

# Characterization of Structure and Dynamics of the Guanidine-II Riboswitch from *Escherichia coli* by NMR Spectroscopy and Small-Angle X-ray Scattering (SAXS)

Tatjana Schamber,<sup>[a, c]</sup> Oliver Binas,<sup>[a, c]</sup> Andreas Schlundt,<sup>[b, c]</sup> Anna Wacker,<sup>[a, c]</sup> and Harald Schwalbe\*<sup>[a, c]</sup>

Riboswitches are regulatory RNA elements that undergo functionally important allosteric conformational switching upon binding of specific ligands. The here investigated guanidine-II riboswitch binds the small cation, guanidinium, and forms a kissing loop-loop interaction between its P1 and P2 hairpins. We investigated the structural changes to support previous studies regarding the binding mechanism. Using NMR spectroscopy, we confirmed the structure as observed in crystal structures and we characterized the kissing loop interaction

upon addition of  $Mg^{2+}$  and ligand for the riboswitch aptamer from *Escherichia coli*. We further investigated closely related mutant constructs providing further insight into functional differences between the two (different) hairpins P1 and P2. Formation of intermolecular interactions were probed by small-angle X-ray scattering (SAXS) and NMR DOSY data. All data are consistent and show the formation of oligomeric states of the riboswitch induced by  $Mg^{2+}$  and ligand binding.

## Introduction

In bacteria, riboswitches are wide-spread regulatory elements primarily located in the 5'-untranslated region (5'-UTR) of mRNAs. They control gene expression mainly at the level of translation or transcription in response to binding of molecules of low molecular weight. While a large variety of riboswitches are known to sense various cell-metabolites<sup>[1]</sup> including TPP,<sup>[2,3]</sup> FMN,<sup>[4]</sup> preQ<sub>1</sub>,<sup>[5]</sup> amino acids,<sup>[6]</sup> or purines,<sup>[7]</sup> as well as the secondary messengers 3',3'-cGAMP,<sup>[8]</sup> c-di-GMP,<sup>[9]</sup> ZMP<sup>[10]</sup> and even small ions ( $F^-$ ,  $Mg^{2+}$ ),<sup>[11,12]</sup> riboswitches sensing the cytotoxic guanidinium cation were only recently discovered.<sup>[13]</sup> Guanidinium ( $Gdm^+$ ) is the protonated form of guanidine, possesses  $D_{3h}$  symmetry and is protonated under physiological conditions due to its high  $pK_A$  value of 13.6. It is able to provide

six hydrogen bonds for binding to hydrogen acceptor sites in RNA.<sup>[14]</sup> The  $Gdm^+$  cation is significantly larger than the ubiquitous RNA-binding cation  $Mg^{2+}$ , also larger than  $F^-$ .<sup>[11]</sup>

The putative biological function of free guanidine has been discussed for long but was not ascertained due to the absence of biological receptors. In fact, new insight came from the discovery of riboswitches that function as  $Gdm^+$ -binding receptors.<sup>[15,16]</sup> These guanidine riboswitches (Gdn) sense  $Gdm^+$  with high sensitivity and specificity.<sup>[17]</sup> Derived from the genomic location of the  $Gdm^+$ -sensing riboswitches,  $Gdm^+$  has then been proposed to play an important role in the detoxification of cells via its carboxylation<sup>[13]</sup> but also to serve as a source of nitrogen.<sup>[18]</sup>

To date, four different classes of  $Gdm^+$ -sensing riboswitches have been discovered termed guanidine-I to guanidine-IV.<sup>[15,18,19]</sup> Many of these  $Gdm^+$ -sensing riboswitches are biologically well characterized and structural models of isolated hairpins were reported for all the four riboswitch classes including Gdn-II from *Escherichia coli* (PDB: 5NDI), from *Gloeobacter violaceus* (5NOM and 5NDH) and from *Pseudomonas aeruginosa* (5VJ9).<sup>[14,20]</sup> The structures reveal striking similarities of ligand binding pockets between these different riboswitches.<sup>[21]</sup>

For a Gdn-II class riboswitch aptamer domain, it was shown that binding of  $Gdm^+$  to the  $Gdm^+$ -sensing riboswitches induces kissing loop-loop interactions leading to the formation of two GC base pairs between its two hairpin elements (Figure 1A). The  $Gdm^+$  ligand binds via hydrogen bonds to the Hoogsteen side of the third loop nucleobase guanosine G13 (O6, N7) and the backbone phosphodiester. In addition,  $\pi$ - and ionic interactions stabilize ligand binding (Figure 1B).<sup>[14]</sup> Up to now, only hairpins with ACGA loop motifs have been investigated at the structural level but not the equally frequent ACGG loop motifs.<sup>[22]</sup> Likewise, the possibility of hetero dimerization involving two independent and different looped hairpins

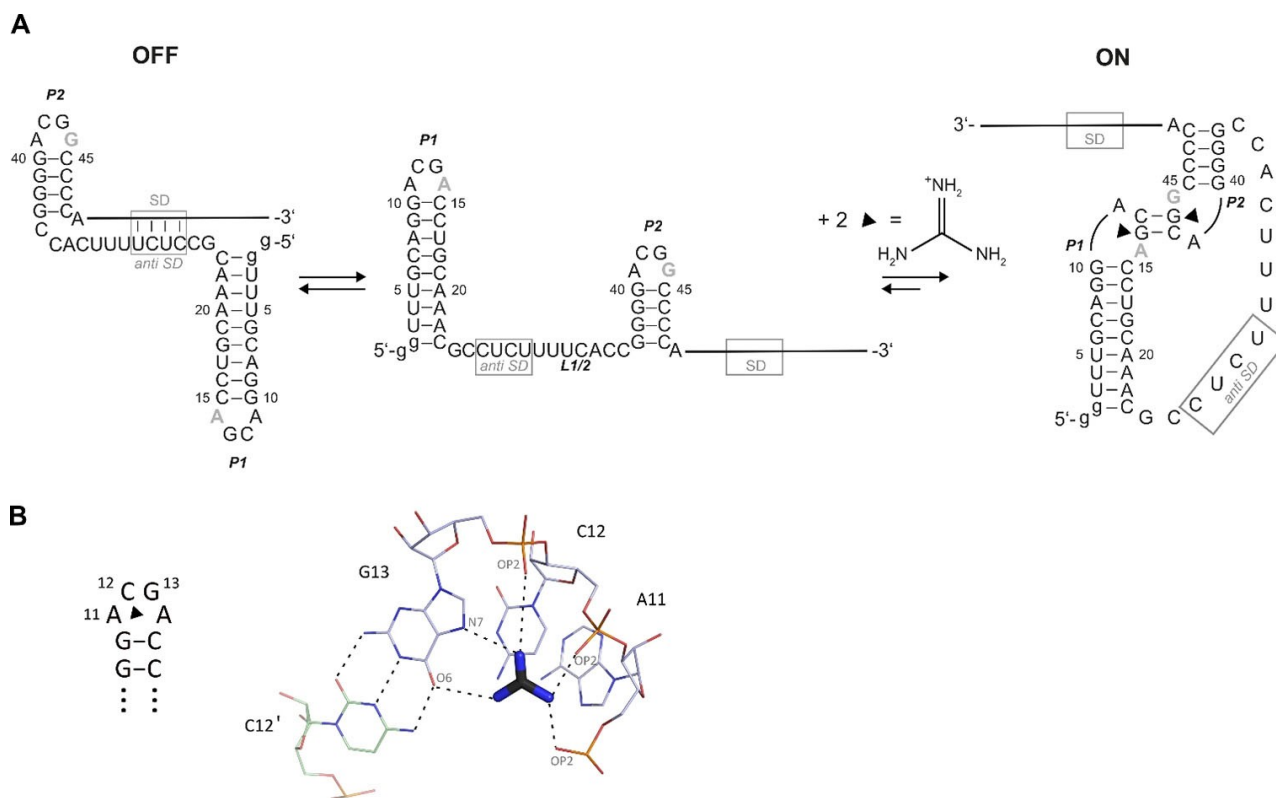
[a] T. Schamber, O. Binas, A. Wacker, Prof. Dr. H. Schwalbe  
Institute for Organic Chemistry and Chemical Biology  
Johann Wolfgang Goethe University  
Max-von-Laue-Str. 7  
60438 Frankfurt/Main (Germany)  
E-mail: schwalbe@nmr.uni-frankfurt.de

[b] Dr. A. Schlundt  
Institute for Molecular Biosciences  
Johann Wolfgang Goethe University  
Max-von-Laue-Str. 9  
60438 Frankfurt/Main (Germany)

[c] T. Schamber, O. Binas, Dr. A. Schlundt, A. Wacker, Prof. Dr. H. Schwalbe  
Center for Biomolecular Magnetic Resonance  
Institute for Organic Chemistry and Chemical Biology  
Johann Wolfgang Goethe University  
Max-von-Laue-Str. 7-9  
60438 Frankfurt/Main (Germany)

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202100564>

© 2021 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** (A) Secondary structures of the investigated aptamer within the Gdn-II riboswitch and the proposed mechanism of switching.<sup>[20]</sup> The structural elements P1, P2 and linker L1/2 are annotated accordingly. 'SD' and 'anti-SD' denote the Shine-Dalgarno and complementary *anti*-Shine-Dalgarno sequences, respectively. The current model of the mechanism of switching of the full-length Gdn-II riboswitch including the expression platform is shown. (B) Representation of the ligand binding pocket of P1 depicting hydrogen bonds for Gdm<sup>+</sup> to the Hoogsteen side of guanosine (via O6 and N7) and the backbone phosphates (right). The corresponding sequence is shown on the left.

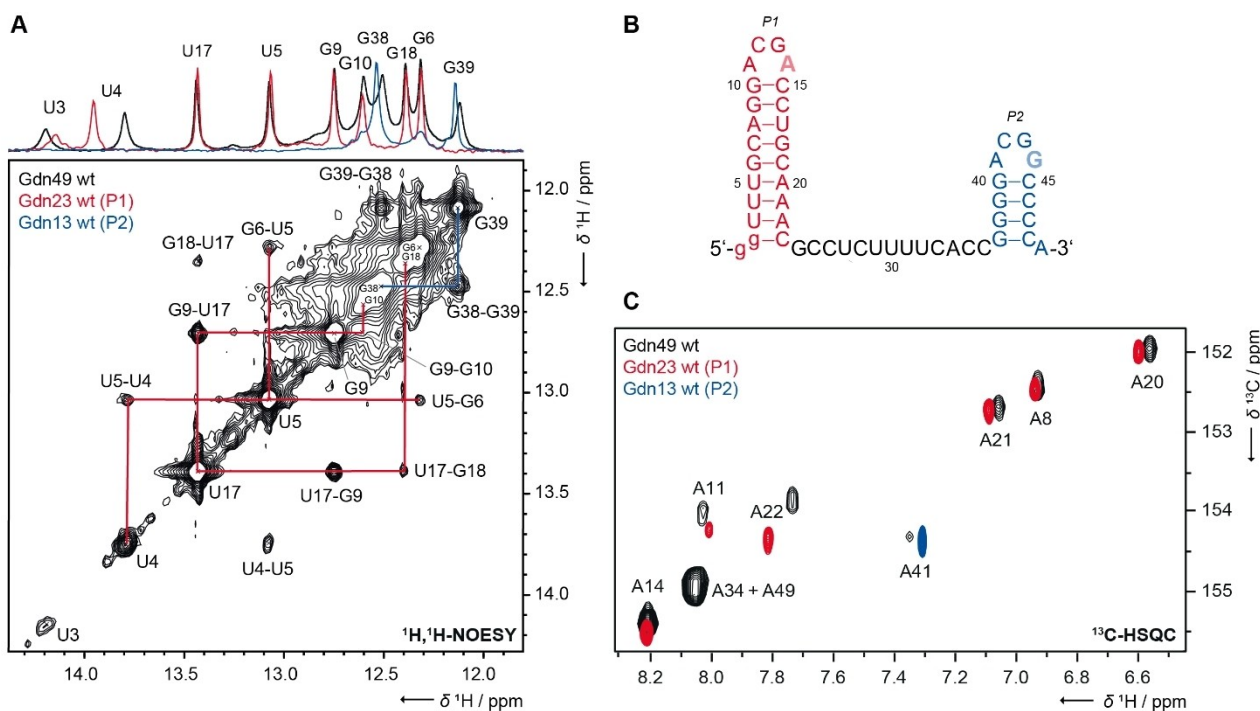
has escaped detailed (structural) analysis. We here investigate the Gdn-II class riboswitch aptamer from *E. coli* (Figure 1A) that is part of the *sugE* gene which codes for a multi drug efflux pump<sup>[17]</sup> by NMR spectroscopy and small-angle X-ray scattering (SAXS). The riboswitch consists of the mini-*ykkC* motif<sup>[22]</sup> that contains two linked GC rich stem-loop domains P1 and P2. Both stems are capped by *ACGR* loop motifs.<sup>[14]</sup> The kissing loop-loop interaction was proposed to up-regulate *SugE* through weakening of the interaction of linker L1/2 with a putative downstream Shine-Dalgarno sequence (Figure 1A). The increased accessibility to the Shine-Dalgarno ribosome binding sites is assumed to promote translation initiation.<sup>[17]</sup> Kissing loop-loop interactions upon ligand binding have previously been reported in various riboswitches but striking differences are observed.<sup>[23]</sup> In guanine- and adenine-sensing riboswitches, the ligand binding site is remote from the kissing loop-loop interaction site. Formation of the kissing loop-loop interaction leads to an allosteric stabilization of the RNA-ligand interaction. By contrast, in the Gdn-II riboswitch, the ligand binding site is located in close proximity to the kissing loop-loop interaction site. While the interaction between P1 and P2 in Gdn-II has been studied systematically on a mimic of the double stem-loop RNA,<sup>[24]</sup> its existence has not been shown for the wild-type sequence, i.e. including the linker L1/2 and two different *ACGR* loops. In this

study, we provide evidence that the two GC base pairs form the putative kissing loop-loop interaction and map binding sites for the ligand to various relevant constructs of the riboswitch at atomic resolution. We further investigate mutant constructs to delineate functional differences between the two hairpin domains.

## Results and Discussion

### NMR resonance assignment

Following a divide-and-conquer-strategy,<sup>[25]</sup> we conducted NMR experiments of the wild-type (wt) Gdn-II aptamer domain from *E. coli* containing 49 nucleotides (nts) RNA (Gdn49 wt) and assigned its NMR resonances. We used two constructs consisting of the first stem-loop with 23 nts (Gdn23 wt as P1) and the second stem-loop with 13 nts (Gdn13 wt as P2). Both stem-loops are present in the full-length aptamer Gdn49 wt (Figure 2B) and we could in fact observe comparable sets of imino signals from nucleobases involved in base pairing in <sup>1</sup>H-1D as well as 2D-<sup>1</sup>H,<sup>1</sup>H-NOESY experiments (Figure 2A). Signals at 12.1 ppm (assigned as G39) and 12.5 ppm (assigned as G38) are present in Gdn13 wt (blue) and Gdn49 wt (black) but not in



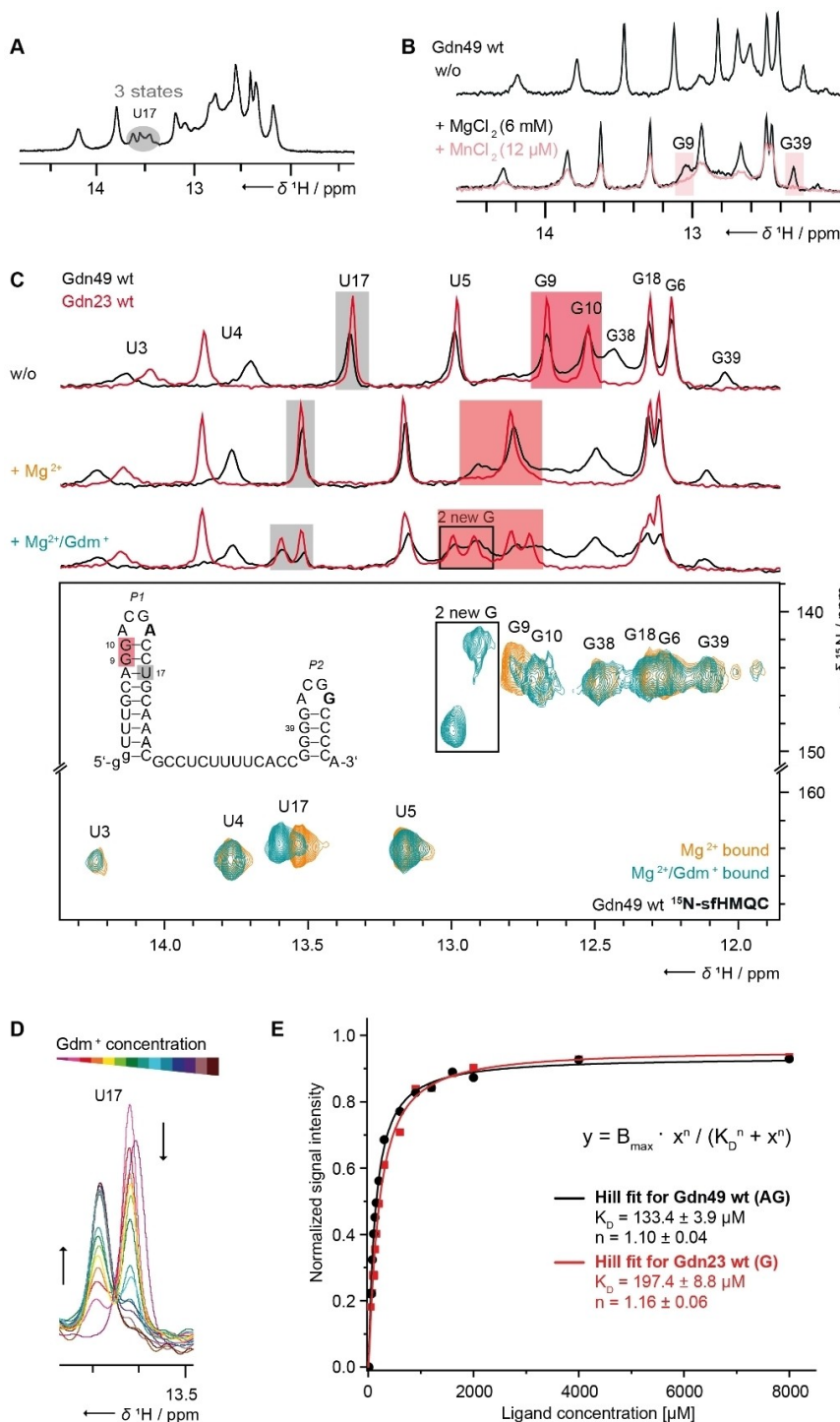
**Figure 2.** NMR resonance assignments. (A) Imino region of  $^1\text{H}, ^1\text{H}$ -NOESY spectra of Gdn49 wt annotated with assignments. Color-coding indicates resonances from either P1 (red) or P2 (blue). Data were measured at 600 MHz,  $2048 \times 768$  points, 160 scans and 150 ms mixing time. The sample contained  $600\ \mu\text{M}$  RNA. (B) Secondary structure of the investigated Gdn-II riboswitch from *E. coli*. Full-length aptamer is Gdn49 wt, P1 is Gdn23 wt with ACGA loop motif in red and P2 is Gdn13 wt with ACGG loop motif in blue. (C)  $^{13}\text{C}$ -HSQC spectra of adenosines C2-H2 recorded from Gdn49 wt and the structural elements P1 (Gdn23 wt) and P2 (Gdn13 wt). Data were measured at 950/238 MHz,  $2048 \times 128$  points, 256 scans with  $720\ \mu\text{M}$  of Gdn13 wt, 800/201 MHz,  $1024 \times 128$  points, 512 scans with  $660\ \mu\text{M}$  of Gdn23 wt and 950/238 MHz,  $1332 \times 128$  points, 2 scans with  $800\ \mu\text{M}$   $^{13}\text{C}, ^{15}\text{N}$ -labeled Gdn49 wt.

Gdn23 wt (red), and signals of Gdn23 wt are present in Gdn49 wt but not in Gdn13 wt, a finding that is also supported from analysis of  $^1\text{H}, ^{13}\text{C}$ -correlations, e.g. C2-H2 (Figure 2C and Supporting Figure S1).

$^1\text{H}$  chemical shift assignment was obtained from analysis of 2D- $^1\text{H}, ^1\text{H}$ -NOESY experiments. Sequentially connected Watson-Crick base pairs AU and GC were identified in the imino-imino cross peak region of the NOESY spectra. The assignment was confirmed by  $^1\text{H}, ^{15}\text{N}$ -correlation spectra (as depicted for the bound state in Figure 3C). As a result, the sequential assignment for all guanosine and uridine residues in the RNA stems was obtained.<sup>[25]</sup> Note that the minor chemical shift differences of the U3 and U4 resonances for the Gdn23 wt and Gdn49 wt are typically observed in divide-and-conquer approaches and are likely due to the different stem stability in presence or absence of the linker. The shift of U4 is larger than of U3 but the reason for this remains unclear. We did not assign the imino proton of G40 which partially overlaps with G6 or G10. Further, imino protons of the terminal helical base pairs in P1 and P2 were not observed due to solvent exchange. We also observed weak NMR signals at 10.3 ppm for the constructs including P2 (Gdn13 wt and Gdn49 wt, Supporting Figure S2), indicating the presence of a non-canonical base pair potentially stabilized by a G–G Hoogsteen interaction.

### Ligand interaction in guanidine-II wild-type constructs of *E. coli*

In NMR titration experiments, we characterized the interaction of the Gdn-II riboswitch with  $\text{Mg}^{2+}$  and the ligand  $\text{Gdm}^+$ , respectively, by monitoring chemical shift perturbations (CSPs) upon increasing concentrations of both ligands. It was previously established that  $\text{Gdm}^+$  binding requires the presence of  $\text{Mg}^{2+}$ .<sup>[26]</sup> We were able to detect the simultaneous population of three different long-lived conformational states (unbound,  $\text{Mg}^{2+}$ -bound and  $\text{Mg}^{2+}/\text{Gdm}^+$ -bound) using the imino proton signal of U17 as a reporter signal (Figure 3A). We deliberately started by adding only 3 mM  $\text{Mg}^{2+}$  which does not suffice to saturate the  $\text{Mg}^{2+}$  binding capacity of the riboswitch at an RNA concentration of  $215\ \mu\text{M}$  for Gdn49 wt. Under this condition, we observed signal changes compared to the unbound state both, for the  $\text{Mg}^{2+}$  bound state and for the  $\text{Mg}^{2+}/\text{Gdm}^+$ -bound state after adding  $430\ \mu\text{M}$   $\text{Gdm}^+$ . Specific  $\text{Mg}^{2+}$  binding sites involving nucleotides G9 (P1) and G39 (P2) could be identified via  $\text{Mn}^{2+}$ -induced paramagnetic relaxation enhancement (PRE)<sup>[27]</sup> (Figure 3B). These  $\text{Mn}^{2+}$ -PRE experiments typically involve quantification of the change in signal intensity when adding sub-stoichiometric amounts of  $\text{Mn}^{2+}$  competing with  $\text{Mg}^{2+}$  for binding under fast exchange conditions.<sup>[27]</sup> The identified nucleotides G9 (P1) and G39 (P2) are located in close proximity to the ligand binding site adjacent to the loop, and consequently the corresponding signals vanished upon addi-



**Figure 3.**  $\text{Mg}^{2+}$  and ligand dependent structural changes observed by CSP. (A) Three states detected for U17 in sub-stoichiometric titration experiments (unbound,  $\text{Mg}^{2+}$ - and  $\text{Mg}^{2+}/\text{Gdm}^+$ -bound) in  $^1\text{H}$ -1D NMR spectra of Gdn49 wt in presence of 3 mM  $\text{Mg}^{2+}$  and 430  $\mu\text{M}$   $\text{Gdm}^+$ . Data were measured at 600 MHz, room temperature, 8192 points, 512 scans with 215  $\mu\text{M}$  of Gdn49 wt. (B) Determination of  $\text{Mg}^{2+}$  binding site through  $\text{Mn}^{2+}$ -induced paramagnetic relaxation enhancement for Gdn49 wt with  $^1\text{H}$ -1D NMR spectra of Gdn49 wt unbound, in presence of 6 mM  $\text{Mg}^{2+}$  and in addition of 12  $\mu\text{M}$   $\text{Mn}^{2+}$  (0.2%). Data were measured at 600 MHz, room temperature, 4096 points, 1024 scans with 20  $\mu\text{M}$  of RNA. (C)  $^1\text{D}$  spectra: Ligand titration states for Gdn23 wt and Gdn49 wt. Imino CSPs marked in gray for U17 and in red for G9, G10 and two new arising signals, respectively.  $^1\text{H}$ -1D NMR spectra unbound, in presence of 5 mM  $\text{Mg}^{2+}$  and in addition of 320  $\mu\text{M}$   $\text{Gdm}^+$ . Data were measured at 600 MHz, room temperature, 4096 points, 1024 scans with 20  $\mu\text{M}$  RNA.  $^2\text{D}$  spectra: Verifying of ligand-induced GC base pairs in  $^{15}\text{N}$ -SOFAST-HMQC spectra of Gdn49 wt before (orange) and after (cyan) addition of  $\text{Gdm}^+$  in presence of 5 mM  $\text{Mg}^{2+}$ . Data were measured at 600 MHz, room temperature, 36 points and 64 scans. Sample contained 100  $\mu\text{M}$  RNA of Gdn49 wt. (D)  $\text{Gdm}^+$  dependent decrease and increase of signal intensity in  $^1\text{H}$ -1D NMR spectra for the  $K_D$  determination. The signal of U17 in Gdn49 wt was analyzed. The color coding corresponds to a rainbow gradient from low concentration of  $\text{Gdm}^+$  in red to high concentration in violet.  $^1\text{H}$ -1D NMR spectra in presence of 5 mM  $\text{Mg}^{2+}$  and in addition of up to 8000  $\mu\text{M}$   $\text{Gdm}^+$ . Data were measured at 700 MHz, room temperature, 2048 points, 768 scans with 50  $\mu\text{M}$  of RNA. (E) Plot of  $K_D$ -determination via  $^1\text{H}$ -1D NMR titration experiments (Supporting Figure S5) for imino signal U17 of Gdn23 wt (red) and Gdn49 wt (black). The resulted values were determined by the Hill equation fit.

tion of  $\text{MnCl}_2$ . For the remaining signals, a significantly smaller decrease in intensity and signal broadening was observed. To define this decrease, we used the integral ratio between spectra before and after adding of  $\text{MnCl}_2$  (see percentages of integrals for remaining signals in Figure 3B).

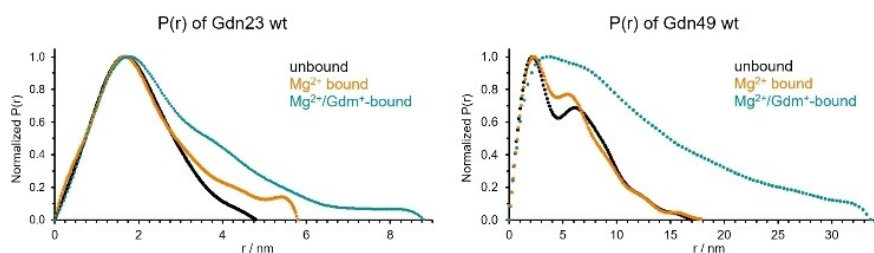
We further characterized  $\text{Gdm}^+$ -binding in the presence of 5 mM  $\text{Mg}^{2+}$ . Imino resonances of U17 (marked in gray, Figure 3C 1D spectra and Supporting Figure S3), G9 and G10 (marked in red) of P1 showed significant CSPs. We noted that U17 shows larger changes than G9 and G10 but we cannot explain this. Upon addition of  $\text{Gdm}^+$  (up to 8 mM), the formation of two GC kissing base pairs is reasoned by the detection of two new imino proton signals (Figure 3C 2D spectra), in line with the base pairs observed in crystal structures.<sup>[19]</sup> For Gdn23 wt and Gdn49 wt, we see shifts both, for  $\text{Mg}^{2+}$  and additional shift of  $\text{Gdm}^+$  interaction. By contrast, in P2, we only observed  $\text{Mg}^{2+}$ -induced CSPs, but no further shifts in effect of  $\text{Gdm}^+$  as reporter signals indicating no interaction of  $\text{Gdm}^+$  with the smaller stem-loop (Supporting Figure 12). When adding only  $\text{Gdm}^+$  in the absence of  $\text{Mg}^{2+}$ , no ligand binding was observed (Supporting Figure S4). Specific  $\text{Mg}^{2+}$  binding to P1 is required for  $\text{Gdm}^+$  binding, and pre-formation of the kissing loop-loop conformation precedes ligand binding. The  $\text{Mg}^{2+}$ -bound state coexists with both, the  $\text{Mg}^{2+}$ -free state and the  $\text{Mg}^{2+}/\text{Gdm}^+$ -bound state.

From quantification of CSPs in  $^1\text{H}$ -1D NMR titration experiments (Supporting Figure S5), we determined a  $K_D$  of 130  $\mu\text{M}$  for the full-length aptamer Gdn49 wt with a Hill coefficient  $n$  of 1.1 and a  $K_D$  of 200  $\mu\text{M}$  with  $n$  of 1.2 for the isolated P1 stem-loop Gdn23 wt (Figure 3D and 3E). While the Gdn49 wt might be expected to bind two ligand molecules the Hill coefficient points to a strong difference in affinity between the two ligand binding stems P1 and P2. Gdn23 wt features an ACGA loop motif and the  $K_D$  values of Gdn23 wt and Gdn49 wt are in a comparable regime (Table 2 and Figure 3E). Accordingly, in Gdn49 wt the ACGA loop motif located in P1 represents the stronger ligand binding site. Since the  $K_D$  value for the isolated stem-loop is higher, we conclude that the kissing hairpin formation upon ligand binding is facilitated by linker L1/2 and the additional interaction with stem-loop P2.

## DOSY and SAXS characterization

Characterization of ligand binding by  $^1\text{H}$ -1D NMR revealed two new G-imino signals, assigned to the formation of two additional GC base pairs. These base pairs may either arise from intramolecular kissing loop interaction or from dimerization of Gdn49 wt via intermolecular P1-P1' kissing loop interaction. The lack of any CSPs after addition of ligand, clearly pointed to the induction of dimerization as the more likely event. Thus, to determine the molecularity of the ligand-Gdn49 wt complex, we measured its hydrodynamic radius ( $R_h$ ) by diffusion ordered spectroscopy (DOSY) NMR experiments.<sup>[28]</sup> We observed an increase of  $R_h$  greater than 60% upon addition of  $\text{Gdm}^+$ , indicating dimerization upon complex formation (Supporting Figure S6). In conjunction with the  $^1\text{H}$ -1D-NMR data showing a single set of signals, we assume the formation of a  $C_2$ -symmetric dimer with the P1 loop as interaction interface (Supporting Figure S7). The addition of only  $\text{Mg}^{2+}$  increased  $R_h$  by 10% (Supporting Figure S6), pointing to a  $\text{Mg}^{2+}$ -dependent transient pre-formation of kissing loop interaction.

In addition, we characterized Gdn23 wt and Gdn49 wt by SAXS focusing on the oligomerization states of the free and complexed RNAs (Figure 4, Supporting Figure S8 and Supporting Table S9), as previously conducted for different riboswitches.<sup>[29,30]</sup> For Gdn23 wt in the absence of  $\text{Mg}^{2+}$  and  $\text{Gdm}^+$ , analysis of SAXS data revealed an apparent molecular weight (MW) of 7.3 kDa, a maximal intramolecular distance ( $D_{\text{max}}$ ) of 48 Å and a radius of gyration ( $R_g$ ) of 14.8 Å consistent with a predominantly monomeric conformation (Table 1). These data also agree with a predicted structure<sup>[31]</sup> as well as the crystal structure<sup>[19]</sup> (Supporting Figure S10). Upon addition of  $\text{Mg}^{2+}$ , we also observed an increase of the apparent MW (16%),  $D_{\text{max}}$  (21%) and  $R_g$  (20%). When  $\text{Gdm}^+$  is added, the apparent MW increases to 14.2 kDa,  $D_{\text{max}}$  to 88 Å and  $R_g$  to 22 Å in accordance with a dimeric species, which is in line with the NMR data. The theoretical molecular weight of the full aptamer domain (Gdn49 wt) is 14.9 kDa, but analysis of SAXS data for free Gdn49 wt revealed an apparent molecular weight of 30.0 kDa in line with predominantly dimeric arrangement. Since the P1-P1' dimerization is not observed in the free Gdn23 wt containing only the P1 stem, we conclude that there is at least a weak interaction of the P2 stems, which, however, is ligand-independent. We speculate that this potential P2-P2'-based dimerization stems from the interaction of G-rich P2 hairpins.



**Figure 4.** SAXS analysis of guanidine-II for Gdn23 wt (left) and Gdn49 wt (right). These curves represent the distance distribution functions  $P(r)$  for unbound (black),  $\text{Mg}^{2+}$  bound (orange) and  $\text{Mg}^{2+}/\text{Gdm}^+$ -bound (cyan) RNA reporting  $D_{\text{max}}$  as measurable geometric parameter.

Table 1. SAXS-derived geometric parameters for guanidine-II aptamer constructs Gdn23 wt and Gdn49 wt. We calculated the expected values using monomer and dimer models, <sup>[31]</sup> respectively.					
SAXS analysis		Unbound	Mg <sup>2+</sup> -bound	Mg <sup>2+</sup> /Gdm <sup>+</sup> -bound	Expected changes from unbound to bound states
Gdn23 wt	MW <sub>exp</sub> [kDa]	7.3	8.5	14.2	7.0 to 14.0
	D <sub>max</sub> [Å]	48	58	88	42 to >80
	R <sub>g</sub> [Å]	14.8	17.8	22.0	13.6 to 21.4
	Conformation	monomer	> monomer	dimer	monomer to dimer
Gdn49 wt	MW <sub>exp</sub> [kDa]	30.0	35.5	> 94.0	14.9 to 14.9 (for intramolecular interaction) or to 29.8 (for intermolecular interaction)
	D <sub>max</sub> [Å]	169	181	> 336	no realistic determination possible due to linker
	R <sub>g</sub> [Å]	40.7	42.0	83.3	no realistic determination possible due to linker
	Conformation	dimer	> dimer	> hexamer?	monomer to kissing loop-loop interaction

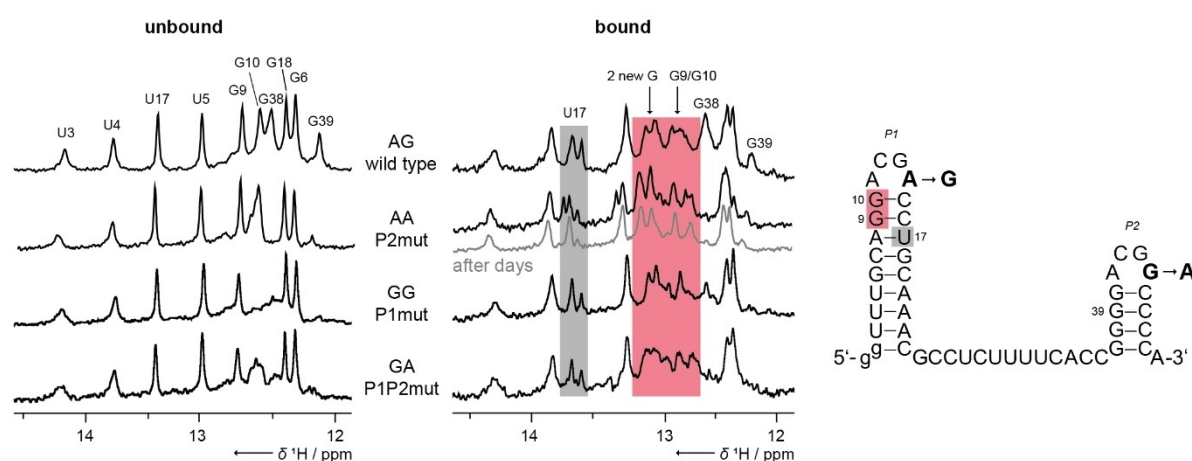
The P2 hairpins might interact via G-mediated interactions in line with the weak NMR signals presumably due to G–G Hoogsteen interaction observed in the corresponding region (10.3 ppm, Supporting Figure S2). Addition of only Mg<sup>2+</sup> to Gdn49 wt led to an increase in molecular weight (18%) consistent with the increase in R<sub>h</sub> observed in the DOSY NMR experiments (see also R<sub>g</sub>/R<sub>h</sub> in Supporting Table S11). When both, Mg<sup>2+</sup> and Gdm<sup>+</sup> were added, SAXS data yielded an experimental molecular weight of at least 94 kDa. This high molecular weight indicates conformations of higher molecular weight, presumably hexamers, most likely stabilized by intermolecular P1-P1' interactions.

### Investigation of mutants

Due to the lack of experimental data for the G variant of the fourth loop nucleotide (G44) we investigated whether G-to-A mutants in P2 could promote P1-P2 interactions. Since Gdn13 wt features a ACGG motif, we first examined the G44A mutant,

both in the isolated stem-loop (Gdn13 G44A, Supporting Figure S12) and in the full-length aptamer (Gdn49 P2mut, Figure 5). This mutation led to ligand binding, as we observed significant CSPs for both constructs (Gdn13 G44A and Gdn49 P2mut) upon Gdm<sup>+</sup> titration (Supporting Figure S13). We conclude that the loop sequence ACGR with an A at loop position 4 enables ligand binding. P1 dimerizes via P1-P1' kissing loop interaction, if P2 contains a ACGG loop motif with impaired kissing loop-loop interaction potential. Accordingly, we further speculated that an intramolecular kissing loop-loop complex could be formed in the presence of two ACGA loop motifs. Therefore, we investigated all four combinations of A/G loop variations (Table 2) in ligand titrations (Figure 5 and Supporting Figure S13) and determined K<sub>D</sub> values (Supporting Figure S14).

In NMR Gdm<sup>+</sup>-titrations with all possible loop combinations – AG (wild-type), AA, GG, GA – we observed approximately the same CSP response – U17, G9, G10 show CSPs, and two new G-signals arise. Consequently, mutants with ACGG loop motifs are ligand-binding competent and the ligand is



**Figure 5.** Investigation of mutated constructs shows distinct CSPs in P2. <sup>1</sup>H-1D NMR spectra of ligand titration experiments for Gdn49 constructs of wild-type and mutants are shown: free forms are shown on the left, and in presence of 5 mM Mg<sup>2+</sup> and 1 mM Gdm<sup>+</sup> on the right. CSPs marked in gray for U17 and in red for G9, G10 and two new arising signals, respectively. Data were measured at 600 MHz, room temperature, 4096 points, 512 scans with 20 μM of RNAs. An overview of the constructs and mutations is given in Table 2.

**Table 2.** Loop variants of the investigated guanidine-II aptamer construct and its binding behavior at RNA concentrations of 50  $\mu\text{M}$  with 5 mM  $\text{Mg}^{2+}$  and up to 8 mM  $\text{Gdm}^+$ .

	Loop sequence	Abbreviation for loop sequence	$K_D$ [ $\mu\text{M}$ ]	Hill coefficient n
P1 wild-type Gdn23	ACGA	A	197.4 $\pm$ 8.8	1.16 $\pm$ 0.06
P2 wild-type Gdn13	ACGG	G	not detected	not detected
Aptamer wild-type Gdn49	ACGA...ACGG	AG	133.4 $\pm$ 3.9	1.10 $\pm$ 0.04
Gdn49 P1mut	ACGG...ACGG	GG	264.3 $\pm$ 27.8	1.35 $\pm$ 0.15
Gdn49 P2mut	ACGA...ACGA	AA	58.9 $\pm$ 8.1	1.01 $\pm$ 0.22
Gdn49 P1P2mut	ACGG...ACGA	GA	165.0 $\pm$ 14.3	1.38 $\pm$ 0.17

clearly sensed by both types of loop motifs in the context of P1. Comparison of  $K_D$ s between constructs (Table 2) shows a clear preference for A-loop interactions, evidenced by the lowest  $K_D$  of 60  $\mu\text{M}$  for the AA-construct (Gdn49 P2mut), and followed by the construct with an A-loop in only P1 (AG, Gdn49 wt, 130  $\mu\text{M}$ ). Accordingly, the  $K_D$  value is higher for mutants with either an A-loop only in P2 (165  $\mu\text{M}$ , GA, Gdn49 P1P2mut), and yet more when only the G-loop is present (260  $\mu\text{M}$ , GG, Gdn49 P1mut). Our NMR and SAXS data showed that only P1 – the A-loop construct – adopts a monomeric conformation and the aptamer wild-type construct (Gdn49 wt) has a dimer conformation in the free form. Comparing the  $K_D$  values relative to wild-type, it becomes apparent that ligand binding and formation of binding pocket are compromised either by shortening the construct to only P1 or when mutated to the G variant. In contrast, a mutation to the A-loop motif improves binding by a factor of 2.

We thus propose that the binding pocket is preferentially formed when adenosines are presented in position four of the loop and a P2-P2' interaction supports dimerization because the stem is GC rich, and especially when it contains the G-loop motif. Possibly, the enthalpic gain from interactions involving G-loops might not sufficiently compensate the entropic penalty compared to sequences involving interactions only between A-loops. Consequently, a G-loop only mutant is less capable of forming binding pockets. We speculate that this is due to a possible weak intramolecular loop base pair interaction between C12-G14 and C42-G44, respectively. This finding implies that a single nucleotide can be responsible for the binding efficiency of the entire riboswitch as one ligand seems to be sufficient to promote the kissing interaction. In fact, a G variation in both loop sequences is rarely found in nature and the ACGA variant of P1 is conserved in most cases.<sup>[23]</sup> Moreover, for the AA-construct (P2mut), we observed that after a few days the bound state shows the exact same CSP scheme as the wild-type (Figure 5, spectra highlighted in gray). We thus conclude, the kissing loop-loop interaction induces an intermolecular  $C_2$ -symmetric dimerization rather than an intramolecular interaction. Such change of conformation takes also place in the unbound states of both, the Gdn13 G44A-mutant and the corresponding aptamer Gdn49 P2mut, which within a few days

show the same set of signals as the wild-type constructs (Supporting Figure S15).

## Conclusion

In this study, we probed several wild-type and mutated constructs of the guanidine-II riboswitch from *E. coli* with a particular focus on ligand binding capacity and molecularity in an *in vitro* context. Assessing the interaction with  $\text{Gdm}^+$ , we identified a striking difference between ACGA and ACGG loop motifs. While ACGA motifs tend to show kissing loop interactions together with  $\text{Gdm}^+$  as postulated in earlier studies, ACGG motifs interact prior to  $\text{Gdm}^+$  addition rendering them unavailable for intramolecular action. This is supported by SAXS and DOSY NMR data which indicate higher-order oligomers of Gdn49 wt under  $\text{Gdm}^+$  influence. In mutational studies, we observed that P1 and P2 are involved in (intramolecular) ligand recognition if adenosines are present in position four in the loop. If a single G variant is present, an intermolecular A-loop to A-loop interaction is preferred. In GG-constructs, the ligand is recognized but we detect a lower affinity. In addition, AG-constructs are likely influenced by P2-P2' interaction, responsible for wild-type oligomerization and therefore might bind ligand intermolecularly. In a biological context, P2-P2' interactions via G-mediated contacts are unlikely to occur but remain important to bear in mind for chemical biology applications. The reduced activity of riboswitch variants bearing an ACGG loop motif however can be considered a highly abundant tuning mechanism towards the extent of translational control. This is in line with the low abundance of GG variants in biological systems.

## Experimental Section

### Constructs of interest

The sequence of the *Escherichia coli* SugE guanidine-II riboswitch aptamer (Gdn-II, mini-ykkC RNA motif) was used (Figure 2B).

The construct is a composite of P1 (23 nt) and P2 (13 nt) hairpins and the native linker L1/2 (13 nt) connecting both. The 5' end was modified with G1 and G2 to enable an efficient *in vitro* transcription<sup>[32]</sup> and stabilize the P1 hairpin.

The mutant P1mut was modified with G instead of A14 and P2mut with A instead of G44.

### RNA preparation

The wild-type RNA constructs were purchased from Dharmacon Inc.

In addition, Gdn23 wt, Gdn49 wt and all its mutants were prepared in house through *in vitro* transcription with T7 RNAP.<sup>[33]</sup>

DNA templates included the necessary T7 promoter sequence and were either obtained from linearizing plasmid from pUC57 vector (GenScript) containing the full-length native sequence of Gdn-II riboswitch or PCR run-off according to the standard protocol by New England Biolabs® (0.5 mM of each primer, 200 mM dNTPs) using homemade Phusion polymerase. The primers were purchased from Eurofins Genomics (Germany).

Transcription reactions contained transcription buffer (100 mM Tris/glutamate pH 8.1), 2 mM spermidine, 20 mM dithiothreitol (DTT), 20% (v/v) DMSO, 5 mM of each NTP, 10 mM Mg(OAc)<sub>2</sub>, 9.6 μg/mL of homemade yeast inorganic pyrophosphatase (YIPP) and 32 μg/mL of homemade T7 RNAP. Unlabeled NTPs were purchased from Carl Roth GmbH + Co. KG (Germany) and <sup>13</sup>C, <sup>15</sup>N labeled NTPs from Silantes GmbH (Germany) or from Eurisotop (France).

The conditions were optimized for yield and sample purity. The *in vitro* transcription was performed in 15 mL scale. Purification was performed either by HPLC, preparative PAGE or buffer exchange to NMR buffer if necessary.<sup>[34]</sup>

### NMR spectroscopy

NMR samples were prepared by adding 8% D<sub>2</sub>O and 7.5 nmol DSS as internal reference to RNA stock solutions in NMR buffer (25 mM potassium phosphate buffer, 50 mM KCl, pH 6.2). All spectra were recorded of 280 μL samples in Shigemi NMR tubes (Shigemi Inc.).

NMR experiments were conducted on Bruker AV600, AV700, AV800, AV900 and AV950 spectrometers, equipped with cryogenic probes. Data were processed with Bruker TopSpin® 3.6.1 (Bruker Biospin) and NMRFAM-SPARKY 3.135.<sup>[35]</sup> Water suppression was achieved using WATERGATE<sup>[36]</sup> or jump-and-return echo<sup>[37]</sup> water suppression pulse schemes.

Analysis of dissociation constants ( $K_D$ ) were carried out by measuring intensities of the imino proton signals of U17 in <sup>1</sup>H-1D spectra for Gdm<sup>+</sup> binding to the Mg<sup>2+</sup>-RNA-complex using the one-site binding hyperbole (equation 1) or Hill formalism (equation 2) as fitting function.<sup>[17,38]</sup> In equation 2, the postulated cooperativity  $n$  is considered. The unbound signal was correlated with the bound signal, then normalized and subsequent fitted while leaving the RNA concentration variable.

$$y_{\text{one-site-binding}} = B_{\text{max}} \cdot x / (K_D + x) \quad (1)$$

$$y_{\text{Hill}} = B_{\text{max}} \cdot x^n / (K_D^n + x^n) \quad (2)$$

DOSY measurements were evaluated either by using the mean values of the aromatic range or all ranges to obtain the logarithmic value for the diffusion constant  $D$  (Supporting Figure S6). Com-

pared to the internal standard 1,4-dioxane with known hydrodynamic radius of 2.14 Å, the hydrodynamic radius of the investigated Gdn49 wt is determined via the equation 3:

$$r_{\text{RNA}} = \frac{D_{1,4\text{-dioxane}}}{D_{\text{RNA}}} \cdot r_{1,4\text{-dioxane}} \quad (3)$$

### SAXS

SAXS measurements were performed on BM29 BioSAXS beamline of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France in a remote measurement mode. Sample buffers were identical to NMR conditions except for D<sub>2</sub>O. Sample volumes were 60 μL and the following concentrations have been used 2.5 mg/mL for free and Mg<sup>2+</sup>-bound and 2 mg/mL for Mg<sup>2+</sup>/Gdm<sup>+</sup>-bound guanidine-II aptamer wild-type constructs (Gdn23 wt and Gdn49 wt). We also measured all samples at higher concentrations of 4–5 mg/mL. We consistently used the more dilute sample for downstream analysis in order to avoid non-specific effects from concentration with respect to monomer-dimer-oligomer equilibria. Notably, also at lower concentrations, both the S/N and data quality were of sufficient quality. We used the respective filtrates during sample concentrating as buffer match to subtract background scattering. Samples were measured by exposure at 12.5 keV beam energy, scattering was acquired on a Pilatus3 2 M detector placed under vacuum, and with local standard setups of distance and beam size. The herein used q-range for RNA samples was 0.25–5.2 nm<sup>-1</sup>. Acquired frames were automatically scanned for sufficient quality and to exclude radiation damage and summed up over a period of 5–10 sec of measurement. Buffer subtraction was performed using an average from prior and posterior buffer-only runs. For data analysis, both the pre-processed data were visualized manually and the estimation of  $R_g$ ,  $D_{\text{max}}$  and MW was carried out using ATSAS 3.0.<sup>[39]</sup> Structural alignment was performed using the SASpy PyMOL plugin.<sup>[40]</sup>

### Acknowledgements

We thank Dr. Martin Hengesbach, Dr. Nadine Schwierz, Dr. Robbin Schnieders, Christin Fuks, Elena Bierwirth and Sarah Grudzinski for insightful discussions. Further, we thank Elke Stirnal and Kerstin Witt for HPLC purification and Dr. Christian Richter for help and maintenance of NMR spectrometers. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation beam time and we would like to thank Mark Tully for assistance in using beamline BM29. Karthikeyan Dhamotharan is also to be thanked for taking over our in-house SAXS organization. Supported by the DFG in the Collaborative Research Center 902; Work at BMRZ is supported by the state of Hesse. A.S. is supported through DFG grant number SCHL2062/2-1 and by the Johanna Quandt Young Academy at Goethe (stipend number 2019/AS01). Open Access funding enabled and organized by Projekt DEAL.

### Conflict of Interest

The authors declare no conflict of interest.



## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon request.

**Keywords:** aptamers · guanidine riboswitch · NMR spectroscopy · RNA · structural biology

- [1] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043–1049.
- [2] W. Winkler, A. Nahvi, R. R. Breaker, *Nature* **2002**, *419*, 952–956.
- [3] T. E. Edwards, A. R. Ferré-D'Amaré, *Structure* **2006**, *14*, 1459–1468.
- [4] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15908–15913.
- [5] A. Roth, R. R. Breaker, *Annu. Rev. Biochem.* **2009**, *78*, 305–334.
- [6] A. Serganov, D. J. Patel, *Biochim. Biophys. Acta* **2009**, *1789*, 592–611.
- [7] A. Wacker, J. Buck, D. Mathieu, C. Richter, J. Wöhnert, H. Schwalbe, *Nucleic Acids Res.* **2011**, *39*, 6802–6812.
- [8] A. Ren, X. C. Wang, C. A. Kellenberger, K. R. Rajashankar, R. A. Jones, M. C. Hammond, D. J. Patel, *Cell Rep.* **2015**, *11*, 1–12.
- [9] N. Sudarsan, E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link, R. R. Breaker, *Science* **2008**, *321*, 411–413.
- [10] O. Binas, T. Schamber, H. Schwalbe, *Nucleic Acids Res.* **2020**, *48*, 6970–6979.
- [11] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, R. R. Breaker, *Science* **2012**, *335*, 233–236.
- [12] A. Ramesh, W. C. Winkler, *RNA Biol.* **2010**, *7*, 77–83.
- [13] J. W. Nelson, R. M. Atilho, M. E. Sherlock, R. B. Stockbridge, R. R. Breaker, *Mol. Cell* **2017**, *65*, 220–230.
- [14] C. W. Reiss, S. A. Strobel, *RNA* **2017**, *23*, 1338–1343.
- [15] R. R. Breaker, R. M. Atilho, S. N. Malkowski, J. W. Nelson, M. E. Sherlock, *Biochemistry* **2017**, *56*, 345–347.
- [16] M. M. Meyer, M. C. Hammond, Y. Salinas, A. Roth, N. Sudarsan, R. R. Breaker, *RNA Biol.* **2011**, *8*, 5–10.
- [17] M. E. Sherlock, S. N. Malkowski, R. R. Breaker, *Biochemistry* **2017**, *56*, 352–358.
- [18] M. Sinn, F. Hauth, F. Lenkeit, Z. Weinberg, J. S. Hartig, *Mol. Microbiol.* **2021**, *116*, 200–210.
- [19] H. Salvail, A. Balaji, D. Yu, A. Roth, R. R. Breaker, *Biochemistry* **2020**, *59*, 4654–4662.
- [20] L. Huang, J. Wang, D. M. J. Lilley, *Cell Chem. Biol.* **2017**, *24*, 695–702.e2.
- [21] R. A. Battaglia, A. Ke, *WIREs RNA* **2018**, *9*, e1482.
- [22] Z. Weinberg, J. E. Barrick, Z. Yao, A. Roth, J. N. Kim, J. Gore, J. X. Wang, E. R. Lee, K. F. Block, N. Sudarsan, S. Neph, M. Tompa, W. L. Ruzzo, R. R. Breaker, *Nucleic Acids Res.* **2007**, *35*, 4809–4819.
- [23] J. Noeske, J. Buck, J. Wöhnert, H. Schwalbe, *Non-Protein Coding RNAs* **2009**, 229–247.
- [24] C. Wuebben, M. F. Vicino, M. Mueller, O. Schiemann, *Nucleic Acids Res.* **2020**, *48*, 10518–10526.
- [25] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem* **2003**, *4*, 936–962.
- [26] C. E. Dann, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, W. C. Winkler, *Cell* **2007**, *130*, 878–892.
- [27] E. Bonneau, P. Legault, *Biochemistry* **2014**, *53*, 579–590.
- [28] C. S. Johnson Jr, *Nucl. Magn. Reson.* **1999**, *34*, 203–256.
- [29] B. Chen, X. Zuo, Y.-X. Wang, T. K. Dayie, *Nucleic Acids Res.* **2012**, *40*, 3117–3130.
- [30] J. Lipfert, D. Herschlag, S. Doniach, *Methods Mol. Biol.* **2009**, *540*, 141–159.
- [31] M. Popenda, M. Szachniuk, M. Antczak, K. J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R. W. Adamiak, *Nucleic Acids Res.* **2012**, *40*, 1–12.
- [32] J. F. Milligan, O. C. Uhlenbeck, *Methods Enzymol.* **1989**, *180*, 51–62.
- [33] J. K. Bains, J. Blechar, V. de Jesus, N. Meiser, H. Zetzsche, B. Fürtig, H. Schwalbe, M. Hengesbach, *Methods* **2019**, *153*, 22–34.
- [34] C. Helmling, S. Keyhani, F. Sochor, B. Fürtig, M. Hengesbach, H. Schwalbe, *J. Biomol. NMR* **2015**, *63*, 67–76.
- [35] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.
- [36] M. Liu, X. Mao, C. Ye, H. Huang, J. K. Nicholson, J. C. Lindon, *J. Magn. Reson.* **1998**, *132*, 125–129.
- [37] G. Marius Clore, B. J. Kimber, A. M. Gronenborn, *J. Magn. Reson.* **1983**, *54*, 170–173.
- [38] D. Leipply, D. E. Draper, *Biochemistry* **2010**, *49*, 1843–1853.
- [39] K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S. Molodenskiy, A. Panjkovich, H. D. T. Mertens, A. Gruzinov, C. Borges, C. M. Jeffries, D. I. Svergun, D. Franke, *J. Appl. Crystallogr.* **2021**, *54*, 343–355.
- [40] A. Panjkovich, D. I. Svergun, *Bioinformatics* **2016**, *32*, 2062–2064.
- [41] M. Baalousha, F. V. D. Kammer, M. Motelica-Heino, H. S. Hilal, P. Le Coustumer, *J. Chromatogr. A* **2006**, *1104*, 272–281.

---

Manuscript received: October 20, 2021  
Revised manuscript received: November 30, 2021  
Accepted manuscript online: November 30, 2021  
Version of record online: December 9, 2021