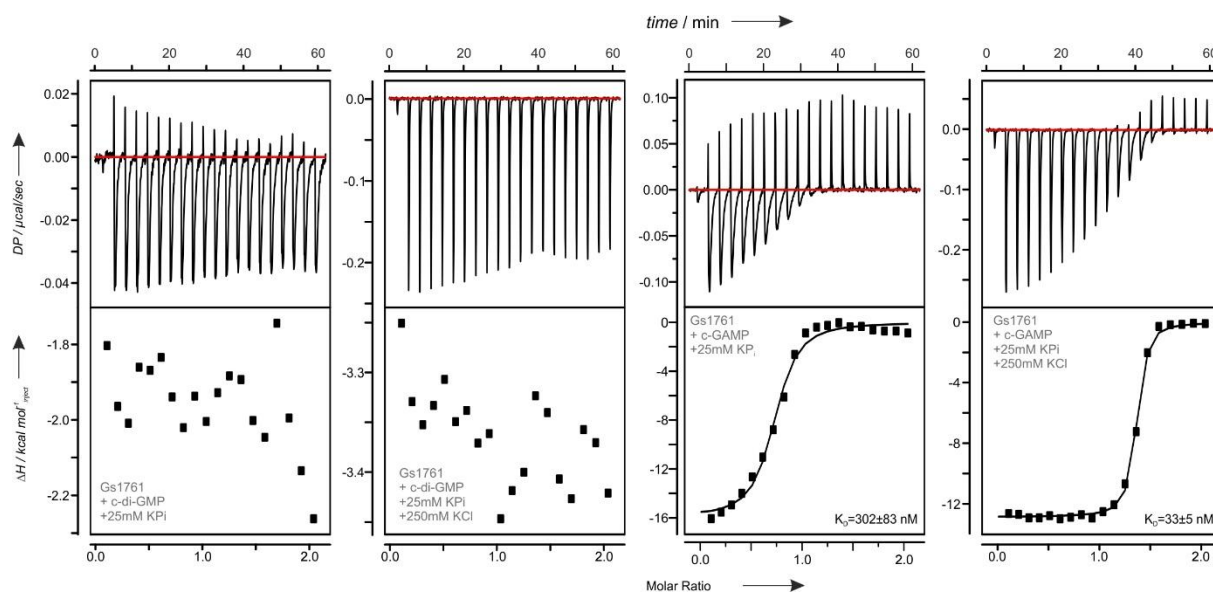
Supplementary Figure S2. High Mg²⁺-concentrations and low pH both promote c-di-GMP binding to the Vc2 G20A riboswitch. (A) *left* Imino proton spectra of Vc2 G20A with 1.3 fold

excess of c-di-GMP without Mg^{2+} (top) and in the presence of 2 mM (middle), 5 mM (middle) or 10 mM Mg^{2+} (bottom) in 50 mM BisTris (pH 6.5). RNA concentrations are indicated on the left. The changes in the RNA concentrations are due to dilution effects upon Mg^{2+} addition. Shaded areas in the spectra highlight chemical shift regions with pronounced spectral changes. *right*

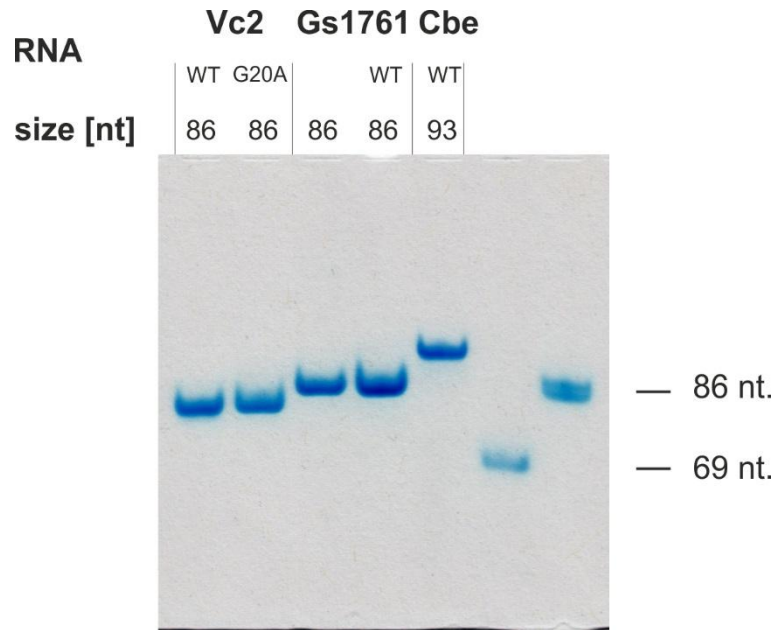
Representative ITC thermograms and fits for the Vc2 G20A riboswitch RNA binding to c-di-GMP in the presence of the indicated Mg^{2+} -concentration in BisTris buffer at pH 6.5. (B) same as in (A) but measurements were performed in 50 mM Tris/HCl buffer at pH 8.3. (C) Bar graph comparing the K_D 's for c-di-GMP and c-GAMP binding to the Vc2 G20A riboswitch aptamer domain under the conditions indicated. See also Supplementary Table 2.



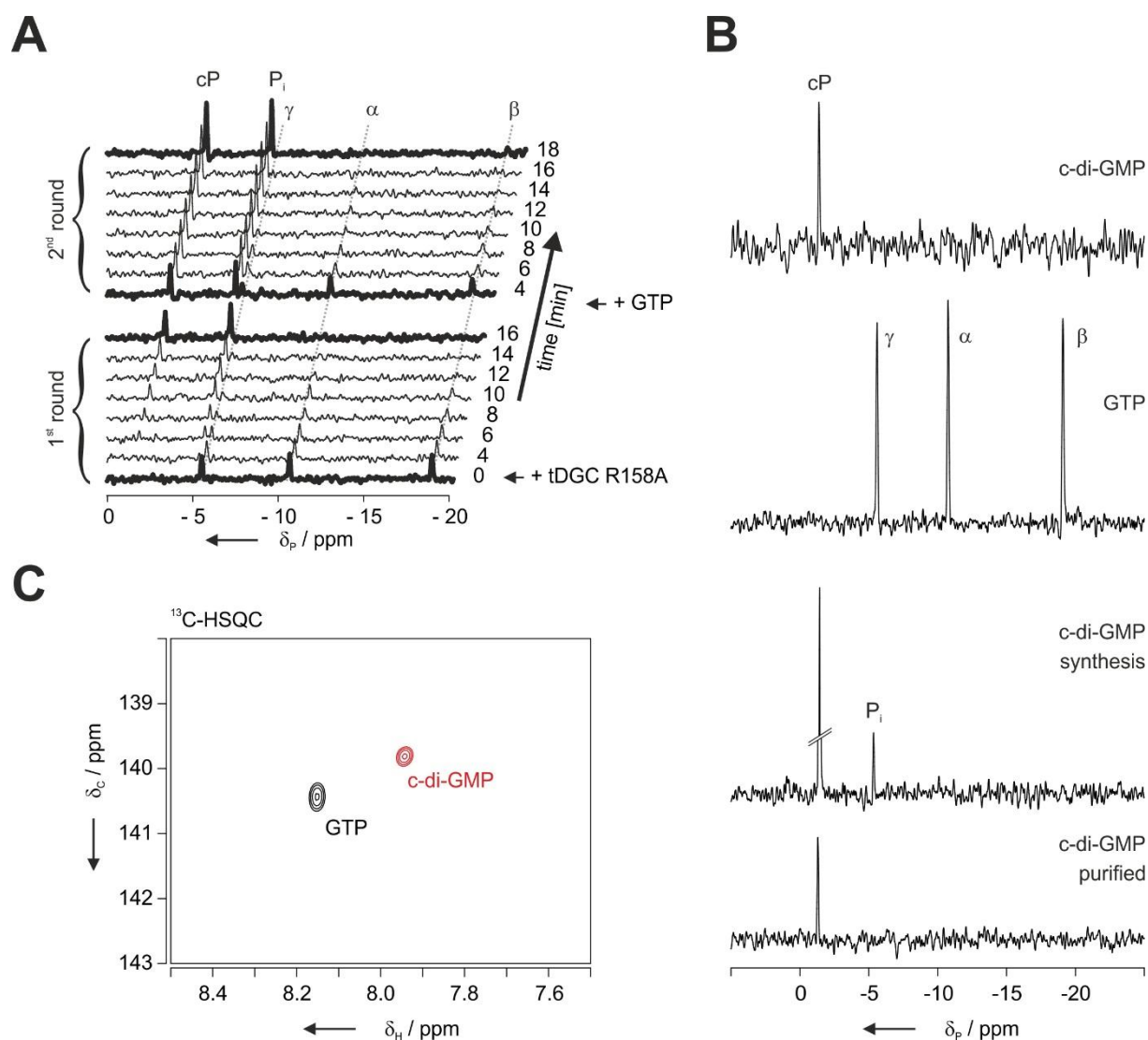
Supplementary Figure S3. Ligand binding properties of Gs1761 WT. (A) Secondary structure of the Gs1761 WT RNA construct used in this study. The two nucleotides that differ between Gs1761 WT and Gs1761 are highlighted with a grey box. As in all other constructs used in this study three G-C base pairs close the P1 stem (open box). (B) Representative ITC thermograms and fits for Gs1761 WT riboswitch RNA binding to c-di-GMP (left) and c-GAMP (right), respectively, in 50 mM BisTris (pH 6.5), 5 mM magnesium acetate. The resulting K_D s are given for both complexes.



Supplementary Figure S4. Effect of potassium ions on ligand binding by Gs1761. Representative ITC thermograms for Gs1761 riboswitch RNA titrated with c-di-GMP in 25 mM KP_i buffer (pH 6.5, left) or in 25 mM KP_i buffer (pH 6.5), 250 mM KCl (middle left) and titrated with c-GAMP in 25 mM KP_i buffer (pH 6.5, middle right) or in 25 mM KP_i buffer (pH 6.5), 250 mM KCl (right). All buffers contained 5 mM magnesium acetate. All measurements were carried out at 25 °C. No binding is observable for c-di-GMP under this conditions and fits are only shown for c-GAMP binding. The resulting K_{DS} are indicated.



Supplementary Figure S5. Native gel electrophoresis of Vc2 (lane 1), Vc2 G20A (lane 2), Gs1761 (lane 3), Gs1761 WT (lane 4) and Cbe 1-2 (lane 5) RNAs after purification and refolding using 10 % polyacrylamide gels containing 89 mM Tris (pH 8.0), 89 mM boric acid, 2.5 mM EDTA and 5 mM magnesium acetate. Two unrelated hairpin forming RNAs with sizes of 69 (lane 6) and 86 (lane 7) nucleotides were used as molecular size markers. Approximately 2 μ g RNA were loaded per lane. Gels were run at room temperature. RNAs were visualized by methylene blue staining.



Supplementary Figure S6. Enzymatic synthesis and purification of $^{13}\text{C},^{15}\text{N}$ -labeled c-di-GMP.

(A) Representative 1D- ^{31}P spectra recorded sequentially at 45 °C directly in TM1788 reaction buffer (see Material & Methods) during multiple cycles of a c-di-GMP synthesis. Conversion of GTP into c-di-GMP is completed after 15 minutes. The next round of synthesis is then reinitiated by the addition of fresh GTP. The time point for each spectrum is indicated with respect to the start of the reaction (1st round) or the last addition of GTP (2nd round). The chemical shifts of the ^{31}P signals of the α -, β - and γ -phosphate groups of GTP, the cyclic phosphate (cP) in c-di-GMP and the inorganic pyrophosphate (P_i) accumulating during synthesis are labeled. (B) ^{31}P spectra of commercially available c-di-GMP (top, SigmaAldrich), $^{13}\text{C},^{15}\text{N}$ -GTP (middle), $^{13}\text{C},^{15}\text{N}$ -c-di-GMP at

the end of the synthesis (middle) and after purification (bottom) recorded at 25 °C in TM 1788 reaction buffer. NMR signals are labeled as in (A). The intensity of the signal for inorganic pyrophosphate is low compared to the signal for the cyclic phosphate of c-di-GMP due to precipitation of magnesium pyrophosphate during synthesis. (C) Overlay of the ^{13}C -HSQC spectra recorded at 25 °C in TM1788 reaction buffer optimized for the H8C8 moiety for $^{13}\text{C},^{15}\text{N}$ – labeled GTP (black, 1 mM) before the cyclization reaction and $^{13}\text{C},^{15}\text{N}$ –labeled c-di-GMP at the end of the synthesis (red, 2.5 mM) indicating complete conversion of GTP to c-di-GMP.

RNA	Ligand	K_D [nM]	Selectivity factor c-GAMP/c-di-GMP	N	ΔH [kcal/mol]	ΔS [cal/mol/deg]
Vc2 WT	c-di-GMP	*				
	c-GAMP	$10,469 \pm 14$ 3		1.63 ± 0.0 4	-13 ± 1	-22 ± 2
Vc2 G20A	c-di-GMP	53 ± 5	0.4	0.9 ± 0.1	-42 ± 4	-109 ± 13
	c-GAMP	128 ± 2		1.04 ± 0.0 2	-31 ± 2	-74 ± 6
Gs1761	c-di-GMP	693 ± 107	2.4	1.1 ± 0.1	-7.0 ± 0.3	-5 ± 1
	c-GAMP	286 ± 22		1.4 ± 0.1	- 15.0 ± 0.3	-20 ± 1
Gs1761 WT	c-di-GMP	589 ± 58	2.5	0.76 ± 0.0 4	-11 ± 2	-7 ± 7
	c-GAMP	236 ± 68		1.2 ± 0.1	-15 ± 1	-22 ± 3
Cbe	c-di-GMP	30 ± 1	0.06	0.8 ± 0.1	-20 ± 1	-34 ± 3
	c-GAMP	463 ± 45		0.9 ± 0.1	-18 ± 2	-30 ± 6

Supplementary Table 1. Overview of all ITC-experiments carried out in BisTris-buffer, pH 6.5 in the presence of 5 mM Mg^{2+} at 25 °C and the derived c-GAMP and c-di-GMP binding parameters for all RNA-constructs used in this study.

Vc2 G20A + c-di-GMP						
Mg^{2+} [mM]	pH	K_d [nM]	N	ΔH [kcal/mol]	ΔS [cal/mol/deg]	Spacing time
10	6.5	39 ± 8	1.1 ± 0.2	-28 ± 2	-61 ± 7	180 sec
	8.3	107 ± 20	1.03 ± 0.04	-24.4 ± 0.2	-50 ± 1	
5	6.5	53 ± 5	0.9 ± 0.1	-42 ± 4	-109 ± 13	

	8.3	123 ± 19	1.0 ± 0.2	-25.4 ± 0.5	-54 ± 1	
2	6.5	134 ± 8	0.9 ± 0.1	-36 ± 1	-90 ± 3	720 sec
	8.3	191 ± 64	1.1 ± 0.2	-31 ± 4	-71 ± 12	

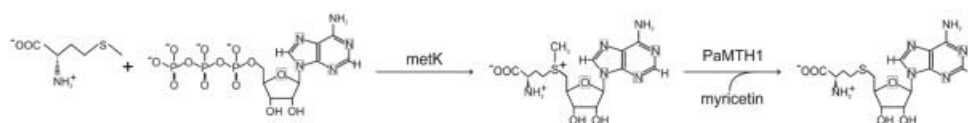
Supplementary Table 2. Overview of ITC-experiments and the derived c-di-GMP binding parameters for the Vc2 G20A mutant at variable Mg^{2+} concentrations at two pH values at 25 °C. At Mg^{2+} concentrations below 2 mM no binding was observable. At a Mg^{2+} concentration of 2 mM the kinetics of c-di-GMP was significantly slowed down. Thus, the spacing between individual ligand injections in the ITC had to be increased in order to reach equilibrium before the next injection.

3. SUPPLEMENTARY INFORMATION WEICKHMANN ET AL (STRUCTURE)

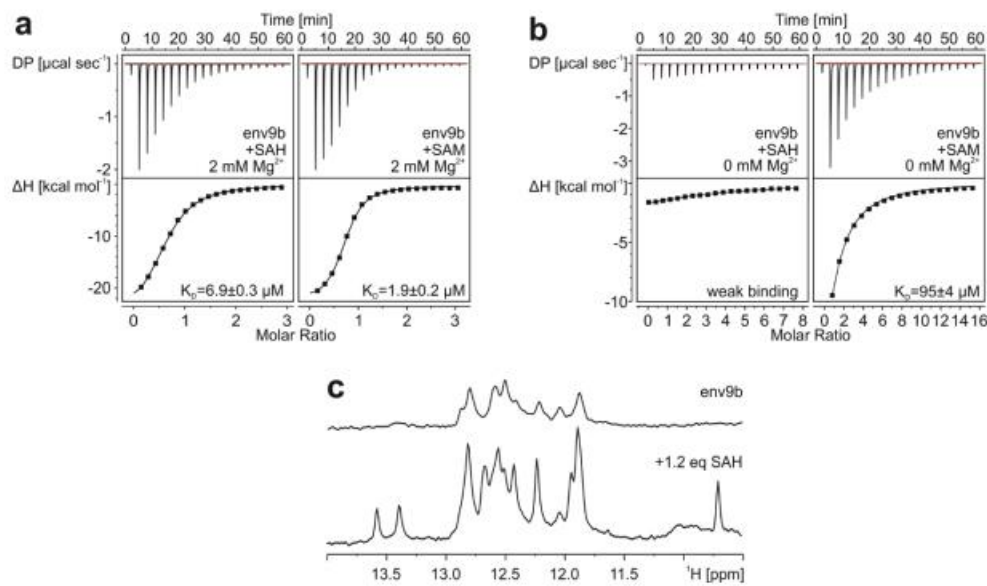
Supplementary Material

The structure of the SAM/SAH-binding riboswitch

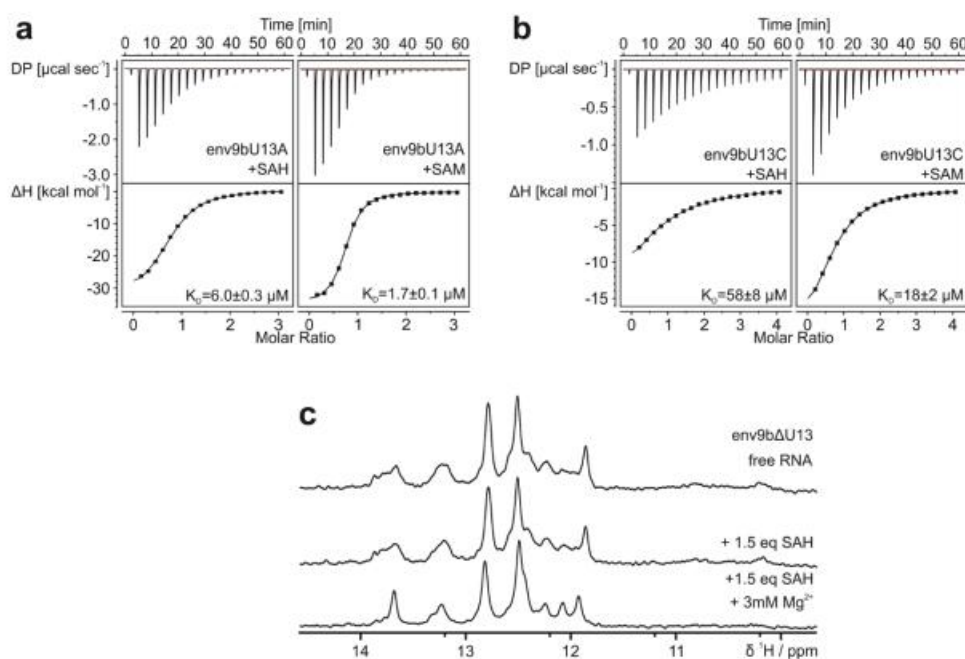
A. Katharina Weickmann¹, Heiko Keller¹, Jan Philip Wurm^{1,2}, Elisabeth Strebiter³, Michael A. Juen³, Johannes Kremser³, Zasha Weinberg⁴, Christoph Kreutz³, Elke Duchardt-Ferner¹, Jens Wöhnert¹



Supplementary Figure S1. Synthesis of isotope-labeled SAM and SAH. Methionine and adenosine triphosphate are used as substrates for the S-adenosylmethionine (SAM) synthetase MetK from *Escherichia coli* to synthesize SAM. The product SAM is then used in a methyl transfer reaction together with myricetin as methyl group acceptor by the methyl transferase *Podospira anserina* MTH1 (PaMTH1) to produce S-adenosylhomocysteine (SAH) and methyl-myricetin. Specific labeling schemes are introduced by using appropriately labeled ATP and methionine.

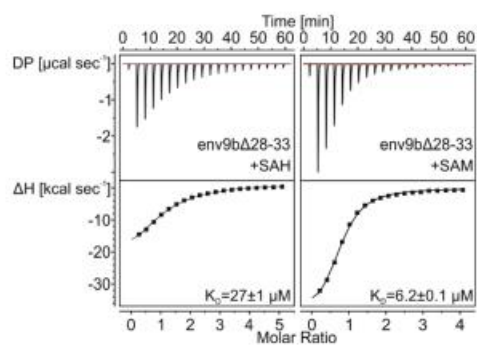


Supplementary Figure S2. The SAM/SAH-binding riboswitch binds both ligands, SAM and SAH, in the presence and absence of magnesium ions. (a)-(b) Representative ITC thermograms and fits for the env9b riboswitch RNA binding to SAH (*left*) or SAM (*right*) at a magnesium concentration of 2 mM (a) or without magnesium ions (b), respectively. The resulting K_D values are indicated. The binding affinity of SAH for the RNA in the absence of Mg^{2+} is too low to obtain a thermogram that can be used to derive a K_D . (c) Imino proton spectra of env9b without (*top*) and with a 1.2-fold excess of SAH (*bottom*) in the absence of magnesium ions clearly demonstrate SAH binding.

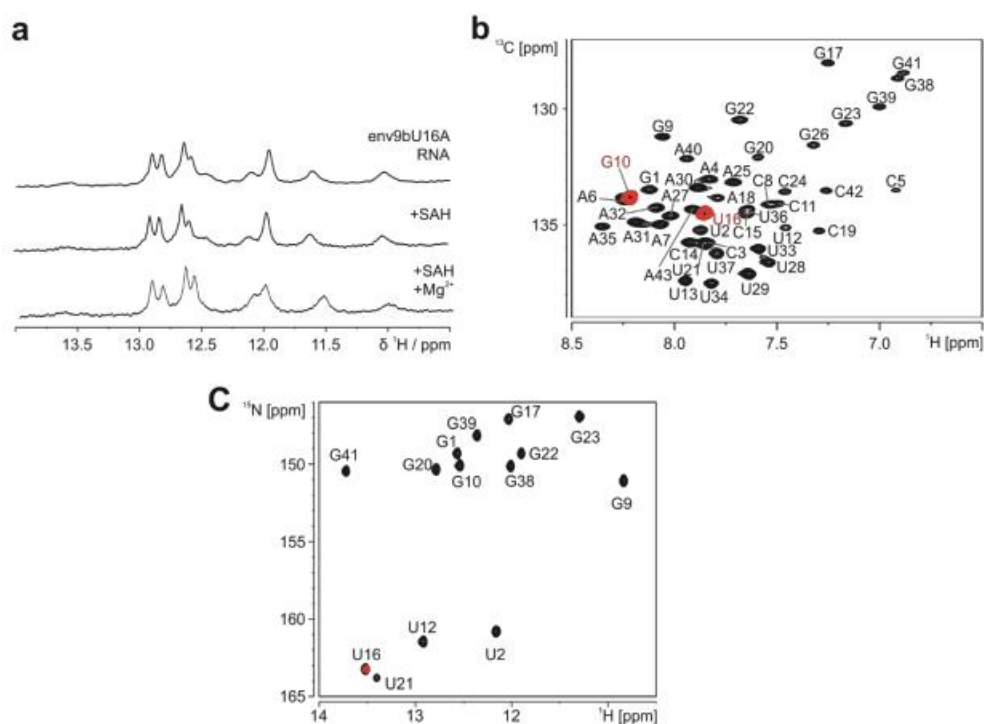


Supplementary Figure S3. The flexible nucleotide U13 acts as backbone spacer nucleotide.

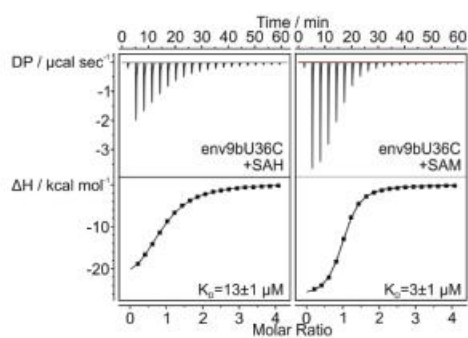
(a)-(b) Representative ITC thermograms for env9bU13A (a) and env9bU13C (b) riboswitch RNA titrated with SAH (*left*) or SAM (*right*) in the presence of 2 mM Mg $^{2+}$. The resulting K_D values are indicated. (c) Imino proton spectra of env9b Δ U13 (*top*) titrated with 1.5-fold excess of SAH (*middle*) and 3 mM magnesium acetate (*bottom*) show that the env9b Δ U13 mutant is no longer able to bind SAH.



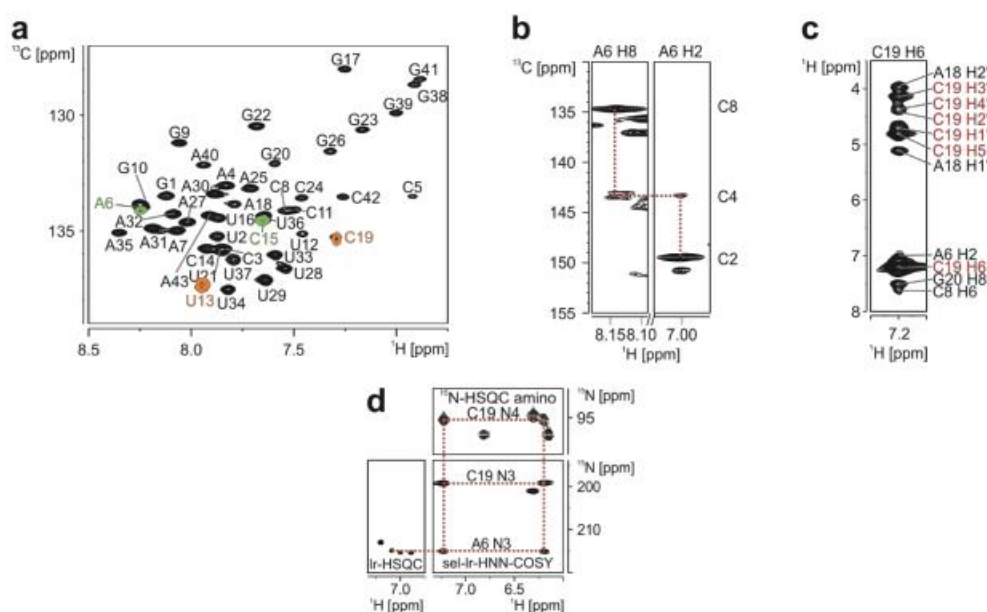
Supplementary Figure S4. A mutant with a truncated linker region (env9bΔ28-33) does show ligand binding. Representative ITC thermograms and fits for the env9bΔ28-33 riboswitch RNA binding to SAH (*left*) or SAM (*right*) at an Mg^{2+} concentration of 2 mM. The resulting K_D values are indicated.



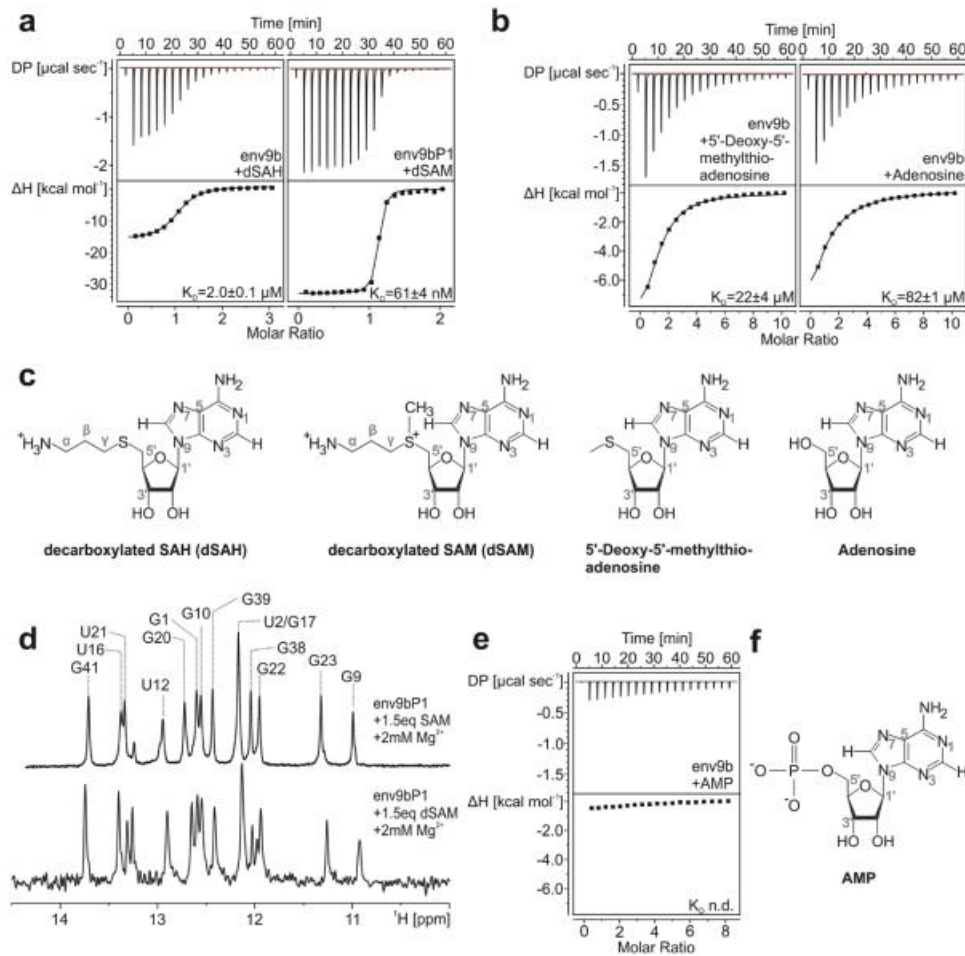
Supplementary Figure S5. Identification of U16 as the SAH binding nucleotide and assignment of key NMR resonances. (a) Imino proton spectra of env9bU16A (*top*) with 1.5-fold excess of SAH in the absence (*middle*) and in the presence of 2 mM magnesium acetate (*bottom*). (b) Overlay of ^1H , ^{13}C -HSQC spectra recorded at 20°C and optimized for the H8C8 and H6C6 moieties of ^{13}C , ^{15}N – labeled RNA (*black*) and site specific U16-H6C6/G10-H8C8 ^{13}C , ^{15}N – labeled RNA (*red*) for the unambiguous assignment of the U16 H6C6 resonance. (c) Overlay of ^1H , ^{15}N -HSQC spectra of the imino region of uniformly ^{15}N -labeled RNA (*black*) and site specific U16- ^{15}N -labeled RNA (*red*) recorded at 10°C for unambiguous assignment.



Supplementary Figure S6. Representative ITC thermograms for the env9bU36C mutant riboswitch titrated with SAH (*left*) or SAM (*right*) in the presence of 2 mM Mg^{2+} . The resulting K_D values are indicated.

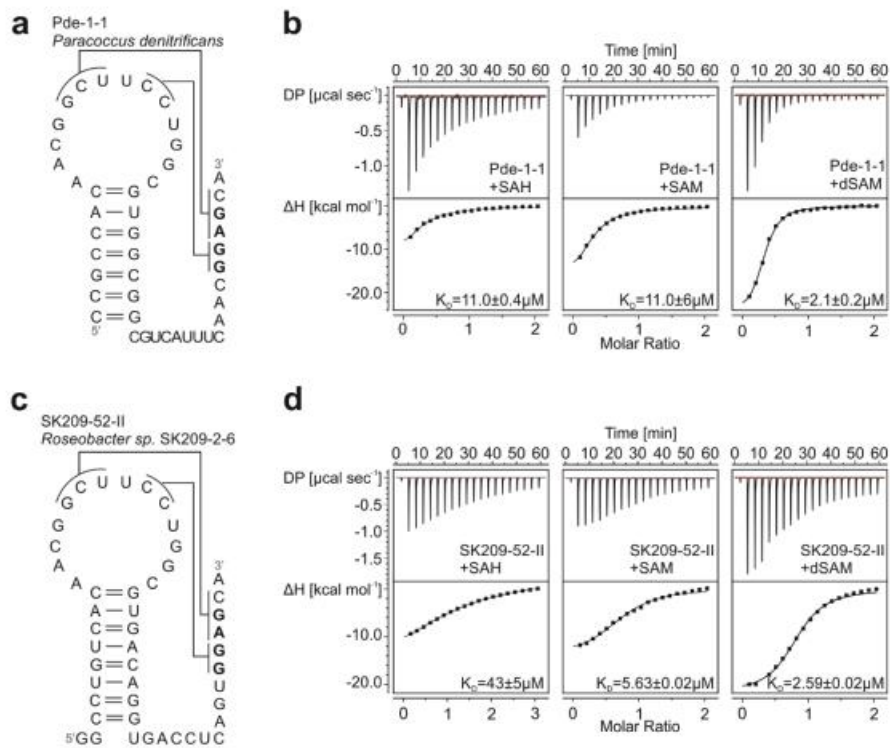


Supplementary Figure S7. Characterization of the A6-C19 base pair. (a) Overlay of ^1H , ^{13}C -HSQC spectra recorded at 20°C optimized for the H8C8 and H6C6 moieties of uniformly ^{13}C , ^{15}N -labeled (*black*), site and atom specifically A6-H8C8/C15-H6C6 ^{13}C , ^{15}N -labeled (*orange*) and U13-H6C6/C19-H6C6 ^{13}C , ^{15}N -labeled env9bP1 RNA (*green*) for unambiguous assignment of A6, U13, C15 and C19 base spin systems in the SAH-bound complex. (b) Strips from a TROSY-relayed-HCCH-COSY experiment recorded for a ^{13}C , ^{15}N -A/C-labeled env9bP1 RNA in complex with SAH to correlate the H2 and H8 resonances of A6 to the C4 resonance. (c) Strip of the C19 H6 resonance from the ^1H , ^{13}C -NOESY-HSQC of an ^{13}C , ^{15}N -labeled env9bP1 RNA in complex with SAH with intranucleotide NOEs (*red*) and internucleotide NOEs (*black*). (d) Ir-HSQC, amino ^{15}N -HSQC and sel-Ir-HNN-COSY experiment recorded for an ^{13}C , ^{15}N -A/C labeled env9bP1 RNA in complex with SAH for the identification of the hydrogen bond between A6 N3 and H41/H42 of C19.



Supplementary Figure S8. (a)-(b) Representative ITC thermograms for env9b and env9bP1 titrated with the decarboxylated analogues of SAH and SAM (dSAH and dSAM), respectively, and analogues with truncated aminocarboxypropyl moieties – 5'-deoxy-5'-methylthioadenosine and adenosine – in the presence of 2 mM Mg^{2+} . The resulting K_D values are indicated. (c) Structural formula of ligands used in (a) and (b) – decarboxylated SAH and SAM, 5'-deoxy-5'-methylthioadenosine and adenosine. (d) Imino proton spectra of env9bP1 titrated with 1.5-fold excess of SAM and 2 mM magnesium acetate (*top*) and 1.5-fold excess of decarboxy-SAM and 2 mM magnesium acetate (*bottom*). (e) Representative ITC

thermogram for an env9b RNA titration with adenosine monophosphate (AMP). No binding event is discernible from this thermogram. (f) Structure of AMP.



Supplementary Figure S9. Ligand binding affinities of other SAM/SAH-binding riboswitches.

(a) Secondary structure of the SAM/SAH-binding riboswitch Pde-1-1 from *Paracoccus denitrificans*. (b) Representative ITC thermograms for Pde-1-1 titrated with SAH, SAM and the decarboxylated analogue of SAM (dSAM), respectively, in the presence of magnesium ions. The resulting K_D values are indicated. (c) Secondary structure of the SAM/SAH-binding riboswitch SK209-52-II from *Roseobacter sp. SK209-2-6*. (d) Representative ITC thermograms for SK209-52-II titrated with SAH, SAM and the decarboxylated analogue of SAM (dSAM), respectively. The resulting K_D values are indicated.

Supplementary Table S1. Overview of ITC-experiments and the derived SAH and SAM binding parameters for the env9b, env9bP1 and their mutant at variable Mg^{2+} concentrations at 25 °C. Titrations with SAM and no magnesium ions showed very low binding ability that where outside the measurable range of the ITC instrument.

RNA	P1 background	Ligand	Mg^{2+} [mM]	K_D [μ M]	N	ΔH [kcal mol ⁻¹]	ΔS [cal mol ⁻¹ deg ⁻¹]
WT	WT	SAM	0	95 ± 4	0.5 ± 0.1	-66 ± 6	-204 ± 20
		SAH		n.d.	n.d.	n.d.	n.d.
WT	2	SAM		1.9 ± 0.2	0.6 ± 0.1	-24 ± 1	-53 ± 3
		SAH		6.9 ± 0.3	0.62 ± 0.02	-26.2 ± 0.4	-64 ± 1
		decarboxy-SAH (dSAH)		2.0 ± 0.1	1.0 ± 0.1	-15.7 ± 0.9	-27 ± 3
		5'-Deoxy-5'-methylthioadenosine		22 ± 4	1.1 ± 0.2	-13 ± 6	-22 ± 21
		Adenosine		82 ± 1	0.55 ± 0.03	-29 ± 2	-79 ± 2
		AMP		n.d.	n.d.	n.d.	n.d.
P1opt	P1 mutant	SAM		1.5 ± 0.1	0.9 ± 0.1	-26.4 ± 0.6	-62 ± 2
		SAH		3.7 ± 0.3	0.796 ± 0.004	-28.9 ± 0.4	-72 ± 2
U13A	WT	decarboxy-SAM (dSAM)		0.061 ± 0.004	1.0 ± 0.1	-32.5 ± 0.7	-76 ± 2
		SAM		1.7 ± 0.1	0.73 ± 0.03	-34.2 ± 0.9	-88 ± 3
		SAH		6.0 ± 0.3	0.78 ± 0.01	-33 ± 2	-85 ± 6
		SAM		18 ± 2	0.71 ± 0.01	-25 ± 2	-63 ± 6
U13C		SAM		58 ± 8	0.7 ± 0.1	-26 ± 4	-67 ± 15
		SAH		40 ± 4	1.07 ± 0.05	-27 ± 3	-71 ± 9
A6G	P1 mutant	SAM		47 ± 2	0.98 ± 0.04	-36 ± 2	-102 ± 8
A6GA18U		SAM		0.8 ± 0.1	0.6 ± 0.1	-36 ± 2	-91 ± 5
A18G	WT	SAM		1.3 ± 0.1	0.51 ± 0.03	-33 ± 1	-85 ± 4
		SAH		3 ± 1	0.8 ± 0.2	-30 ± 2	-74 ± 8
U36C		SAM		13 ± 1	0.9 ± 0.1	-29.2 ± 0.6	-76 ± 8
		SAH		6.2 ± 0.1	0.78 ± 0.02	-40 ± 1	-110 ± 5
del28-33		SAM		27 ± 1	1.06 ± 0.04	-27.0 ± 0.1	-69.6 ± 0.4
		SAH		11 ± 6	0.2 ± 0.1	-39 ± 26	-108 ± 89
Pde-1-1	WT	SAM					

SK209-52-II	ΔG28	SAH	11.0 ± 0.4	0.24 ± 0.01	-15 ± 2	-29 ± 5
		decarboxy-SAM (dSAM)	2.1 ± 0.2	0.30 ± 0.01	-27 ± 1	-63 ± 3
		SAM	5.63 ± 0.02	0.74 ± 0.02	-16 ± 1	-31 ± 3
		SAH	43 ± 5	1.20 ± 0.09	-20 ± 1	-46 ± 2
		decarboxy-SAM (dSAM)	2.59 ± 0.02	0.81 ± 0.02	-21 ± 1	-44 ± 5

Supplementary Table S2. Relevant enzymes in organisms containing SAM/SAH, according to the KEGG database. Organisms are listed that have predicted SAM/SAH riboswitches, whose genome sequences in RefSeq version 87 are complete and that are annotated in the KEGG database. None of the organisms listed below contains a recognizable gene for SAH nucleosidase (EC 3.2.2.9).

Organism	SAM decarboxylase (EC 4.1.1.50)	SAH hydrolase (EC 3.3.1.1)
<i>Jannaschia</i> sp. CCS1	Yes	Yes
<i>Ketogulonicigenium vulgare</i> WSH-001	No	No
<i>Ketogulonicigenium vulgare</i> Y25	No	Yes
<i>Phaeobacter inhibens</i> DSM 17395	Yes	Yes
<i>Rhodobacter sphaeroides</i> 2.4.1	No	Yes
<i>Rhodobacter sphaeroides</i> ATCC 17025	No	Yes
<i>Rhodobacter sphaeroides</i> ATCC 17029	No	Yes
<i>Rhodobacter sphaeroides</i> KD131	No	Yes
<i>Ruegeria</i> sp. TM1040	Yes	Yes