

Structural and functional investigation of nucleic acids with NMR spectroscopy

Dissertation
zur Erlangung des Doktorgrads
Der Naturwissenschaften

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

von
Oliver Binas
Aus Frankfurt am Main

Frankfurt am Main

2020

(D30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der Goethe Universität Frankfurt als
Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Erster Gutachter: Prof. Dr. Harald Schwalbe

Zweiter Gutachter: Prof. Dr. Alexander Heckel

Datum der Disputation:

Meiner Großmutter, Anneliese Fiedler

This thesis was prepared under the supervision of Prof. Dr. Harald Schwalbe between November 2015 and August 2020 at the Institute for Organic Chemistry and Chemical Biology of the Goethe University in Frankfurt, Germany.

Table of contents

1. Introduction and Summary	9
2. About research on nucleic acids	11
3. RNA structures.....	13
3.1. Primary structure	13
3.2. Secondary structure.....	13
3.2.1. Helices	14
3.2.2. Hairpins	15
3.2.3. Bulges and internal loops	15
3.2.4. Junctions	16
3.2.5. G-quadruplexes	16
3.2.6. Other structural motifs.....	17
3.3. Tertiary interactions	19
3.3.1. Coaxial stacking interactions	19
3.3.2. Loop-Loop interactions.....	19
3.3.3. Pseudoknots	19
4. Natural RNAs and their role in gene regulation	21
4.1. Overview	21
4.2. Small molecule-dependent regulation in prokaryotic 5'-UTRs: Riboswitches.....	24
4.2.1. Riboswitch types and modes of regulation	24
4.2.2. Common riboswitches: Structural and functional properties	25
4.3. Gene regulation by AU-rich sequences and constitutive decay elements in the 3'-UTR of eukaryotic mRNA.....	35
4.4. G-quadruplexes in gene regulation.....	36
5. Non-natural RNA: Practices and applications	41
5.1. Artificial miRNA and RNA based drugs.....	41
5.2. Artificial functional RNA: Aptamers and Ribozymes.....	42
5.3. Nano applications and molecular machines.....	43
6. NMR spectroscopy of nucleic acids	45
6.1. NMR methods applied to nucleic acids.....	45

6.1.1.	¹ H-1D experiments.....	45
6.1.2.	Reporter signals and probes	46
6.1.3.	Resonance assignment and NOESY	47
6.1.4.	2D-HSQC and related hetero correlated experiments.....	47
6.1.5.	Isotopic labeling	48
6.1.6.	2D- and 3D-NMR methods aiding resonance assignment and structural characterization.....	49
6.2.	Structure determination of nucleic acids by NMR spectroscopy	50
6.3.	NMR of G-quadruplexes	51
7.	RNA and NMR in drug development.....	53
7.1.	Targeting RNA with small molecules	53
7.2.	Drug discovery by NMR	54
7.3.	Fragment-based drug discovery.....	56
8.	References	58
9.	Research article I: The conformational landscape of transcription intermediates involved in the regulation of the ZMP-sensing riboswitch from <i>Thermosinus carboxydivorans</i>	74
10.	Research article II: Structural basis for the recognition of transiently structured AU-rich elements by Roquin	94
11.	Research article III: Structure validation for G-rich RNAs in non-coding regions of the human genome.....	128
12.	Research article IV: Photoresponsive Formation of an Intermolecular Minimal G- Quadruplex Motif.....	140
13.	Research article V: ¹⁹ F-NMR-based fragment screening for 14 different biologically active RNAs and 10 DNA and protein counter-screens.....	200
14.	Research article VI: Book chapter: NMR on Ribozymes in “Ribozymes” (Wiley-VCH)....	282
15.	German Summary – Zusammenfassung in deutscher Sprache	304
16.	Danksagung	310
17.	Lebenslauf	312

1. Introduction and Summary

Nucleic acids are involved in numerous biological processes including the long-known storage of genetic information (DNA) and transfer of genetic information (RNA) but also many, regulatory roles.^[1] The ability of nucleic acids and especially RNA to form three-dimensional structures and complex recognition motifs allows for manifold interaction modes with protein cofactors^[2] but also metabolites^[3]. These features are increasingly recognized and spark not only the scientific interest in understanding the underlying biological processes, but also in functionalization of the properties of nucleic acids^[4–6]. In the scientific studies conducted as part of this thesis, nucleic acids are studied with liquid state nuclear magnetic resonance (NMR) spectroscopy, a highly versatile biophysical method, which despite its comparatively high instrumental efforts and cost is widely used in life sciences.^[7–10] It combines flexibility in experiment choice and resulting applicability to various experimental approaches with atomic resolution and monitoring of molecules in the liquid state, a property often desired for investigation of biological processes.

This cumulative thesis includes distinct studies, which all involve NMR spectroscopy and nucleic acids, structured into two parts. The introductory chapters 2 – 7 feature state-of-the-art structural and functional research on nucleic acids as well as NMR spectroscopy applied to nucleic acids. They comprise insights into structural characteristics of nucleic acids, diverse roles of RNA in gene regulation, non-natural RNA in medicine and nanotechnology, NMR spectroscopy of nucleic acids and NMR-based drug discovery with RNA as a target. These chapters are meant to establish an introductory basis for the following research articles, as introductions in those articles themselves are often highly concise and specific. Notably, since research article VI is a book chapter and therefore serving an educational approach already, no introduction on the topic of “NMR on Ribozymes” is provided apart from the short introduction directly before the article. In the second part (Chapters 9 – 14), six research articles are featured, which include the scientific advances achieved during the research period. In research article I, functional studies by NMR spectroscopy were performed on a riboswitch, which senses the cell metabolite ZMP^[11]. Riboswitches are RNA regulatory elements located in the non-coding region of many bacterial mRNAs^[12]. NMR studies allowed the detailed characterization of RNA secondary structures attained during transcription in an *in vitro* model system. Research article II comprises structural and functional studies on a human regulatory RNA element, which is recognized by a protein cofactor^[13]. The structural basis for the protein interaction and the undocked element was established by X-ray and NMR structure determination, respectively. Additionally, the RNA-protein interaction was investigated utilizing differential biophysical methods, including NMR spectroscopy. Structurally distinct human RNA regulatory elements are investigated in research article III. Six RNA sequences were investigated for their ability to form highly stable structural elements called G-quadruplexes^[14]. In a short biophysical protocol, NMR-spectroscopy and circular dichroism spectroscopy was utilized to obtain information on G-quadruplex formation and additional insight into the G-quadruplex topology. G-quadruplex formation also played a role in research article IV, which includes the development and structural characterization of a light-switchable DNA G-quadruplex. NMR- CD- and UV-VIS- spectroscopic methods were applied to this interesting system to characterize switching behavior, reversibility and the three-dimensional

structure. Research article V comprises fragment-based drug discovery^[15] studies carried out on 24 biological targets including 14 RNAs. The screening was performed using a library of fluorine-containing compounds allowing ¹⁹F-NMR-based observation. Experiments additionally included various counter-screens, characterizations of binding and chemical follow-up experiments. The final research article VI is a book chapter about ribozymes^[16] researched with NMR spectroscopy. It is reported how these self-catalyzing RNAs were investigated by NMR spectroscopic methods and the chapter features not only many literature examples, but also general guidelines for NMR sample preparation and experiments.

Since very different types of nucleic acids were investigated spanning from natural regulatory riboswitch RNA to chemically modified light-controllable G-quadruplex DNA, it was possible to address many upcoming and highly differential experimental tasks with NMR-spectroscopy. Thus, the thesis not only highlights the structural and functional flexibility of nucleic acids, but also the immense flexibility of NMR-spectroscopy as a scientific method.

2. About research on nucleic acids

At latest since the discovery of the DNA structure by Watson and Crick 1953, research on nucleic acids became a major field of scientific advances.^[17] This finding led to numerous studies with the aim of deciphering biological processes taking place in cells and ultimately the postulation of the central dogma of molecular biology.^[18] This dogma clearly defined DNA as the molecule enabling information storage, RNA acting as a carrier molecule (as messenger RNA, mRNA) or template molecule (ribosomal or transfer RNA, rRNA/tRNA) and proteins being the cellularly active species. Consequently, intensive research was focused on proteins, for a considerable time span. Additionally, at that time, it was consensus, that a single gene always codes for a single protein (One-Gene-One-Enzyme hypothesis)^[19]. Many of these considerations were proven wrong later, after several small RNAs were found in biochemical fractionation experiments^[20] and with the discovery of splicing in 1970^[21] the One-Gene-One-Enzyme hypothesis was disproven. Still, with these discoveries and postulations a basis for a deeper understanding of biological processes was established.

Especially the understanding of the role of RNA was subject to changes in modern days science, after many occasions were observed in which RNA is involved as a regulating molecule, interacts with proteins or even catalyzes biological processes without the influence of protein cofactors. Examples are peptide bond formation in the ribosome, miRNA (micro RNA) which is involved in transcriptional regulation, snRNA (small nuclear RNA) which acts as part of the spliceosome or riboswitches and ribozymes, which commonly occur in prokaryotes.

While the different types of RNA molecules fulfil a multitude of cellular functions, they consist of sequences of only four different basic building blocks, namely the nucleobases adenine (A), cytosine (C), guanine (G) and uracil (U) and therefore are can be considered less complex than proteins. It is a widespread hypothesis that the evolution of live as we know it took a route over an intermediate step of an RNA-world in which information storage but also chemical catalysis was RNA based.^[12] Prokaryotic enzymatically active RNA such as ribozymes are often considered relics of these developments. Nevertheless, RNA plays an important role in highly developed eukaryotic cells exceeding its formerly assumed function of carrier and template molecule. In the subsequent chapters, structural and functional aspects of RNA in nature as well as synthetic RNA are discussed.

3. RNA structures

The multitude of functions employed by nucleic acids in the cellular context requires strong structural polymorphism. Opposed to the mostly double stranded DNA, RNA molecules are observed in highly distinct conformations, often described as folds. While these folds are in many cases based on canonical base pairing, which is also prevalent in DNA, many non-canonical patterns of base pairing are observed in RNA, which are often highly important for the overall structure and the specific characteristics of the RNA. Additionally, these structures frequently show secondary structural motifs exceeding the complexity of helices and additional stabilizing tertiary interactions. Such structural characteristics are discussed in the following chapter.

3.1. Primary structure

The sequence of bases, connected by the phosphate backbone of nucleic acids is described as primary structure. As per convention, the 5'-end of the DNA or RNA strand is defined as the beginning of the sequence and the 3'-end as the end of the sequence. This perspective is related to the natural direction of translation in which the RNA is synthesized beginning with the 5'-end. In DNA the primary structure already contains all genetic information. According to its biological function as carrier of information, the function of a specific DNA is in principle clear as soon as the sequence is known. Contrarily, as RNA molecules adopt many different structures and carry out many biological tasks, knowledge of their sequence is insufficient to understand its function in most cases. Accordingly, secondary and tertiary structural motifs must be considered in order to explain how RNAs exert their function biologically.

3.2. Secondary structure

Secondary structure in nucleic acids is stabilized by base-pairing. It forms as two nucleobases interact by developing hydrogen bonds (H-bonds), irrelevant if both nucleobases are part of the same RNA molecule. The by far most common nucleobases are the canonical nucleobases adenine (A), guanine (G), cytosine (C) uracil (U, in RNA) and thymine (T, in DNA instead of U), even though modified residues also occur in nature. Interaction occurs by the formation of H-bonds at one of the three interaction sites of each base, which are described as Watson-Crick edge, Hoogsteen edge and sugar-edge.^[22] Per definition, base pairing requires the formation of at least two H-bonds. This allows, according to the definition of Westhof et al., the formation of six different base pairs each in trans and cis configuration. Saenger et al. define even 29 different possible base pairs (with at least 2 H-bonds) as they take not only the edges involved, but also their orientation into account.^[23]

Of the many possible base-pairs, the canonical Watson-Crick base pair is by far the most important. In these base pairs, guanine and cytosine or adenine and uracil pair with their Watson-Crick edges with average binding free energies of -5.8 kcal/mol (G-C) or -4.3 kcal/mol (A-U), amounting for approximately 2 kcal/mol per H-bond.^[24] According to Westhof et al. they are described as *cis*-Watson-Crick (*cis*-WW) base pairs.^[22] Apart from the canonical base pairs, the most abundant natural base pairs are *trans*-WW, *trans*-Watson-Crick-Hoogsten (*trans*-WH) and three different interactions of the sugar edge.^[25] **Figure 1** shows three exemplary base pairs and

their description according to Westhof or Saenger. If only the basic geometry is observed, the G-U Wobble base pair would be the same as a canonical base pair. However, the base G and U interact in a shifted geometry to be able to form two H-bonds. The *cis*-WC-H base-pair though, is constructed entirely different and can be distinguished from the other two in both classification systems.

Apart from base pairing, stabilizing interactions are observed in stacking interactions of two nucleobases, as a result of interaction of their π -systems. These stacking interactions are highly prevalent in helices (*vide infra*). Quantum mechanical calculations estimate energies for the interaction of two nucleobases to be -10 kcal/mol in the gas phase. However, under influence of solvent, free enthalpies amount to -0.3 to -1.1 kcal/mol. Worth noting is the relatively high contribution of purines with the most advantageous interaction being between G and G with -1.1 kcal/mol.^[26] These interactions account for stabilization of the secondary structure.

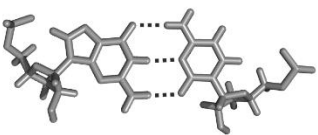
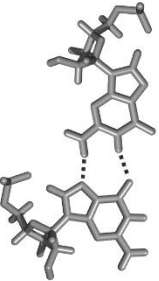
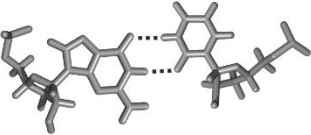



Structure			
General description	G-C- (depicted) / A-U- base pair	G-G Hoogsteen base pair	G-U wobble base pair
Bases involved	Guanine-cytosine oder adenine-uracil (G-C / A-U)	Guanine-guanine (G-G)	Guanin-uracil (G-U)
Description according to Miao & Westhof ^[22]	<i>cis</i> Watson-Crick / Watson-Crick (<i>cis</i> -WW)	<i>cis</i> Watson-Crick / Hoogsteen (<i>cis</i> -WH)	<i>cis</i> Watson-Crick / Watson-Crick (<i>cis</i> -WW)
Symbol according to Miao & Westhof ^[22]			
Symbol according to Saenger ^[23]	XIX / XX	VI	XXVIII
Number of H-bonds	3 / 2	2	2

Figure 1 | Overview of three base pairs commonly found in RNA structures. Schematic showing G-C Watson-Crick, G-G Hoogsteen and G-U Wobble base pairs and their different nomenclatures according to Miao & Westhof or Sanger.

3.2.1. Helices

The most important secondary structure motif for nucleic acids is the helix. In helices, DNA or RNA strands interact through complimentary base pairing and are aligned in a wound helical structure, stabilized by stacking interactions. The strands are, apart from some unusual occurrences^[27], antiparallel aligned. Helical interaction leads to the formation of two grooves that are described as major and minor groove according to their size. In nature, three types of helices, A-, B- and Z- helices are observed (**Figure 3** top). The most notable differences between the two right-handed types A and B are observed in the tilt of the nucleobases with respect to the main axis (roll) and the shift of the base pairs orthogonally to the main axis (slide). Additionally in A-helices, the configuration of the sugar (sugar-pucker) is *C3'-endo*, while it is *C2'-endo* in B-

helices.^[28] In both cases the nucleobases are *anti* configured with respect to the sugar moiety. The distinctly rarer Z-helix is constructed differently. Its winding is left-handed and the sugar-pucker as well as the base-sugar alignment is dependent on the base (C: *anti* and *C2'-endo*, G: *syn* and *C3'-endo*). Additionally, the nucleobases are packed 14% more tightly compared to B-DNA.^[29] Naturally occurring DNA is found mostly double stranded and in B-conformation. Contrarily, double stranded RNA is generally observed in A-conformation. Z-conformation is attained only in cases of alternating deoxy purine and deoxy pyrimidine bases especially in CG repetitions. Accordingly, the principle building block of Z-DNA consists of two base pairs.

In the biological context, RNA is involved in many catalytic and regulatory processes (see Chapter 4). Consequently, in order to carry out their functions, RNAs adopt many additional folding motifs apart from simple helical structures, which are described in the following passage.

3.2.2. Hairpins

Hairpins form from self-complimentary single stranded RNA regions. Through canonical base pairing a double stranded helical stem and a partially structured loop region are formed (**Figure 3**). The length of the loop is variable and, in most cases, lies between three and eleven nucleotides, while loops with four and five nucleotides are the most common^[30] and also most stable.^[31] Tetraloops are the best studied systems and many sequentially different tetraloops were characterized according to their phylogenetical conservation and structure.^[32-35] Especially tetraloops with the sequences UNCG (N: A/C/G/U) and GNRA (R: C/U), which resemble 70% of tetraloops in ribosomal RNA (rRNA)^[36] are well studied. In GNRA tetraloops the conserved bases G and A interact in a *trans*-Hoogsteen-Sugar (*trans*-HS) base pair while residues R and N are stacked on top of the A residue.^[37] In analogy to GNRA loops, the outer U and G bases of the UNCG tetraloop interact in a comparable way, with the imino proton H1 of the G-residue interacting with the O2 atom of the U residue. The C2'-endo configured sugar of N interacts with the pi-system of G and C is located above the interaction site of G and U.^[38]

3.2.3. Bulges and internal loops

If a single stranded region in RNA helices is not located apical to a stem (as in hairpins), but in the middle, it is referred to as a bulge or an internal loop (**Figure 3**). The terms are used in accordance to the geometry of the unpaired region.

If the unpaired region is symmetrical on both strands of the helix it is referred to as an internal loop, but if it is unsymmetrical, especially if only one strand contains bulged nucleotides, it is referred to as a bulge. Because of the disturbance of the helical structure, bulges lead to stem destabilization of about 1 kcal/mol per single nucleotide bulge.^[39] The disturbance of the structure can be observed as a kink in the helix when a single bulged nucleotide intercalates between two Watson-Crick base pairs.^[40] In larger bulges, the respective residues are often turned outside of the helix where they are involved in stacking interactions.^[41] This leads to deformation of the RNA backbone and negatively charged backbone atoms, which are usually exposed to the solvent, point to the inside of the helix where they can interact with positively charged metal ions^[42] or

ligands.^[43] In general, deformation of the RNA structure by bulges can be of critical importance for the function of an RNA.^[44,45]

Internal loops are characterized by the abundance of non-canonically base paired or unpaired residues on both sides of an RNA double strand. According to the attained symmetry, energetical differences exerted by these motifs can be striking. Symmetrical internal loops can stabilize an RNA structure by up to 2 kcal/mol, while unsymmetrical loops lead to destabilization of a comparable degree.^[46] An important example for a stabilizing internal loop is the three nucleotide GGA-AAG internal loop, which occurs in rRNA and is composed of three t-HH base pairs.^[47] Comparable G-A interaction also play a role in the wide-spread kink-turn motifs. These structural motifs, firstly observed in rRNA, feature at least two interacting G-A base pairs followed by a three nucleotide bulge and lead to a 120° kink between two adjacent stems.^[48] If residues in the internal loop are only weakly bound or fully exposed to the solvent, they are often involved in inter- or intermolecular interactions.^[49–51]

3.2.4. Junctions

Regions in which a single helix transitions into two or more other helices are described as junctions (**Figure 3**). By far the most common junctions are three-way junctions and the prevalence decreases with an increasing number of helices involved.^[52] Still, even six-way junctions are described, for instance in the Group-I intron.^[53] In the well-studied three- and four-way junctions coaxial stacking (see **Chapter 3.3.1**) is often observed, which stabilizes the overall structure.^[52] Accordingly, many three-way junctions, despite being displayed as a Y-shape in schematics, do in fact assume a T-shaped structure. Lescoute and Westhof classified three-way junctions into three structural families A, B and C defining rules which allow a rough estimation of the type of junction from the RNA sequence.^[54] In the more complex field of four-way junctions, nine structural families, observable in nature, were defined by Laing and Schlick.^[55] They differ in the arrangement of the helices and their connectivity and include in almost every case one or two instances of coaxial stacking. As a result of the high abundance of unpaired residues in junctions as well as their complex structural space, they are often crucial for catalytical or regulatory action of RNA as in ribozymes^[56] or riboswitches (see **Chapter 4.2**).^[57]

3.2.5. G-quadruplexes

An unusual structure formed by nucleic acids is the G-quadruplex (G4). This structure is formed from guanine-rich RNA and DNA sequences under influence of monovalent cations such as K⁺, Na⁺ or NH₄⁺. The G-rich sequences self-align in tetrads in which each G-residue interacts by its Watson-Crick edge with the Hoogsteen edge of another G-residue (*cis*-WH). A stable G4 is formed when at least two of these tetrads stack on top of each other stabilized by constructive stacking interactions (**Figure 3**). The cations are located in the central ion channel formed in the middle of the nucleobases, where they interact with the O6 atoms of the G-residues in order to stabilize the G4. The exact localization of the ions depends on their ion radius, with smaller ions located in the middle of the tetrads and larger ions between the tetrads.^[58] In principle, very large assemblies are possible, which are often described as G-wires.^[59] More common and also more extensively studied are G4s with two to four stacked tetrads.^[60] In principle, G4s can be formed by

four distinct nucleic acid molecules. In this case, each strand would not be connected to any other strand and accordingly, bimolecular G4s can exist as well. However, G4s are most commonly formed from a single nucleic acid strand.^[58] In this case, other residues located between the G-residues involved in tetrads, form topologically distinct loops according to the respective orientation of the G-strands. Three loop geometries are prevalent in G4s, called propeller, lateral and diagonal loops (**Figure 2** left). Strands can be oriented in parallel, antiparallel or 3+1 hybrid topology (**Figure 2** right). Additionally, the sequence of loops can be oriented in both directions around the G-tetrad block, starting from the 5'-end. Permutation of these characteristics allows 26 different overall topologies for a monomolecular G4.^[61] Nucleobases, with respect to their sugar moiety, can adopt *syn* or *anti* conformation, achieved by variation of the glycosidic bond angle χ . Since lateral and diagonal loops always connect antiparallel strands and propeller loops always connect parallel strands, the configuration of the glycosidic bond angle relates to the loop topology.^[62] RNA G4s are with very few exceptions^[63,64] found in parallel topology with three propeller loops. This is a result of the *syn*-conformation being much more unfavorable for RNA than for DNA. Antiparallel RNA G4 conformations are obtainable by chemical modification at the 8-position of the G-residues, though.^[65] In case of DNA all 26 topologies are possible.

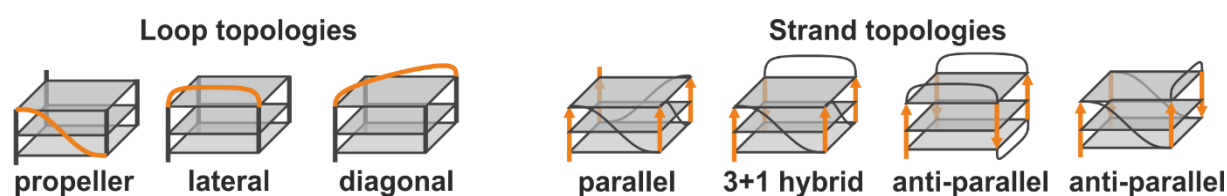


Figure 2 | Possible topologies for loops and strand in G-quadruplexes. Loops and strands the topology refers to are marked orange.

3.2.6. Other structural motifs

Another structural motif which does not include the formation of typical helices is the i-motif (intercalated motif). It is based on the interaction of hemiprotonated cytosines in C-rich sequences. The proton is localized at the N3 position and two C-residues interact by their Watson-Crick edges (*cis*-WW). The C-residues are part of four antiparallel aligned strands in which base-pairs are alternating between two pairs of strands.^[66] Since C-residues must be protonated in order to form an i-motif, its formation is highly dependent on low pH. Recent studies though show formation of an i-motif under special conditions even at neutral pH.^[67] In nature, i-motifs are suggested to form mainly from the antisense strands of G4 forming sequences^[67], but also in untranslated regions (UTRs) of mRNA.^[68] The stability of an i-motif *in vivo* could be confirmed via NMR.^[69]

Upon interaction of single stranded RNA with an A-form RNA-Duplex, an RNA Triplex can form. In these structures, the single stranded RNA can interact with the major or minor groove of the helix. In major groove triplexes stabilizing U-A-U base triplets are formed in which nucleobases of the formerly single-stranded RNA interact with the Hoogsteen edges of the canonically interacting residues (**Figure 3**). The structure is formed from A-U rich sequences and is stable under

influence of Mg^{2+} ions.^[70] Isolated minor groove triplexes are unstable and only form as part of larger RNA structures like in self splicing introns^[71], riboswitches^[72] or rRNA^[48]. They are often described as A-minor triplex.

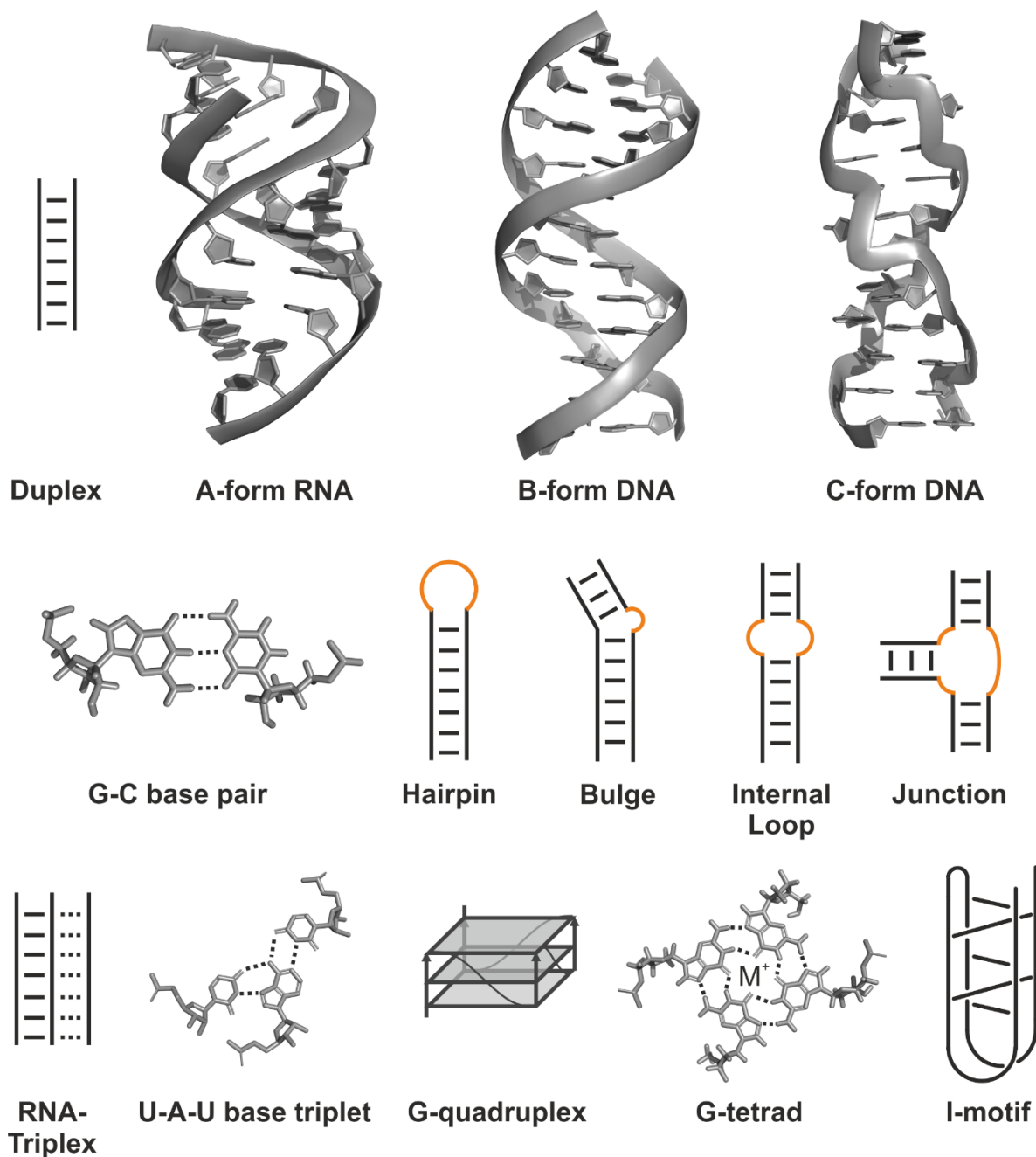


Figure 3 / Structural motifs formed by nucleic acids. Different forms of duplexes and base arrangements are shown as three-dimensional structures, while more complex structures are shown schematically.

3.3. Tertiary interactions

As already mentioned earlier, RNA shows, contrary to DNA, a broad spectrum of possible structural motifs, which allow for the diverse biological functions employed by RNA. In larger structures commonly several of the secondary structural motifs mentioned above can be observed, especially hairpins, bulges, internal loops or junctions which display many unpaired or weakly paired residues. The interaction of these residues which lead to the formation of the three-dimensional overall structure are described as tertiary interactions.

3.3.1. Coaxial stacking interactions

In structure formation of RNA, interaction of base-pairs through stacking is energetically advantageous (See **Chapter 3.2.1**). Accordingly, upon interaction of two principally distinct helices constructive interaction between their terminal base-pairs is observed. This leads to stacking of two helices with a colinear main axis, described as coaxial stacking. The interaction is especially common in the context of junctions (see **Chapter 3.2.4**), but is also observed in conjunction with other tertiary interactions described later in this chapter. Generally, coaxial interactions are a common tertiary structural motif whose stabilizing effects are often the foundation of the formation of more complex three-dimensional structures.^[73,74]

3.3.2. Loop-Loop interactions

Two sequentially separated, but spatially close loop structures can interact with each other through their unpaired nucleobases. This type of interaction is described as loop-loop interaction or kissing-loop interaction in the literature. It was observed for the first time in tRNA, in which the T-loop interacts with the D-loop^[75] and occurs in numerous biologically interesting RNAs like riboswitches^[76], ribozymes^[77] and viral RNA.^[78,79] Upon interaction of two long loops, coaxial stacking can also be involved to stabilize the overall structure.^[80]

3.3.3. Pseudoknots

When loop nucleobases interact with complementary nucleobases of sequentially separated formerly unpaired residues such as longer bulges, the interaction is described as a pseudoknot. A helix is formed, which is coaxially stacking to the helix, the unpaired strand arises from.

4. Natural RNAs and their role in gene regulation

As formerly stated, the breaking of the central dogma of cell biology represented a turning point in research on nucleic acids, leading to many advancing and sometimes groundbreaking discoveries, such as splicing^[21], the tertiary structure of tRNA^[81] or recently the discovery of the CRISPR/Cas9 system^[82]. Up to now, over 20 different types of RNA have been classified according to their respective functions many of which are involved in gene regulation, spanning from regulation of transcription and translation to degradation or even catalytic capabilities. In this chapter some of the most important regulatory active RNAs and their functions are discussed, while focusing especially on regulatory active RNA which resides in the untranslated regions of eukaryotic and prokaryotic mRNA.

4.1. Overview

Messenger RNA (mRNA) was one of the earliest discovered types of RNA, right after the discovery of genetic coding in 1961^[83] and can be considered the most important type of RNA, as it holds the genetic information between the two major cellular processes of transcription and translation. In eukaryotes the product of DNA transcription is termed pre-mRNA, which subsequently undergoes splicing and capping processes resulting in the generation of mature mRNA which can then be translated into proteins. In prokaryotes though, splicing does not take place and the transcribed mRNA is directly translated. Since mRNA is directly involved as a product in transcription and as the template in translation, sequences located on the mRNA can consequently act as key players in regulation of these processes.

Transcriptional control in eukaryotes is exerted by trans-acting protein transcription factors sensing promotor elements of which the most important are the TATA-box and the Inr initiator sequence (**Figure 4**, top).^[84,85] The mRNA itself usually does not play a role in these regulatory processes. In prokaryotes, transcription is controlled by comparable trans-acting, but often also cis-acting elements, such as riboswitches, which will be discussed in detail later in this thesis (see **Chapter 4.2** and **Figure 4** bottom, marked red).

In order to control eukaryotic translation, many partially competing pathways exist involving control of cap- or IRES-dependent initiation^[86,87] but also mRNA degradation^[88]. Most regulatory effects in eukaryotes are exerted on the proteins side. However, a common process observed in eukaryotic mRNAs is the action of so-called upstream open reading frames (uORF, **Figure 4** top, marked red). These elements are located in the 5'-untranslated region (5'-UTR) of a gene and upon translation, their peptide product attenuates transcription.^[89] By comprising sequences tuned to "sense" the metabolic status of the cell they can lead to preferential expression of certain protein isoforms^[90] or at certain lengths can inhibit expression altogether.^[91] 3'-UTR mediated processes (**Figure 4** top, marked green) in eukaryotes are mainly based on interactions with micro RNAs, which are discussed later in this chapter, or interactions of AU-rich elements (ARE). These elements feature long sequences rich in adenine and uracil residues and act as binding hotspots for regulatory RNA and proteins^[92]. The specifics of those interactions are discussed in

Chapter 4.3. In prokaryotes, post-transcriptional control is achieved mainly by riboswitches (see **Chapter 4.2**) or regulatory RNAP recognition motifs in the 5'-UTR^[93] but also by 3'-UTR interactions with RNases or sRNA (**Figure 4** bottom).^[94]

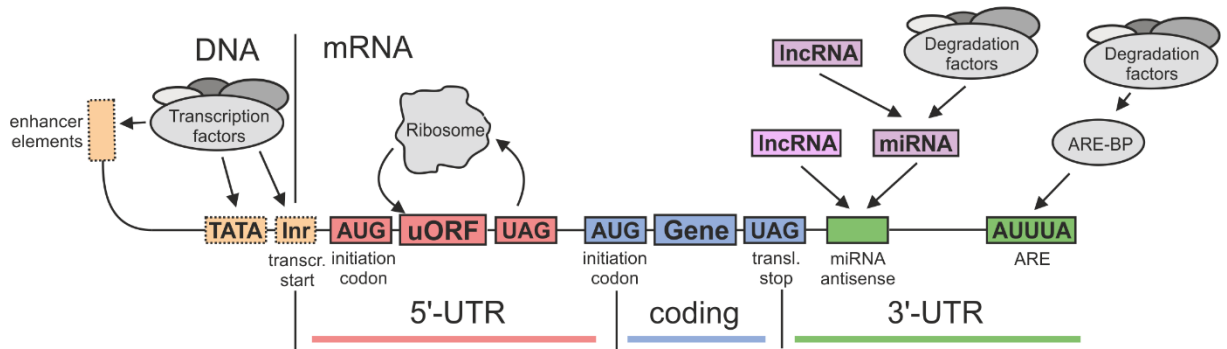
Micro RNA (miRNA) is a naturally existing type of RNA, prevalent in cells with lengths around 20 nt. In animals, primary miRNAs (pri-miRNAs) are transcribed in the nucleus, mainly by RNA polymerase II and processed by the protein Drosha, resulting in hairpins about 70 nt in length known as pre-miRNA or small hairpin RNA (shRNA), which are exported into the cytoplasm. They are subsequently processed into duplex RNA by Dicer, which cleaves the miRNA to a duplex of a specific length, featuring two characteristic 3' overhangs. One strand of the mature miRNA is then loaded onto an AGO family protein.^[95-97] Target sites of miRNA are located in the 3'-UTR of the target mRNA and are characterized by strong complementarity (**Figure 4** top, marked green).^[98] The target mRNA sequence is engaged by the AGO-miRNA complex in a long-lived bound state if sufficient complementarity (usually of nucleotides 2-8) is given.^[99] The bound AGO then recruits proteins initiating deadenylation, decapping and finally degradation^[100], while a different pathway interferes with the eukaryotic initiation factors eIF4A-I and eIF4A-II repressing translation.^[101] In both cases, the gene is silenced.

Transfer RNA (tRNA) is mainly known for its role in decoding the gene sequence and delivering of amino acids to the ribosomal complex in order to allow gene expression. However, regulatory mechanisms in which tRNAs play a key role are also known. The size range of tRNAs is 70 to 90 nucleotides and contains several modified nucleobases, required for proper function. They adopt a well characterized cloverleaf structure including hairpins, carrying the anticodon sequence and facilitating ribosome recognition, as well as the acceptor stem on which amino acids are loaded by acylation.^[102] In eukaryotic cells, regulatory function of tRNA fragments has only recently been observed, mainly as a result of differential modification and subsequent generation of tRNA fragments.^[103] In prokaryotes regulation involving tRNA, specifically unacylated tRNA, can be observed for genes involved in tRNA acylation and amino acid generation.^[104] In this type of regulation, the 5'-UTR of a gene carries a specifier sequence (**Figure 4** bottom, marked red) which recognizes the anticodon stem of the cognate unacylated tRNA. The 3' acceptor end of the tRNA then specifically binds to the 5' strand of an antiterminator helix in the mRNA sequestering the formation of a terminator helix. Without tRNA binding, the terminator helix would lead to termination of transcription.^[105] In another common mechanism termed stringent response, presence of unacylated tRNA acts as the primary effector molecule for protein interactions leading to production of the second messenger ppGpp which then interferes with global gene expression.^[106]

Long non-coding RNA (lncRNA) is a collective term for RNAs which are not coding for proteins, do not fit in any other RNA category (e.g. miRNA, siRNA, snRNA, snoRNA) and are long, typically ranging from 200 to 10000 residues.^[107] Consequently, the regulatory roles of lncRNAs are a vast field, which is only starting to be fully understood. lncRNAs are mostly researched in eukaryotes. Here, for instance they can interfere with transcription, downregulating chromatin interaction^[108], but also having enhancing effects which led to the term enhancer RNA (eRNA).^[109] Additionally,

lncRNAs are known to interact with splicing factors.^[110] Completing the set of possible regulatory interaction processes, modulation of transcriptional efficiency by lncRNAs can also be achieved, passively by reduction of miRNA levels^[111] or actively modulating stability of a target mRNA.^[112]

RNAs involved in eukaryotic gene regulation



RNAs involved in prokaryotic gene regulation

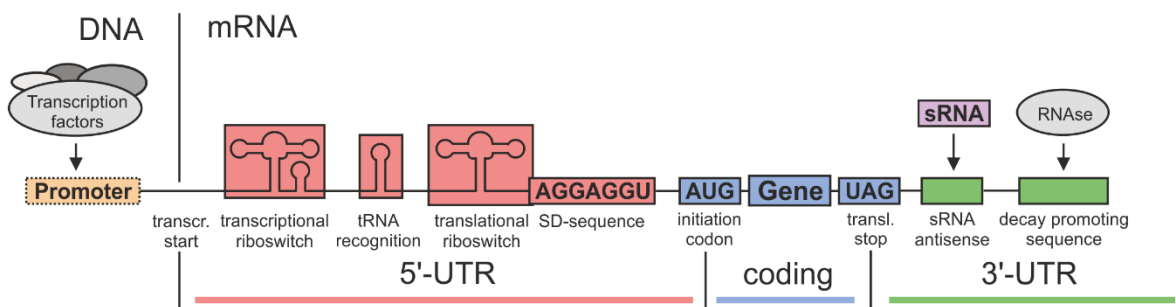


Figure 4 | Common processes of RNA-based gene regulation observed in eukaryotes and prokaryotes. RNA sequences are drawn as boxes, while protein factors are drawn as round shapes. Sequences in dashed boxes are only found on the DNA template as they precede the transcription start site (orange). RNA sequences are grouped according to their location on the mRNA in either the 5'-UTR (red), coding region (blue) or 3'-UTR (green). *Trans-acting* RNAs are colored light purple.

4.2. Small molecule-dependent regulation in prokaryotic 5'-UTRs: Riboswitches

While a multitude of complex mechanisms are involved in RNA-mediated gene regulation which were briefly introduced in the last chapter, in this chapter one particular RNA-based structural element will be discussed, which is of great importance in prokaryotes: the riboswitch. Riboswitches are cis-acting regulatory elements located in the 5'-UTR of mRNAs. Per definition, a riboswitch consists of an RNA based aptamer domain, which binds a small molecule passing a signal to a downstream sequence, often described as expression platform, in order to exert an effect on gene expression by conformational switching. Small molecule binding in riboswitches is achieved without the aid of protein cofactors.^[12] Riboswitch aptamer domains, in order to sense a specific small ligand, show a high phylogenetical conservation, which allows categorization by their structure and the cognate ligand.^[13] The modes of action in riboswitches, especially when the details of structural alteration are appreciated, can be highly diverse, allowing control of transcription or translation, up- or downregulation upon ligand binding and varying degrees of regulatory strength and affinity. Therefore, the main categories for riboswitches and their unique properties are introduced first in this chapter, before examples of the most common riboswitch types are shown and discussed in detail.

4.2.1. Riboswitch types and modes of regulation

Transcriptional riboswitches control the transcription of the nascent mRNA in a ligand-dependent manner. This is achieved by structural alteration of the aptamer domain triggering the formation or disassembly of a rho-independent terminator upon ligand binding. The terminator structure consists of a hairpin, followed by a poly-U stretch. Interaction of this hairpin with the RNA-polymerase and the weakness of A-U bonds destabilizes the transcription bubble and lead to dissociation of the complex, ending transcription.^[14,15] The alternative structure of the riboswitch, which permits termination is often termed antiterminator structure. Depending on the effect, exerted on transcriptional efficiency upon ligand binding, riboswitches are termed on-switches (gene on, when ligand is bound) or off-switches (gene off when ligand is bound). For an on-switch to be able to operate, the unbound state must lead to terminator formation, which means that the terminator state is thermodynamically more stable. In this case the ligand must add additional stability to the antiterminator state, making it the most stable structure, to be able to trigger the on-state. For off switches, the relative stabilities of antiterminator and terminator are inverted. Since the decision for gene-on or gene-off must be made until the RNAP polymerase has reached the terminator, ligand binding must only be effective during a characteristic time-window. Accordingly, it is sufficient for the antiterminator to be metastable during this time frame. Because of these time constraints, transcriptional riboswitches are often described to be under kinetic control.^[16] These systems can become increasingly complicated when all the possibilities of interconversion between stable and metastable states are considered.^[17] Further, transcription speed and therefore kinetics can be impacted by factors such as intrinsic pause sites, uridine-rich sequences which lead to delays in transcription, or Nus proteins stalling the RNAP at those pause-sites.^[18]

Thus, transcriptional riboswitches are able to control mRNA abundance at the transcriptional level and therefore gene expression.

Translational control by riboswitches is generally achieved by interfering with the accessibility of the ribosome binding site, often termed Shine-Dalgarno (SD) sequence. These prokaryotic purine rich sequences are located upstream of the translation start codon and serve as the main ribosome docking site, specifically by being recognized by the 16S ribosomal RNA (rRNA).^[119] They are highly abundant in prokaryotes and were detected in over 77% of protein coding genes.^[120] Translational riboswitches operate by two major mechanisms. The first one is based on the SD sequence being part of the aptamer domain and positioned in such a way that upon ligand addition the SD sequence is involved in intramolecular base-pairing and therefore obscured to ribosome recognition. In the unbound state, the SD sequence would be accessible. This mechanism describes the mode of action in translational off-switches such as the adenosyl cobalamin (AdoCbl) riboswitch from *E. coli*.^[121] The second option involves an anti-SD-sequence, which can interact with the SD-sequence in order to sequester or deploy it. Depending on the consequences of ligand binding on the structural environment of the anti-SD-sequence this mechanism can allow on-switch^[122] or off-switch behavior.^[123,124] While transcriptional riboswitches only act cotranscriptionally, translational riboswitches are able to bind the cognateligand and refold during the whole lifetime of the mRNA. Consequently, gene regulation is considered thermodynamically controlled in translational riboswitches, which implies that its structural states are in thermodynamic equilibrium, which is then affected by ligand interaction. Additionally, external factors such as temperature can also influence this equilibrium and therefore switching efficiency.^[125] An interesting effect is observed for riboswitches with SD-sequences located in a potential rho-independent terminator forming sequence. In this case, riboswitches can act on the transcriptional and the translational level, allowing sequestering of the SD-sequence while simultaneously modulating transcription efficiency by termination.^[126]

4.2.2. Common riboswitches: Structural and functional properties

In order to carry out their biological function, riboswitches must be able to selectively sense their respective cognate ligand. This is achieved by folding of a phylogenetically highly conserved aptamer structure. The high degree of conservation in the aptamer domain allows classification of riboswitches by their aptamer structure and ligand. Consequently, riboswitch classes are mostly named after their cognate ligand, rather than a specific gene they are located upstream of. In this chapter a selection of the most common riboswitch classes is introduced, specifying unique properties of each aptamer, while also discussing reoccurring structural themes and the biological context of each riboswitch. The respective structures are shown in **Figure 5**.

TPP riboswitch

The most common class of riboswitches is sensing thiamine pyrophosphate (TPP), a cellular cofactor in carbohydrate and amino acid metabolism. To date, TPP riboswitches are the only riboswitches also observed in eukaryotes.^[127] The aptamer consists of five helical stems enclosing the ligand from two sides, with the stems P2 / P3 binding the pyrimidine part and P4 / P5

interacting with the pyrophosphate residues. Detailed smFRET studies showed large dynamics in the P2 / P3 and P4 / P5 sides between proposed open and closed conformations, which facilitate association and dissociation of the ligand, making regulation efficiency susceptible to sequential modifications or alterations in the cell environment.^[128,129] Prokaryotic TPP riboswitches typically regulate thiamine related genes, mostly transporters, by the typical transcriptional and translational modulation mechanisms discussed above.^[126] In eukaryotes however, it was shown that the TPP aptamer interacts with a conserved, splicing related element located 530 nt downstream in order to promote gene expression. Comparative gene analysis showed, that this mechanism is spread over several fungal species.^[130]

SAM riboswitch

SAM riboswitch is a collective term for eight different riboswitch subclasses termed SAM-I to SAM-VI, SAM I/IV and SAM/SAH which all bind S-adenosyl methionine (SAM, AdoMet) a common cosubstrate acting as a donor of methyl groups in enzymatic methylation and being involved in many other reactions.^[131] The eight subclasses differ structurally but employ the same function, binding SAM while also discriminating against its demethylation product SAH (except SAM/SAH, *vide infra*). The SAM-I structure is characterized by a four-way junction connecting four helices, which is highly convoluted and features several tertiary interactions, such as a base-triplet and a pseudoknot. SAH discrimination is achieved by a highly conserved AU-base-pair whose carbonyl oxygens interact with the positively charged sulfur atom of SAM.^[132,133] The SAM-II aptamer is built from two stacked helices. Helix P2 interacts with the linker L1/2 to form a short triple helix segment, which forms the binding pocket.^[72] SAM-III^[134], SAM-IV^[135] and SAM-VI^[136] aptamers all feature a three-way junction and are comparable to the structure of SAM-I, SAM-IV showing the strongest similarities.^[137] Additionally, so called SAM I/IV riboswitches show features of both types.^[138] SAM-V aptamers share similarities to SAM-II since they are also built from two stacking helices but feature a longer triple-helical segment.^[139] The last subclass are SAM/SAH riboswitches which are extremely small, consist of two stacked helices, do not discriminate between SAM and SAH and have just recently been structurally characterized by NMR.^[140] Biologically, SAM-riboswitches most commonly regulate genes involved in the synthesis of cysteine and methionine or general sulfur metabolism and exert control on the transcriptional or translational level.^[141–143] An interesting exception is the SAM-I riboswitch from *Xanthomonas campestris*, which is a translational on-switch. But, in addition to SAM recognition by the aptamer domain, the expression platform of the riboswitch recognized uncharged F-Met-tRNA. Even if no SAM is bound, the tRNA recognition event leads to liberation of the Shine-Dalgarno sequence as the anti-SD-sequence is sequestered and thus upregulates gene expression.^[144]

AdoCbl (Coenzyme B12) riboswitch

Adenosylcobalamin (AdoCbl or Coenzyme B12) is a cellular cofactor involved in catalysis of radical-based carbon rearrangement reactions. Its complex structure is sensed by a class of riboswitches with remarkable specificity.^[145] The highly conserved core of their aptamer structure consists of a four-way junction joining the helices P3, P4, P5 and P6. The bridging sequences

between helices P3 and P4 as well as between P3 and P6, called J3/4 and J3/6 interact with AdoCbl from the bottom and from the side, respectively. The AdoCbl riboswitch features two highly conserved adenosine residues interacting with the adenosyl group which are responsible for discrimination against the photolysis product aquocobalamin (AqCbl).^[121] AdoCbl riboswitches mostly regulate genes involved in the B12 metabolism in biological feedback loops by direct modulation of transcriptional or translational efficiency.^[146] A well-studied example is the *btuB* riboswitch from *E. coli*, which controls translation by sequestering the Shine-Dalgarno sequence upon AdoCbl binding.^[76,147] A more complex and fascinating example of gene regulation by AdoCbl riboswitches is the *EutX* (*eut* = ethanolamine utilization) riboswitch from *Enterococcus faecalis*. *EutX* is a trans-acting RNA which includes the AdoCbl riboswitch in control of its transcriptional efficiency and a double hairpin sequence, which can interact with *Eut* gene termination sites. By binding and sequestering these termination sites, *EutX* and therefore its AdoCbl riboswitch are in control of transcriptional efficiency of at least 19 different genes on polycistronic mRNAs in ethanolamine metabolism.^[148]

Lysine riboswitch

Lysine is one of the two amino acids which are known to be bound by riboswitches. It is an essential amino acid and in bacteria is crucial for cell wall biosynthesis or virulence.^[149,150] The structure of the lysin aptamer is composed of five helices forming a five-way junction with the lysine binding pocket located in its core between helix P1 and the linker sequence of helices P2 and P3 (J2/3). Ligand recognition is achieved by several directly interacting residues, which form a small binding pocket for the ϵ -amino group majorly contributing to selectivity against amino acids with longer and shorter side chains.^[151] Analogues with modifications in the γ -position however can bind, as the binding pocket is only loosely packed at this position.^[152] Recognition of the α -carboxy group is mediated by potassium ions, which could be shown to have a beneficial effect on lysine recognition.^[153] Other than for example in humans, lysine is biosynthesized in bacteria, mostly over the diaminopimelate (DAP) pathway.^[154,155] Consequently, the most common genes regulated by lysine riboswitches are involved in this pathway and general lysine biosynthesis.^[156]

Glycine riboswitch

The second amino acid, known to be involved in gene regulation by riboswitches is glycine. Like lysine, glycine is an essential amino acid in bacteria. It is biosynthesized from serine.^[157] Being the simplest amino acid, interaction sites in its molecular structure are limited and only a small energetic difference between the bound structure and the unbound structure of glycine riboswitches can be expected. It could be shown, that this energetic difference amounts to only 2-3 kcal/mol for a single aptamer domain.^[158] As opposed to other riboswitches discussed in this chapter, the glycine riboswitch aptamer structure was described as a tandem aptamer in which two aptamer domains bind glycine and cooperatively regulate transcription^[159], though singleton riboswitches are likewise abundant in prokaryotic species^[160]. However, singleton glycine riboswitches feature single hairpins upstream or downstream of the full aptamer, which were

termed ghost-aptamers and show stabilizing effects on formation of the aptamer domain or act as part of the expression platform.^[158] The structure of each individual aptamer consists of three helices of which P2 and P3 are coaxially stacked, with the glycine binding pocket formed as an A-rich bulge region of helix P3. Glycine interacts with a G-residue, a U-residue and two metal ions, which are coordinated by several backbone atoms. Interaction of both aptamer domains is facilitated by inverted repeat sequences of helix P1 of the upstream aptamer and helix P3 of the downstream aptamer. A specific residue interfacing between the binding pocket and the tandem interaction site allows crosstalk between the two aptamers and was shown to be responsible for cooperativity in binding.^[161] Glycine riboswitches are generally observed in the context of genes involved in the glycine cleavage system or transport proteins. Interestingly, it could be shown that the position of a ghost-aptamer (upstream or downstream of the aptamer) is related to the gene context of the riboswitch.^[162]

FMN Riboswitch

The metabolite flavine mononucleotide (FMN) is biologically synthesized from riboflavin (vitamin B₂) and acts as a cofactor in hundreds of proteins involved in the redox metabolism of bacterial cells.^[163] It is recognized by an aptamer domain composed of six helical stems which are arranged in a complex shape around the binding pocket forming two large domains composed of helices P2/P6 and P3/P5. The two helices of each domain are held together by tertiary interactions. Interestingly, though both domains appear largely symmetric, they bind the ligand unsymmetrically, with the phosphate group pointing towards the P3/P5 domain and the isoalloxazine ring to the P2/P6 domain.^[164] Studies by Batey et al. showed, that FMN binding is achieved by conformational selection, where the riboswitch is conformationally flexible in order to be able to adopt a bound-like structure, which is then recognized by FMN with only minimal structural alteration.^[165] Already in 2005, kinetic control in transcriptional modulation was discovered on an FMN riboswitch from *B. subtilis*. It could be observed, that the level of genetic control, which was expected from the high affinity aptamer was not reached at a given concentration, since thermodynamic equilibrium was not reached while transcription proceeded.^[118] This type of kinetic control is nowadays the estimated standard model for transcriptional regulation by riboswitches (see **Chapter 4.2.1**). FMN riboswitches mostly precede genes involved in the riboflavin, FMN and flavin adenine dinucleotide (FAD) biocycles, where they control transcription, translation or even both at the same time.^[166]

yybP-ykoY riboswitch

The yybP-ykoY riboswitch is a special case among the riboswitches reviewed in this chapter. Firstly, it is not, like most riboswitch classes named after its cognate ligand, but after the genes it was first discovered regulating, the membrane protein coding genes yybP and ykoY from *Bacillus subtilis*.^[167] Secondly its cognate ligand is not a classic cell metabolite, but doubly charged manganese ions (Mn²⁺). Given the unusualness of the ligand, it is not surprising, that only recently its identity could be confirmed^[168] and the structure of the aptamer domain obtained.^[169] Despite having to bind only an ion, the aptamer structure is complex, being comprised of six individual

helices which form two arms by coaxial stacking. The binding pocket features two metal binding sites, in both of which metals are coordinated octahedrally in order to be able to accommodate Mn^{2+} but also Mg^{2+} ions. Selectivity is achieved by one of the two metal binding sites featuring softer interacting atoms (nitrogen over oxygen) and a currently not fully understood preference of higher ion radius. The other binding site is non-selective and is expected to bind preferably Mg^{2+} at cellular ion concentrations.^[169] The known gene context of the *yypP-ykoY* riboswitch includes mostly genes involved in the manganese metabolism such as efflux pumps in *e. coli*^[170] or *xanthomonas oryzae*.^[171]

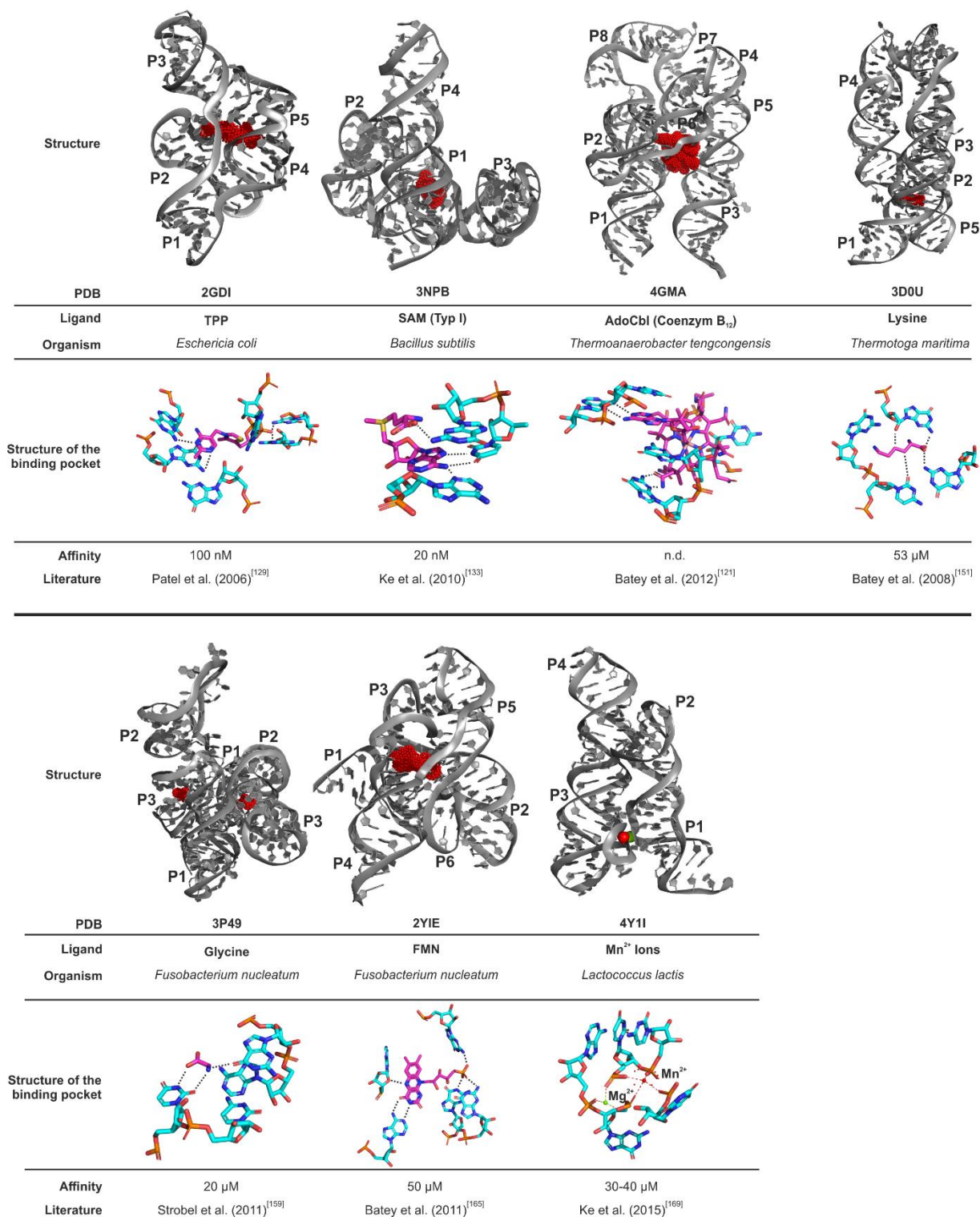


Figure 5 | Aptamer structures of seven common riboswitch classes. The full three-dimensional crystal structure of the aptamer is shown on top, while the interactions in the binding pocket of each aptamer are delineated in detail below.

Purine and purine derivative sensing riboswitches

Though most riboswitches share structural similarities, like the binding pocket being located in a structurally complex area, mostly a junction, such similarities are even more striking throughout the riboswitch classes sensing purines or purine derivatives. Additionally, these riboswitches share strong similarities regarding their gene context and are consequently brought up in a single subchapter here (see **Figure 6** for structures), though separate classes are defined in literature. The first purine riboswitch was suggested by Christiansen et al. already in 1997.^[172] They observed a 160-fold gene repression in the *pubX* gene, which codes for a xanthine transport protein, in presence of hypoxanthine or guanine. The hypothesis was proven later in the Breaker lab, including determination of the secondary structure, scanning of the ligand space and classification of the gene context of the guanine sensing riboswitch.^[173] Shortly afterwards, riboswitches binding adenine were discovered as well, which share the same structure.^[174] It consists of three helical stems, P1 P2 and P3 which form a three-way junction encompassing the ligand binding pocket. Stems P1 and P2 are held together at the top of the structure by loop-loop tertiary interactions (marked orange in **Figure 6**), which are phylogenetically conserved.^[175,176] Interestingly, in order to facilitate discrimination between guanine and adenine by the respective classes, the ligand is directly recognized by formation of a Watson-Crick base-pair and a single uridine/cytosine mutation enables switching of the selectivity.^[177] Two more residues form direct hydrogen bonds with the ligand and stabilize the interaction. Another rare purine binding riboswitch class showed high selectivity for 2'-deoxyguanosine and was first observed in the organism *Mesoplasma florum*.^[178] Its structure is highly similar to the guanine and adenine sensing riboswitches, however a slight shift of the cytosine residue which recognizes the sugar face of the ligand allows fitting the 2'-deoxyribosyl group into the binding pocket.^[179]

The guanine derivative preQ₁ is sensed by structurally fundamentally different riboswitches. This ligand is a precursor of the modified guanosine derivative queuosine, which is part of tRNA anticodon loops in prokaryotes and eukaryotes.^[180] The aptamer domain shares the layout of the binding pocket with purine riboswitches in regard of the type of coordination of the ligand, which is recognized by Watson-Crick base-pairing to a cytosine residue. However, the aptamer domain lacks the two arms and the kissing-loop interaction between them. It is built from two coaxially stacked helices with the binding pocket at the stacking position. The 3' single stranded end is folded into the minor groove of the main helices and provides the residue contacting the sugar face of the ligand.^[181]

Two distinct classes of riboswitches sense secondary messengers composed of two purine residues and two cyclic phosphates, such as cyclic diguanosine monophosphate (c-di-GMP), cyclic diadenosine monophosphate (c-di-AMP) and cyclic guanosine monophosphate-adenosine monophosphate (c-GAMP). Second messengers are molecules known to transduce environmental signals in cells in order to trigger cellular response.^[182] Consequently, they are involved in response to environmental change for example by lifestyle changes, motility or biofilm formation.^[183–185] According to their genetic context, one structural motif of the aptamer domain sensing those second messengers is called Genes for the Environment for Membranes and for

Motility (GEMM) motif.^[186] The aptamers are divided in two subcategories GEMM-I and GEMM-II. GEMM-I riboswitches are, regarding their general structure, highly comparable to purine riboswitches, with three helices, joined in a three-way junction. Interaction of P2 and P3, which, in purine riboswitches, is realized through loop-loop interactions, is observed in GEMM-I aptamers as an interaction of the highly conserved GNRA Loop of P2 with a loop recognition motif in the bulge of helix P3. The ligand is sensed with, compared to other riboswitches, very high affinity as the K_D was determined to be around 10 nM. Two G residues of c-di-GMP, G_α and G_β , in c-di-GMP are recognized by a G residue and a C residue, G_α by WC-Hoogsteen interaction and G_β in a canonical G-C *cis*-WW base-pair. Additionally, the ligand is sandwiched by two WC base-pairs.^[187] In GEMM-I c-GAMP sensing riboswitches the G residue which interacts with G_α is exchanged for an A in order to recognize A_α .^[188]

Moving on to the second class, GEMM-II motifs show a general organization of the binding pocket, comparable to GEMM-I. However, no canonical base pairing is observed but G_α and G_β are recognized by non-WC interactions. The sandwiching interactions with base pairs and the single residue located between both ligand residues are highly similar. The binding pocket is located in a three-way junction, but instead of interacting by tertiary interactions, the two upper helices P2 and P4 are directly connected through another helix P3.^[189]

Interestingly, riboswitches sensing c-di-AMP are fundamentally different compared to GEMM riboswitches. They adopt a highly symmetric structure composed of two domains including three helical units each. The three helices of each domain form a T-shaped three-way junction with two of them coaxially stacked. At each of the two interfaces of the two domains, a ligand is bound and coordinated by residues of both domains.^[190,191] Such symmetry is rare in functional RNAs and the most notable other example is the FMN riboswitch (*vide supra*).^[191] The gene context of c-di-AMP sensing riboswitches is, comparable to GEMM riboswitches, mostly lifestyle related with genes involved in cell-wall synthesis or sporulation.^[192]

This subchapter is closed with the discussion of two riboswitches regulating genes at the connection of purine- to 1-carbon metabolism, a metabolic pathway essential for cell survival often targeted by anti-cancer or antibiotic agents.^[193–195] In prokaryotic purine synthesis, formyl groups are donated by the metabolite 10f-THF (10-formyltetrahydrofolate) to act as the C2 or C8 atom in the purine heterocycle. To add a C2 atom, a formyl group is transferred from 10f-THF to ZMP (5-aminoimidazole-4-carboxamide ribonucleotide, AICAR) resulting in the formation of the direct inosine precursor 5-formamidoimidazole-4-carboxamide ribotide (FAICAR). Consequently, in cases of 10f-THF deficiency ZMP would accumulate in the cell. Therefore, already 1982, ZMP and its three times phosphorylated derivative ZTP were defined as signal molecules for 10f-THF deficiency, so called alarmones.^[196] It took until 2015 though, that a receptor for ZMP was found which is able to transduce the signal of increased ZMP concentration into cellular response, the ZMP-sensing riboswitch featuring the *pfl* structural motif.^[197] The *pfl* motif was named after the controlled gene, which codes for pyruvate-formate lyase, a protein responsible for the biosynthesis of formate, which is used in 10f-THF regeneration.^[198] The structural motif consists, in its minimal form, of two helical stems (here P1 and P3), which interact in a pseudoknot and are connected by a linker of unconserved length. The well-studied example from *thermosinus*

carboxydivorans features an additional small stem in the long P1 loop region termed P2, which is phylogenetically not conserved and an 8 nt A-rich linker.^[199] Helix P3 partially consists of residues which are part of an intrinsic terminator when ZMP is not present, making the riboswitch a transcriptional on-switch. Upon ZMP interaction the pseudoknot structure is stabilized hindering formation of the intrinsic terminator, leading to transcription and gene expression. Ligand recognition is achieved by a binding pocket located directly below the pseudoknot, the interaction site of P1 and P3, contrary to purine and di-purine sensing riboswitches (*vide supra*) in which a three-way junction region detects the ligand. The ring N3 and the amide group of ZMP interact directly with the WC-face of a conserved U residue directly beneath the pseudoknot, with the ZMP heterocycle sandwiched between the lowest pseudoknot G-C base-pair and a conserved G residue of the P3 loop. Recognition of the carboxamide group is facilitated by a magnesium ion, which is coordinated to two phosphate oxygens. A gap in the backside of the binding pocket allows recognition of ZTP, which has a longer phosphate chain.^[200] In addition to the *pfl* gene, ZMP sensing riboswitches are also associated with a number of other genes also involved in purine biosynthesis, such as *PurH* coding for an AICAR formyltransferase or *Fhs* coding for a formate tetrahydrofolate ligase.^[197] In research article I (see **Chapter 9**), the NMR-based characterization of the ZMP-sensing riboswitch from *t. carboxydivorans* is discussed.

As explained earlier, detection and gene control upon 10f-THF deficiency is achieved passively by riboswitch mediated sensing of the alarmone ZMP. However, genes involved in production of the 10f-THF precursor THF are not controlled by *pfl* riboswitches. Instead, an additional riboswitch class, which directly senses THF, is utilized. It controls genes involved in folate biosynthesis or folate transport and it is speculated, that THF riboswitches could colocalize with *pfl* riboswitches in order to achieve recognition of two different ligands.^[201] The aptamer structure is comparable to GEMM-I riboswitches with two helical arms, here achieved through coaxial stacking of two helices each, a three-way junction and tertiary interactions between both arms. In opposition to the GEMM-I motif the main tertiary interaction is a pseudoknot in the lower part of the riboswitch coaxially stacking with the P1 helix. A major difference to most riboswitches is the recognition of two ligands at two distinct locations in the three-dimensional structure one located near the pseudoknot and one in a bulge region of P3. At both binding sites, the pterin domain is the majorly recognized motif of the ligand and is coordinated by three or two residues always involving interaction of the amino group. Mutational studies showed, that the pseudoknot binding site is more important for regulatory activity than the upper binding site.^[202]

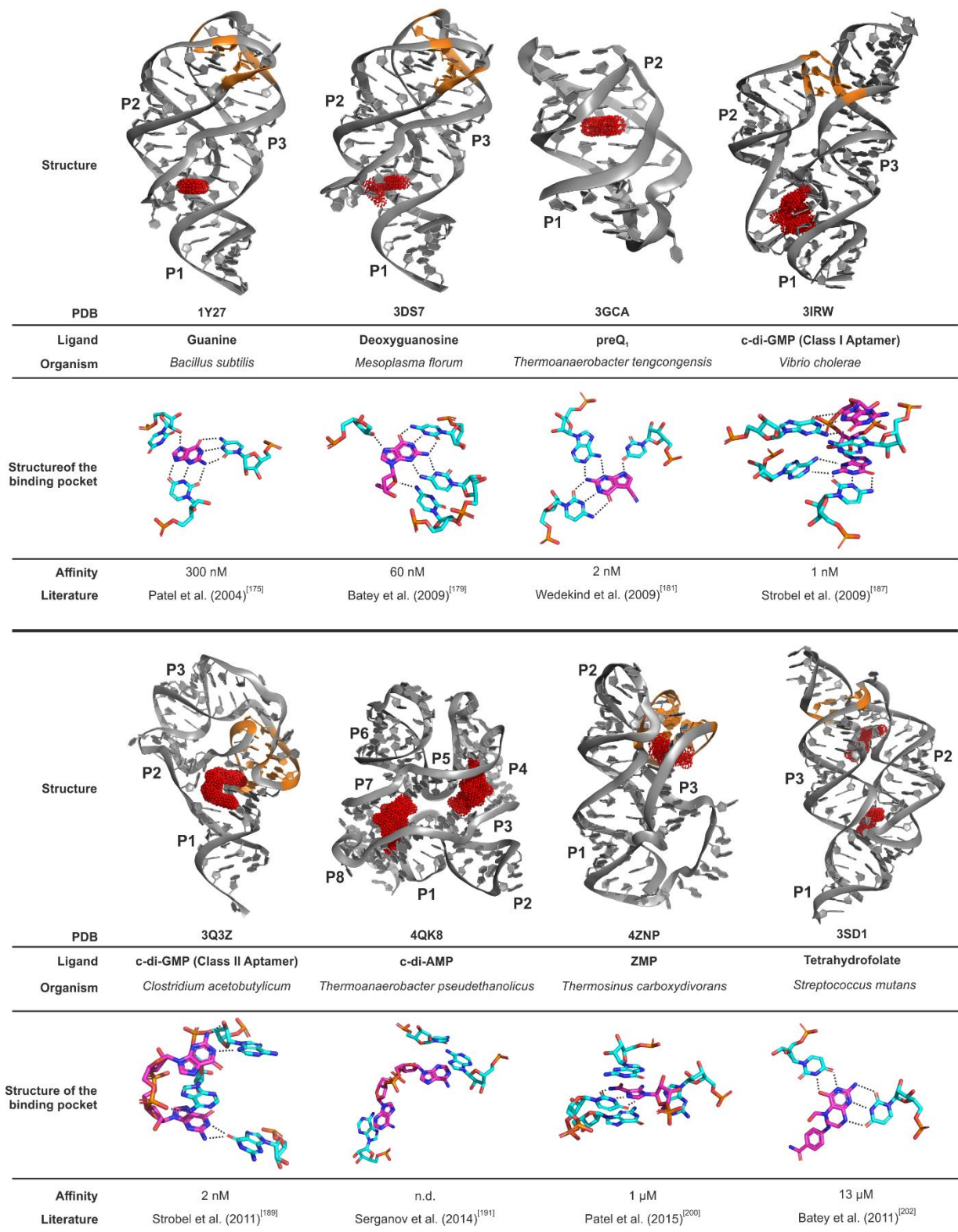


Figure 6 | Aptamer structures of eight common riboswitch classes sensing purines or purine derivatives. The full three-dimensional crystal structure of the aptamer is shown on top, while the interactions in the binding pocket of each aptamer are delineated in detail below. The important tertiary interaction, connecting the stems in most classes are marked orange.

4.3. Gene regulation by AU-rich sequences and constitutive decay elements in the 3'-UTR of eukaryotic mRNA

Eukaryotic gene regulation is a vast topic, as can be observed from the examples introduced in the beginning of this chapter. Here, a specific important mechanic of eukaryotic gene regulation is discussed in further detail. AU-rich sequences (AREs) as regulatory elements were first discovered in 1986 in a human lymphokine gene where it could be shown, that mRNA harboring a 51 nt AU-rich sequence in its 3'-UTR was rapidly degraded.^[203] Today it is estimated, that 5-8% of human genes code for mRNAs including AREs. The affected genes are involved in tightly controlled and often disease relevant processes such as cell proliferation, differentiation or stimuli response (immune-, inflammatory response, environmental factors, cancer) and AREs ensure a low expression level outside of these rare circumstances.^[204,205] Consequently, the understanding of the processes underlying ARE-based regulation can be crucial for research in pathology and drug development.^[206–208]

AREs were initially defined as 50 to 150 nt long unstructured elements featuring an enrichment of uridines and the consensus sequence AUUUA, with the shortest single sequence yielding significant effect being the nonamer UUAUUUAYY (Y: A/U).^[209] Repeats of AU-rich motifs like AUUUA (Class I ARE) or the described nonamer (Class II ARE) lead to more efficient regulation.^[210,211] Additionally, AREs harboring none of these sequences but unspecific AU-rich regions are nowadays known, which still show efficient regulatory effects (Class III ARE).^[212] In opposition to bacterial riboswitches, the eukaryotic AREs are unable to solitarily regulate gene expression but are targeted by a variety of ARE-binding proteins (ARE-BP), which recruit the cellular machinery employed in the process of regulation. ARE-BPs interact by different RNA recognition motifs (RRM) and their action can have stabilizing or destabilizing/degrading effects on the mRNA.^[92] The first identified ARE-BP was AUF1 (ARE-binding protein 1, also heterogeneous nuclear ribonucleoprotein D, hnRNP D).^[213] It interacts with the RNA with two tandem RRM repeats, so called K-homology (KH) domains^[214], and recruits the AUF1- and signal transduction-regulated complex (ASTRC) which is required for mRNA decay^[215], but can also facilitate stabilization.^[216] The particular circumstances of these differential effects and the molecular mechanism are unknown.

Contrarily, the molecular mechanism of a different ARE-BP Tristetraprolin (TTP) is better understood. It binds AREs through interaction of a tandem zinc finger domain^[217] and recruits the CCR4-NOT (carbon catabolite repression - negative on TATA-less) complex, which promotes degradation by harboring (among other enzymes) a poly-A specific ribonuclease (PARN).^[218] Additionally TTP interacts with a decapping complex^[219] and a complex which competes with eukaryotic initiation factors for the 5'-end cap.^[220] All those differential factors contribute to mRNA degradation.

RNA degradation is also facilitated by ARE recognition of K-homology splicing regulatory protein (KSRP), which, like AUF1, uses KH-domains as RRM.^[221] Comparable to TTP though, it recruits a PARN, and additionally promotes the exosome, which degrades mRNA.^[222,223]

Degradation of mRNA facilitated by ARE-BPs as mentioned above is competed by proteins of the

human antigen family (HuR, HuB, HuC and HuD, also: Elav-like proteins), of which only HuR is ubiquitously expressed and the others are only found in neurons.^[224] All of the Hu proteins include 3 RRMs, of which the first two are crucial for ARE recognition. The third RRM was predicted to bind the poly-A tail^[225], but the affinity was confirmed to be lower than the affinity to AREs, rendering the role of RRM3 a matter of scientific debate.^[226] HuR stabilizes mRNAs through competition with TTP and KSRP.^[227] Additionally, it promotes cap-dependent initiation through recruitment of eukaryotic initiation factors and poly-A interaction.^[228,229]

Interestingly an overlap of ARE-mediated and miRNA-mediated gene control was discovered, as the presence of Dicer (see **Chapter 4.1: miRNA**) was shown to be required for gene control based on an ARE in HeLa cells. Action of human miRNA 16, which contains a sequence complementary to the investigated ARE additionally required the presence of TTP confirming miRNA-ARE cooperativity.^[230]

Initially it was assumed that, in order to function as an ARE, to be bound by ARE-BPs and to facilitate mRNA degradation or stabilization, AU-rich sequences must be unstructured. This still holds true for most AREs, however a 2005 study sparked discussions by showing, that the ARE from tumor necrosis factor α (TNF α) forms a hairpin structure *in vivo* and thus modulates the affinity of ARE-BPs and therefore regulation efficiency.^[231] Interestingly, in a different study in 2003 a new regulatory element was discovered on the same mRNA, termed constitutive decay element (CDE), which was shown, to promote mRNA decay, but its structure or interacting proteins were unknown.^[232] Ten years later, it was found that the CDE is a stem-loop which interacts with the protein Roquin.^[233] Genes regulated by CDE-Roquin interaction are commonly involved in inflammation, which can also be regulated cooperatively with a comparable stem-loop system involving the protein Regnase-1.^[234] Roquin interacts with CDEs by its unique Roq domain and recruits the CCR4-NOT complex in order to promote mRNA decay.^[233] The consensus sequence of CDEs was initially defined as a 5 – 8 bp stem with three highly conserved upper base-pairs and a Y-R-Y tri-loop (Y: U/C, R: A/G), however recent studies widened the spectrum of possible Roquin binders to a 6 – 8 bp stem with no stem conservation other than base-pairing and a Y-R-N tri-loop (N: any). Interestingly, this allows a CDE to also be a structured ARE.^[13] Research article II features in depth structural and functional characterization of two of such CDEs, one AU-rich and one even consisting solely of A and U residues (see Chapter 10).

4.4. G-quadruplexes in gene regulation

A special case among the many regulatory active RNAs showcased in this chapter are G-quadruplexes (G4s), as they form fundamentally different structures composed of G-tetrads. These highly stable structures form in G-rich regions of nucleic acids under influence of monovalent cations (see **Chapter 3.2.5**). Though the whole chapter comprises mostly the discussion of regulatory RNA, in this subchapter some examples for regulatory G4-based DNA are presented, in order to show the broad scope of regulation exerted by G4s, especially since both RNA-based and DNA-based regulation sometimes overlap in this field (*vide infra*).

While the formation of G4s has long been proven for natural and unnatural G-rich sequences *in vitro*, *in vivo* evidence for widespread formation of G4s has long been scarce. The topic emerged

when *in silico* studies showed that putatively G4 forming consensus motifs, such as G₃₋₅N₁₋₇G₃₋₅N₁₋₇G₃₋₅N₁₋₇G₃₋₅, were present in over 300.000 sequences of the human genome^[235,236], enriched in regulatory relevant regions, like promoters.^[237] Later studies expanded these endeavors, in order to recognize so-called imperfect G4s, widening the scope of *in silico* G4 recognition. Biochemical studies employing reporter assays made the *in vitro* formation of G4s more evident, utilizing the stalling potential of the very stable G4 structure on reverse transcriptases. With these methods, over 700.000 potentially G4 forming sites were detected in the human genome.^[238] When combined with computational mapping, results of biochemical reporter assays could relate thousands of non-B-DNA sites with regulatory relevant locations.^[239] Some G4 forming sequences in promotor regions of oncogenes are particularly well studied such as G4 sequences from *c-Myc*^[240], *KRAS*^[241] and *c-Kit*^[242]. G4s are involved in several distinct cellular mechanisms outlined in the following passages.

In general, formation of G4 structures is only possible if the natural base-pairing of a nucleic acid is at least transiently interrupted or the nucleic acid is principally single stranded. Therefore, an important mechanism where G4s can interfere is transcription, where the B-DNA strand is disentangled and a single stranded and, if at all, only partially structured mRNA strand is released. G-rich strands that potentially form G4s can impede transcription in distinct ways: Direct blocking of RNAP by G4 formation in the template strand, formation of a G4 between the non-template strand and nascent RNA and G4 formation on the template strand facilitating interaction between the RNA and the non-template strand (a so called R-loop). The same principles apply to DNA replication as well.^[243] Especially R-loops are prone to G4 formations, as only two neighboring G-tracts of two or more Gs would be required on the DNA to allow formation of a RNA:DNA hybrid G4. Such sequences were found in >97% of human genes, enriched downstream of transcription start sites^[244] and their regulatory effect has been confirmed *in vitro* and *in vivo*.^[245]

Additional hotspots for G4 formation are telomeres, which contain the repeating G-rich sequence TTAGGG and play a role in chromosome stability, preventing unwanted protein interactions (e.g. exonucleases) and destabilization.^[246,247] The antisense transcripts of telomeres, called telomeric repeat-containing RNA (TERRA) were suggested to play an additional role in regulating telomere length and telomere histone modification, scaffolding proteins involved in these processes.^[248] Increased telomeric length and action of telomerase, the protein synthesizing telomeres, is a hallmark of most cancers. Consequently, G4 stabilizing compounds, which inhibit telomerase activity, are developed as chemotherapeutics.^[249,250]

RNA G4s are known to be even more stable than their DNA counterparts, despite their ability to form predominantly a single G4 topology.^[251] Since RNAs are single stranded for the most part, their potential to form G4s is increased compared to DNA. Most studies employ cell-based assays and show the actual formation of G4s (if at all) *in vitro* and while formation of RNA G4s can be easily confirmed *in vitro* through various biophysical methods, the evidence of their *in vivo* formation is much weaker. First observations of *in vivo* formation of RNA G4s were established with G4 specific antibodies while the study also explored the potential of interaction of small-

molecule compounds with RNA G4s.^[252] More recently, next generation sequencing allowed more detailed studies, since G-quadruplex formation stalls reverse transcriptases in RNA sequencing (RNA-seq) approaches.^[253] Coupling *in silico* prediction with chemical modification and next generation sequencing, essentially allowed the observation of G4 forming sites over the whole human transcriptome.^[254–256] However, these studies were carried out *in vitro* and similar *in vivo* data showed, that despite numerous indications, G4s are largely unfolded in eukaryotic cells, possibly due to action of a cellular machinery, unwinding those structures.^[257] Unsurprisingly, this study sparked a scientific debate about the existence of RNA G4s *in vivo*, including considerations like transient population of G4s, too high sensitivity thresholds of RNA-seq and the selection of only a subset of potentially G4 forming sequences.^[258] While the cellular machinery, supposed to unwind RNA G4s *in vivo* is still largely undescribed, several proteins which generally bind G4s are known, which can stabilize, or destabilize G4s. Still, in many cases these proteins could just sense G-rich regions and not G4s and a lot of these interactions were only observed *in vitro*.^[258]

Many RNA G4s or G-rich regions with potential of G4 formation were shown to have distinct regulatory roles, mostly in modulation of translational efficiency either in the 5'- or 3'-UTR of mRNAs. When G-rich sequences are located in the 5'-UTR, translational efficiency can be directly regulated. The first studies on this field explored disease relevant genes for the effects of G-rich sequences on expression levels. G-rich sequences were found to amount for 1.5-fold decrease of translational efficiency in the mRNA coding for the well-studied fragile x mental retardation protein (FMRP)^[259] and a 4-fold decrease in the oncogene NRAS.^[260] Interestingly, FMRP could also be shown to recognize a G4 motif in its own mRNA in order to control alternative splicing.^[261] Such systems have been shown to be a promising target for drug-development.^[262] A CGG-repeat containing motif on the mRNA influences transcriptional levels as well.^[263]

In more detailed studies, RNA G4s were shown to interfere with the action of eukaryotic initiation factors (eIFs)^[264], in dependence of their relative position to the cap structure and AUG start codon.^[265] Interestingly, this effect also affects the efficacy of the anti-cancer agent silvestrol, which blocks eIF4A. Ribosome profiling revealed, that G-rich sequences are enriched in mRNAs downregulated upon silvestrol application in mice.^[266] G-rich sequences from this study were biophysically investigated in research article III (see **Chapter 11**). With these studies observing downregulation exerted by possible G4 formation, the opposing effect is also reported. In the 5'-UTR of mRNA encoding the transcription factor NRF2, G4 formation has a beneficiary effect on translational efficiency.^[267] RNA G4 in the 3'-UTR are shown to modulate translation efficiency as well, but the underlying mechanisms and cofactors could not be determined unambiguously.^[268] It is suggested, that G4s interfere with ribosome relocation on the same mRNA to a different open reading frame.^[269,270]

Different studies also show rather indirect G4 mediated modulation of translational efficiency, for example affecting mRNA maturation or localization. A short G-rich sequence in the 3'-UTR of the cancer relevant TP53 mRNA modulates maturation efficiency through interaction with an auxiliary protein, which binds the polyadenylation factor CstF (cleavage stimulation factor), thus sequestering its recognition site.^[271] Additionally, G4s seem to play an important role in mRNA

localization, as modification of a G4 containing 3'-UTR leads to localization failure in postsynaptic proteins in neurons.^[272] Apart from UTRs, recently evidence emerged, that G4s might also be involved in regulation by lncRNAs^[273] as well as miRNAs.^[274,275]

5. Non-natural RNA: Practices and applications

As shown above in a biological context, nucleic acids are structurally and functionally diverse molecules, which fit in a multitude of biological roles. However, there are also extensive efforts to utilize their ambivalent properties by designing or developing new and artificial nucleic acids to carry out specific tasks in i.E. nano or medicinal applications. In this chapter an overview of applications and types of artificial RNA is given.

5.1. Artificial miRNA and RNA based drugs

As mentioned earlier, miRNAs modulate protein expression on the translational level by binding to mRNA in order to regulate translation efficiency. It is estimated that the expression of 60% of human protein-coding genes could be influenced by action of miRNAs.^[276] Additionally, altered expression of miRNAs is associated with diseases such as hepatitis C^[277] and cancer^[278,279]. These properties, along with the relatively small size of miRNAs, inspired the development of RNA based drugs, which are sometimes actual miRNAs or so called small interfering RNAs (siRNAs). By bringing siRNA into the cell, mRNA expression levels can be altered in a process described as RNA interference (RNAi), which ultimately leads to mRNA degradation. Both types can be summarized under the category: antisense drugs.

While nucleic acids, when used as drugs, are in principle well tolerated in cells, they are prone to nuclease degradation, have low cell wall permeability and can potentially trigger an immune response.^[4] To overcome the hurdle of low cell wall permeability, nucleic acid therapeutics are often encapsulated or surface-attached to nanoparticles that feature cholesterol moieties, which are efficient trans-membrane transporters.^[280] The other drawbacks can be efficiently tackled by mimicking the cells natural miRNAs. It could be shown, that alteration of nucleotides within a natural miRNA sequence has no impact on maturation efficiency.^[281] Consequently, artificial miRNA mimics can be brought into the cell as shRNA and will then undergo maturation by Dicer and AGO to be subsequently exported into the cytoplasm where they can bind the target mRNA. This elegant pathway is exploited in the development of new RNA based drugs, targeting the gene huntingtin, whose mutation triggers Huntington's disease.^[282] Dicer products are characterized by 3'-overhangs on each stem, which can be mimicked, to allow RNA drug application at this point in the pathway.^[283] Another way to prevent rapid nuclease degradation is the introduction of chemical modifications into nucleic acid based drugs. One of the most common and also to oldest modification is the so-called PS modification in which a non-bridging oxygen atom is replaced by sulfur^[284], however this modification can lead to toxicity introduced by increased affinity to cell-wall proteins^[285]. An alternative is the boranophosphate group in which the oxygen is replaced with a borane.^[286] Additionally, the common sugar modifications 2'-OMe, 2'-O-methoxyethyl and 2'-flouro can be utilized to stabilize the RNA in the preferred C3'-endo sugar pucker, which leads to an overall stabilization and increased binding efficiency to the target.^[287] A comparable effect is achieved by the implementation of locked nucleic acids LNA, in

which the 2' and 4' positions are bridged by a methylene linker forcing a C3'-endo sugar conformation.^[288]

Nucleotide based antisense drugs can be used to battle a multitude of diseases, most of which are genetic diseases. For example, drugs against spinal muscular atrophy (SMA)^[289], hypercholesterolemia (HoFH)^[290] or Duchenne muscular atrophy (DMD)^[291] have already been commercialized, with multiple more in clinical trials^[292]. Additionally, it could be shown, that several other diseases could potentially be combatted utilizing RNA based therapeutics. Shim et al. developed an artificial miRNA mimic, based on the natural miRNA-155 to inhibit metastasis and invasion of breast cancer cells by increasing expression of the gene CXCR4^[293]. With a DNA based vaccine vector, which delivers an artificial miRNA into the cell, Purcell et al. could show promising effects on the development of T-cell based immune response to HIV-1 envelope antigens^[294]. All these studies show the immense success of RNA-based therapeutics, which only recently emerged.

5.2. Artificial functional RNA: Aptamers and Ribozymes

Thought to be reminiscent of an ancient RNA-world, functional RNA molecules are highly prevalent in nature, with riboswitches mainly in prokaryotes and ribozymes spanning all kingdoms of life (see **Chapter 4.2**). Inspired by nature's ability to create functional molecules from four simple building blocks, scientists managed to create functional RNA sequences themselves. Two main applications of artificial functional RNA molecules are biotechnological and medicinal utilizations. Compared to alternatives such as small molecules or proteins, pure RNA implementations bear advantages such as being directly transcribable in an organism or modular composition in which parts of the functional molecules can be exchanged.^[295] For example cleavage activity of a hammerhead ribozyme could be made ligand dependent, by combining it with a riboswitch aptamer domain.^[296] While the mode of action in self-cleaving ribozymes and riboswitch expression platforms is known and the sequences can be rationally altered, the design of entirely new aptamer domains for riboswitches or completely new ribozymes is not carried out by rational design approaches. Instead, a method called *in vitro* selection is applied in which a solution of random RNA sequences of desired length is exposed to the desired ligand (in case of riboswitches)^[297] or Mg²⁺ as a cofactor (for ribozymes)^[298]. After that, non-binders are removed, while binders are PCR amplified and used for another selection round in which only the tightest binders are selected. This process is often also described as *in vitro* evolution or the common tradename SELEX.^[299] It is worth noting, that this method selects for affinity rather than specificity^[300], which results in the necessity to test selected aptamers additionally in appropriate assays such as *in vivo* screenings.^[301]

It has been shown, that the obtained functional RNAs can be utilized in a multitude of different applications. For instance, a translational riboswitch could be implemented into eukaryotic cells, where it acted by inhibiting cap dependent initiation in the 5'-UTR rather than AUG or SD masking, which is the primary mode of action in prokaryotes.^[301] This includes yeast but also mammalian cell cultures in which activity of a programmable riboswitch was observed.^[302] The control of cap independent translation could also be demonstrated *in vitro*.^[303] Artificial

riboswitches can additionally play a role in bioengineering of bacteria. Gallivan *et al.* developed an *e. coli* stem, capable of biodegradation of the herbicide atrazine in which an artificial riboswitch responsive to atrazine was coupled to a motility controlling gene in order to respond to atrazine concentration gradients.^[304] Of course, medicinal applications for artificial functional RNAs do exist as well. In a particularly interesting example Smolke *et al.* were able to interfere with human T-cell proliferation, with an RNA device located in the 3'-UTR of a gene coding for a proliferative cytokine.^[305] The construct contains an aptamer domain functionally coupled to a ribozyme. Upon binding of an external trigger compound, self-cleavage of the ribozyme results in degradation of the transcript and lowered proliferation. Such strategies could find application in cell-based therapies^[306], where direct control via external trigger signals is a highly appreciated feature.

5.3. Nano applications and molecular machines

In the recently emerging field of nano technology nucleic acids play an important role in building large scaffolds and nano assemblies but can also carry out functional tasks. Recent key studies showed, how DNA molecules can be assembled to resemble virtually any shape on nano scale.^[307] While natural nucleic acids carry already preferential properties for nano structures, their flexibility posed a problem for larger structures. This was overcome, by designing self-assembling sequences which include cross-over strands in order to introduce additional rigidity to the system.^[308] In further advances, templating of larger DNA nanostructures with short DNA scaffolds was introduced, which was named DNA-origami, allowing much larger three-dimensional nano assemblies.^[309]

Nanostructures built from DNA are not only scaffolded by DNA, but in addition can act as a scaffold themselves in order to accommodate not only other nucleic acids, but also proteins^[310] or other nanoparticles. The ability to organize enzyme complexes can be utilized in biological systems in order to modulate enzymatic turnover rates.^[311] Additionally, in order to introduce dynamics to these systems, they can be combined with functional RNA to allow various applications also in *in vivo* environments^[312], which lead up to biocomputing in which small circuits and logic gates were already shown to be active *in vitro* and *in vivo*.^[313,314] Such logic gates find application in drug delivery. Church *et al.* were able to construct a “nano-robot”, carrying two aptamer domains involved in an AND logic gate, allowing drug release only upon twofold ligand binding.^[315] In the field of biosensors, nanostructures and an aptamer domain were combined with two FRET dyes, to construct a fluorescent probe responsive to external stimuli.^[316]

A particularly interesting building block for nucleic acid nanostructures is the G-quadruplex (G4). As this very stable structural element can be composed of only single, but also up to four different DNA or RNA strands it can serve as three or four way junction, as shown by Seeman *et al.*^[317] The very small G4 structure can serve as a connecting element for large DNAs such as the 168 nt DNA minicircles employed by Heckel *et al.*, which can subsequently be visualized via atomic force microscopy (AFM).^[318] By introducing the G4 based, only 15 nt long thrombin binding aptamer an additional layer of control can be added to a desired system.^[319] Additionally, G4s form under the

influence of monovalent cations, which is exploitable using a G4 as the control element, sensing these ions in a cellular environment.^[320]

As seen above, nanostructures can not only act as rigid scaffolds, but can also carry out tasks such as drug delivery, sense metabolites or act as a molecular sensor. While the possibilities to construct dynamic nanostructures, so called nanomachines are vast, all of them require some type of fuel to operate, which can be the sensed metabolite, another nucleic acid strand or a ligand, which acts as an external trigger. The addition of these fuels or their consumption could have unexpected and undesired effects in a cellular environment ranging up to cell toxicity. The use of light as an external trigger to power nanomachines is opposingly considered safe, reliable and non-invasive.^[321] In research article IV the development and characterization of a light-driven G4 element, which could find application in nano devices, is discussed (see **Chapter 12**).

6. NMR spectroscopy of nucleic acids

Many structural and functional properties of nucleic acids were highlighted in the preceding chapters and many biochemical, bioinformatical and biophysical experimental methods allow their characterization. Herein, the application of the particularly versatile method solution state nuclear magnetic resonance (NMR) spectroscopy on nucleic acids is discussed. NMR allows monitoring of molecular alterations and interactions on a single atom level and couples unmatched resolution with the ability to observe a multitude of biomolecular properties including, but not limited to molecular interactions, structure, size or kinetics. In this chapter a few important experiments for the characterization of nucleic acids are explained from an application standpoint. For simplicity, biomolecular solution state NMR spectroscopy is referred to simply as “NMR spectroscopy” in this chapter, despite the existence other NMR spectroscopic methods, of which some, like solid state NMR, are also employed for nucleic acids.^[322]

6.1. NMR methods applied to nucleic acids

6.1.1. ¹H-1D experiments

The most basic NMR experiment is the 1D spectrum, in organic molecules and biomolecules specifically the proton (¹H) 1D spectrum. As nucleic acids are built from repetitive building blocks, the spectrum contains several regions where atoms, located at the same position in the nucleobase, cluster, displaying only a certain chemical shift dispersion (**Figure 7**). Between 10 and 15 ppm in the ¹H spectrum, imino protons of G and U residues are observed. These, with a pK_a value of 9.2^[323], acidic protons rapidly exchange with the solvent (usually buffered water). Since these exchange events are faster than the NMR experiment, they are broadened beyond detection in unfolded nucleic acids, due to detection of an ensemble of chemical shifts. Upon base-pairing and subsequent structure formation though, they become increasingly protected from solvent exchange and are therefore detectable in structured nucleic acids. Still, imino signals are generally weaker and slightly broadened compared to signals of other residues. A single imino proton resonance is observed for each sufficiently exchange protected G or U residue (each base pair in canonical base pairing), which leads, together with the inherently large shift dispersion, to a highly dispersed region, well suited for investigation in 1D-¹H experiments. Additionally, the imino region can be further divided, as canonical base-pairs resonate at lower field compared to non-canonical base pairs such as G-U wobble- or Hoogsteen base-pairs (e.g. cis-WH of G4s). Aromatic protons include H2 (only adenine) and H8 in purines and H5 and H6 in pyrimidines, which resonate between 6 and 9 ppm. Solvent exchange of these residues is extremely slow^[324], which leads to no additional broadening and therefore comparatively sharp lines. However, since one or two aromatic protons of each base are observed, the region is usually too cluttered in 1D experiments to be able to observe all signals individually, especially with increasing size of the nucleic acid investigated. This problem is even larger when the sugar region is observed, in which all sugar atoms resonate at 3.5 to 6.5 ppm. The H1' protons are shifted downfield compared to the remaining sugar protons by about 1.3 ppm and therefore overlap with H5 protons of pyrimidine residues. In DNA, due to the lack of the 2'OH-group, H2' and H2'' are observed at lower ppm values of 2 to 3 ppm.

In terms of resolution, ^1H -1D experiments are therefore far behind more sophisticated 2D or 3D experiments, like 2D-HSQC spectra (see Chapter 6.1.4). However, because they can be acquired rapidly (measurement times are in the order of minutes or even seconds), they are well suited to study comparatively fast effects, as in kinetic studies in the order of minutes or to be employed repetitively, for example in titration experiments.

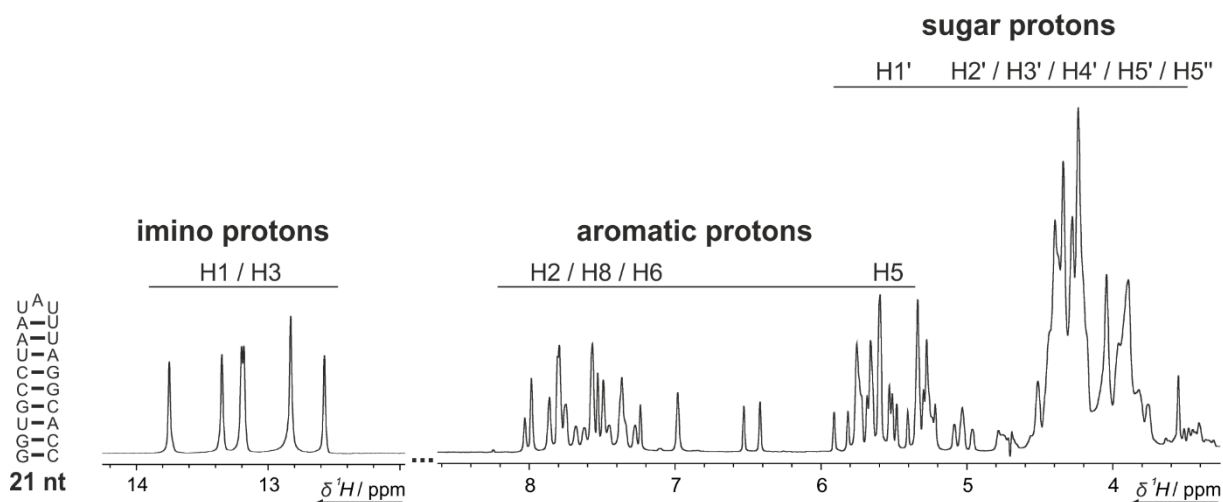


Figure 7 | ^1H -1D spectra of a 21 nt hairpin (depicted left) with annotation of the respective regions. The Imino proton spectrum was acquired with H_2O (10% D_2O) as solvent to allow monitoring of exchanging protons. The remaining spectrum was acquired with D_2O as solvent to reduce signal perturbation, arising from suppression of water signals. Resonance regions for imino and non-exchangeable proton are annotated. Imino protons are displayed with increased scaling, as these signals are usually much smaller.

6.1.2. Reporter signals and probes

NMR signals show a distinct chemical shift, based on the chemical surroundings of each individual nucleus. Consequently, molecular events like structural alteration or intermolecular interaction are best monitored on signals which stem from a nucleus located ideally in the affected residue or in close vicinity. These signals are termed reporter signals. Depending on the exchange rate between the states observed, perturbation of the chemical shift (fast exchange) or vanishing of an original signal and parallel rise of a new signal (slow exchange) is observed. In case of intermediate exchange, signals might be broadened beyond detection, which requires the use of more complex methods.^[325] In nucleic acids imino protons are often best suited, since they are highly dispersed and are therefore common reporter signals^[326–328]. Since they are only observable if the respective residue is in a structured surrounding, investigation of insufficiently structured strands becomes more complex. Detection of aromatic residues in 2D-HSQC spectra is certainly a possibility in this case, but requires several hours of measurement time or isotopic labeling (see Chapters 6.1.4 and 6.1.5) and even then, at higher molecular weight unambiguous identification of signals can be troublesome. An interesting alternative, though requiring chemical modification of the nucleic acid, is introduction of molecular probes. In this approach residues featuring atoms which resonate in a single separate signal such as methyl-groups^[329] or fluorine atoms^[330–332] are utilized. They are introduced in proximity to the interaction site and allow observation of clear reporter signals. While being preparatively challenging, this method allows

direct observations of dynamics and does not require sophisticated NMR experiments or resonance assignment.

6.1.3. Resonance assignment and NOESY

For most NMR studies it is crucial to be able to assign observed resonances to their respective residues and atoms and while atoms can often be easily distinguished by their characteristic chemical shift range, the effect of the chemical surrounding of a specific residue on its chemical shift can only be roughly predicted.^[333] An experiment called two-dimensional nuclear Overhauser effect spectroscopy (2D-NOESY) facilitates assignment of nucleic acids, especially the often crucial assignment of imino protons. It exploits the quantum chemical nuclear Overhauser effect (NOE)^[334], which allows magnetization transfer by cross-relaxation of two different nuclei. The effect depends on the spatial distance between the two nuclei ($\sim r^{-6}$) and is independent of chemical bonds. In a two-dimensional NMR-experiment it leads to formation of cross-signals between two signals, with their intensity directly related to the distance between the respective atoms up to a threshold of approximately 6 Å. In the majority of cases 2D-NOESY spectra are recorded as proton-proton experiment (2D-¹H,¹H-NOESY), which is also applied for resonance assignment of nucleic acids.

In helices, imino protons of two consecutive base pairs are distanced below the 6 Å threshold and accordingly, cross-signals from an imino proton to the n-1 and n+1 base pair are observed. This allows the assignment of imino resonances through an “imino walk” through sufficiently exchange protected (stable) stems. Signals of G and U/T residues can be distinguished by characteristic cross-signal patterns at C-amino and A-H2 resonances respectively. Further assignment of 2D-NOESY spectra is performed in the aromatic region, where a “walk” comparable to imino protons is possible for the aromatic protons as well, though these cross-signals are usually faint. Additionally, correlation between aromatic protons and H1' protons of n and n-1 residues are utilized. In principle assignment of sugar resonances, starting from assigned H1' resonances, is also possible, but in practice the spectral region shows severe overlap even in small RNAs, limiting the feasibility of this approach. Apart from assignment, NOESY spectra are highly important for collecting experimental data in NMR structure determination. Since the intensity of NOESY cross-signals depends directly on the interatomic distance, they are well suited to determine distance restraints, which can be used in structure calculations (see Chapter 6.2).

6.1.4. 2D-HSQC and related hetero correlated experiments

2D-hetero correlation experiments are another widespread subset of NMR experiments, prevalent in nucleic acid research. They are usually based on the correlation of proton nuclei with nuclei of their covalently bound hetero atom (¹³C or ¹⁵N) by J-coupling^[335], allowing a substantial increase of resolution compared to 1D experiments. Many variants of hetero correlated experiments exist operating with single quantum (HSQC)^[336] or multi quantum magnetization (HMQC)^[337], exploiting the interplay between relaxation mechanisms to gain resolution (TROSY)^[338] or different mechanisms to increase sensitivity (SOFAS, sensitivity improvement)^[339,340]. Here the example of HSQC will be used, since it is the most commonly used experiment, but all techniques mentioned lead to comparable spectra from an application standpoint.

For imino groups, ^{15}N -HSQC experiments facilitate the discrimination between guanine and uracil, since N3 atoms of uridine resonate approximately 15 ppm low field compared to guanine N1 atoms. In larger nucleic acids imino regions can appear heavily crowded in $1\text{D-}^1\text{H}$ spectra due to the high amount of base paired residues. In this case acquiring ^{15}N -HSQC spectra can be advantageous to allow unambiguous identification of all signals. Additionally, they are often applied to monitor exchange protected amino residues, though guanine and adenine residues often show broad signals, which can only be overcome by utilization of more sophisticated NMR experiments.^[341]

Sufficient resolution of aromatic resonances can be achieved in ^{13}C -HSQC experiments in which pyrimidine C6 and purine C8 resonances resonate at comparable ppm values between 135 and 145 ppm, making their discrimination problematic, though C8 resonances are found few ppm upfield compared to C6 resonances. Fortunately, C5 resonances of pyrimidines and C2 resonances of adenines appear well separated from the remaining aromatic resonances and especially adenine H2-C2 peaks can be used as reporter peaks^[342], since resonance assignments of canonically base-paired adenines can be directly obtained from uracil imino proton correlation. Acquiring ^{13}C -HSQC spectra also helps deconvoluting sugar signals and is required, if acceptable resolution in this region is desired. Different sugar atoms resonate at separate positions, allowing easy discrimination between each ribose position. However, due to their low dispersion, the regions are still highly convoluted in larger nucleic acids and only the H1'-C1' signals are frequently used for assignment purposes in larger RNAs (see Chapter 6.1.3).

HSQC experiments are well suited to characterize structural alterations for example induced by interactions with ligands^[343], proteins^[344] and cofactors^[345] or when showing general dynamics^[346] as they provide a good trade-off between measurement time and resolution provided. Measurement times can in addition be significantly shortened if isotopic labeling is involved.

6.1.5. Isotopic labeling

While NMR experiments based on protons can yield many interesting results on their own when applied to nucleic acids, the field of possible applications is significantly enlarged if isotopic labeling is employed, by increasing the amount of NMR active nuclei in the sample from natural abundance of approx. 1.1% ^{13}C and 0.4% ^{15}N to almost 100%. This procedure enhances the efficiency of hetero correlation significantly and makes magnetization transfer over hetero atoms feasible beyond ^1J -coupling as used in e.g. HSQC experiments. Since nucleic acids for NMR samples are usually produced either by solid-phase synthesis (most DNA)^[347] or *in vitro* transcription (most RNA)^[348], isotopically labeled building blocks for each method can be applied to achieve the labeling. Solid-phase synthesis allows isotopic labeling at any position in the nucleic acid, but is only feasible up to a length of approximately 55 nt.^[348] *In vitro* transcription enables preparation of larger RNA but allows, in its basic form, only residue specific strategies. However, preparative methods were developed which allow segmental labeling^[349] or even site specific labeling^[350,351]. As the efficiency of hetero correlation is increased by isotopic labeling of residues, the application of more sophisticated NMR experiments becomes practicable which aid resonance assignment as well as determination of structural parameters. It should be noted that isotopic labeling, despite all the benefits can also have disadvantages. First, isotopically labeled

samples are expensive in preparation, though manageable. Second and sometimes even more striking, when working with isotopically labeled samples decoupling of unwanted magnetization pathways is required to obtain clean spectra. The most common ways to achieve decoupling are broad band heteronuclear decoupling during the acquisition time by repetitive pulse schemes^[352] and the implementation of a constant time period in order to refocus homonuclear coupling^[353]. Both methods can lead to loss of resolution compared to spectra acquired on natural abundance samples, as they limit the amount of points which can be acquired in the indirect or direct dimension of spectra respectively.

6.1.6. 2D- and 3D-NMR methods aiding resonance assignment and structural characterization

On isotopically labeled samples numerous experiments were developed which rely on different mechanisms to relay magnetization, most importantly INEPT (Insensitive nuclei enhanced by polarization transfer)^[354], TOCSY-like isotropic mixing^[355,356] (Total correlation spectroscopy) and NOE (see Chapter 6.1.3). In this chapter, some of the most popular advanced experiments for nucleic acids are discussed, which facilitate resonance assignment and determination of angular parameters meant for structure calculation.

While imino-imino correlation can in most cases be achieved from ¹H,¹H-NOESY data, further assignment becomes increasingly complicated and even then, might stay ambiguous due to signal overlaps. Additionally, imino resonances are only observed for base paired residues. Direct correlation of resonances can help resolving these issues for isotopically labeled samples. To directly correlate neighboring residues magnetization needs to be relayed over the bridging phosphorous leading to experiments such as 3D-HCP^[357], which correlates H3',H4' and H5'/H5'' sugar protons of neighboring residues over their respective carbons and the bridging phosphorus. By addition of a TOCSY mixing time, it can be extended to a 3D-HCP-CCH-TOCSY^[358] experiment, which allows correlation of H1' atoms of neighboring residues. Their assignment can serve as basis for the assignment of aromatic protons, but also the remaining sugar protons, which already can be partially assigned from a 3D-HCP experiment.

To obtain the assignment of aromatic protons from H1' resonances, HCN^[359] experiments are employed in which magnetization is transferred with multiple INEPT steps over the N1 or N9 of pyrimidines and purines respectively. Depending on the size of the RNA and correspondingly the resolution required, these experiments can be run as 2D or 3D versions omitting the ¹³C plane in the 2D version.

In order to assign non-H1' sugar resonances, ¹H,¹H-NOESY strategies routinely fail even for small RNAs, due to the low signal dispersion and strong signal overlap in this region. A more convenient way to assign sugar resonances are 3D-HCCH-TOCSY spectra^[360], which provide the required resolution to unambiguously assign sugar resonances at least in smaller RNAs.

Imino resonances can be correlated with aromatic protons by HCCNH experiments^[361] leading to an assignment of most relevant protons. These strategies combined can be applied to achieve an assignment alone but are even more valuable to complement NOESY assignments in order to identify and remove ambiguities.

In order to collect data for a structure calculation, restraints for interatomic distances are routinely derived from NOESY experiments (see Chapter 6.1.3) but can be complemented by angular restraints. This angular information is usually obtained by two different methods, quantitative determination of coupling constants^[362–364] or determination of cross-correlated relaxation rates^[365]. Specific experiments have been developed, which allow the determination of these quantities and subsequently dihedral angles in nucleic acids. For example, C-P coupling can be determined in quantitative HCP experiments.^[366] Cross-correlated relaxation rates affected by the ³¹P chemical shift anisotropy (CSA) and the C-H bond dipole are obtained from Γ -HCP experiments^[367]. Additionally, H-P coupling constants can be obtained directly from ¹³C-FIDS-HSQC experiments.^[368] These experiments allow the determination of backbone angles α , β , γ and ζ . In analogy, angles related to the sugar-pucker are determined from coupling constants by HCCH-TOCSY-CCH-E.COSY experiments^[369] or from cross-correlated relaxation rates by Γ -HCCH-TOCSY experiments.^[370] Cross-correlated relaxation rates acquired in Γ -HCN experiments additionally allow determination of the glycosidic bond angle χ .^[371] In combination, these methods can substantially improve the quality of a structure calculated from NMR data.^[372]

In larger structures global alignment, for example between two helices, is often insufficiently sampled in the short-distance information from NOE data and long-distance restraints are required. These restraints are often obtained from residual dipolar couplings. Dipolar coupling is related to the angular difference of a dipole between two NMR active nuclei to the external magnetic field as well as the interatomic distance. While the effect averages to zero in isotropic solutions, it can be exploited in alignment media to obtain additional information about the overall structure, complementing a structure calculation.^[373]

6.2. Structure determination of nucleic acids by NMR spectroscopy

De novo determination of three-dimensional structures of biomolecules by NMR was pioneered by Kurt Wüthrich in 1985 when they released the structure of proteinase inhibitor IIA^[374], which was closely followed by the first nucleic acid structures.^[375] Today, while NMR structures account for less than 10% of total structures in the wwPDB (wwpdb.org^[376]), over 40% of annual submissions of small RNA structures (<155 nt) are NMR structures^[377] making NMR the second most popular method in this field behind X-ray crystallography.

In general, NMR structures are determined by molecular dynamics calculations which are restrained by structural parameters obtained from NMR experiments (**Figure 8**). The calculation is carried out in a simulated annealing approach where, from a number of calculated structures, a subset of structures with the lowest overall energy is selected. The structures are often presented as a bundle and the root-mean-square deviation (RMSD) of all atoms or all heavy atoms is calculated and used as a quality factor. Accordingly, the divergence of structures and therefore the RMSD largely depends on the quality and amount of NMR derived restraint data plugged into the calculation. The most commonly used data are distance restraints derived from NOE-based experiments, routinely 2D-¹H,¹H-NOESY but also 3D-NOESY-HSQC^[378] experiments. The strength of the observed proton-proton cross-peaks is correlated with the interatomic distance, which can be determined roughly for all signals based on the intensity of a well characterized fixed distance such as the H5-H6 vector.^[379] Alternatively, more exact distance restraints can be

obtained using iterative computer assisted approaches such as ARIA (Ambiguous Restraints for Iterative Assignment).^[380] The structure is then calculated in a simulated annealing protocol, often with the programs CNS^[381] or X-PLOR^[382], repetitively in order to remove errors in the initial data, which are sorted out and omitted. The structure can additionally be improved by plugging angular restraints obtained from J-couplings or cross-correlated relaxation rates (see Chapter 6.1.6) into the calculation.

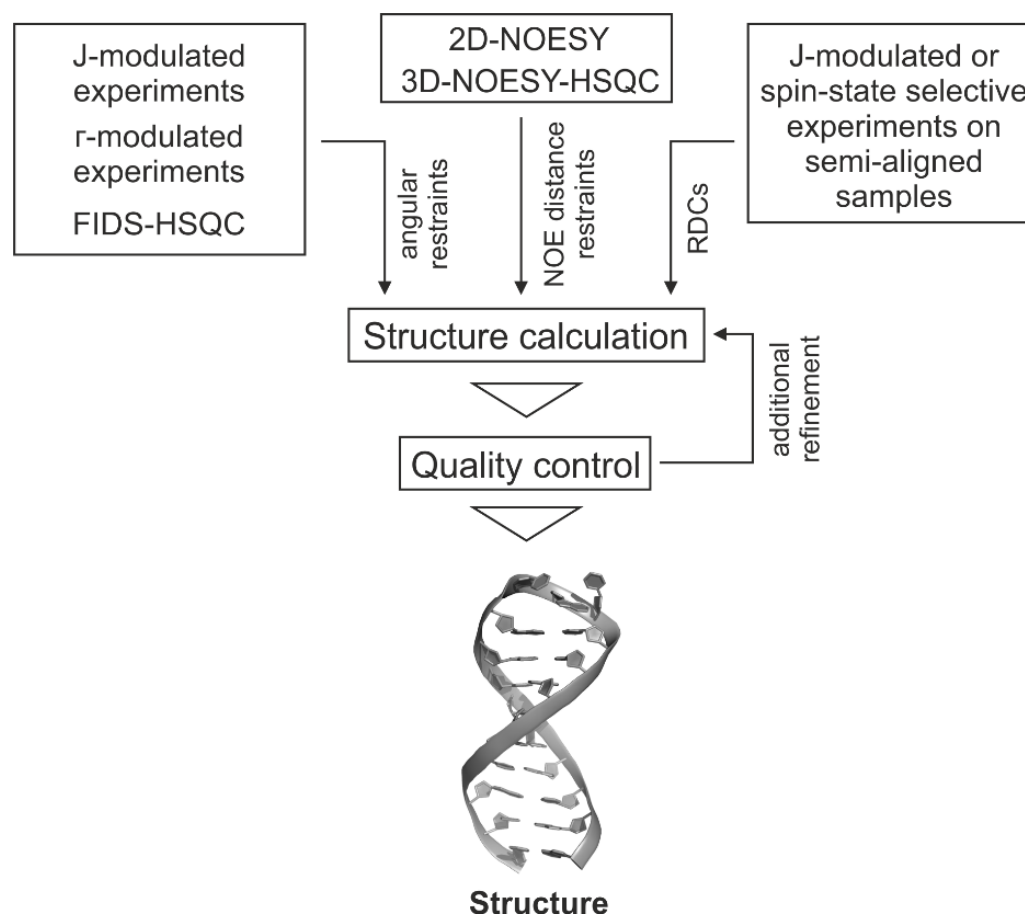


Figure 8 | Flowchart showing a typical process of structure calculation carried out for nucleic acids.

6.3. NMR of G-quadruplexes

As helices are the by far most common structural elements in nucleic acids, they are also the most common motif investigated with NMR methods. While most of the NMR experiments developed for RNA work comparably also in G4s, there are some differential aspects which should be noted when these specific structural elements are investigated. G4s are composed of G-tetrads, which are involved in very stable Hoogsteen (*cis*-WH) bonds often only bridged by short loop sequences. The guanine imino protons in G4s resonate at around 10.5 to 12 ppm and feature often less dispersion than helical RNA imino protons, especially if the guanines are comparably structured (for example all-trans all-parallel in RNA G4s). Additionally, increased signal overlap can be expected in ¹³C-HSQC spectra. Fortunately, a lot of G4s investigated are

small (<25 nt), which counteracts these problems. Imino assignment is achieved mostly through NOESY, but is complicated compared to helices, since each imino proton always shows two intra-tetrad interactions and often additional intra-strand interactions, depending on its structural context. Establishing intra-strand connectivity through H1'-H8 interactions is likewise complicated, as the visibility of those interactions depends largely on the base topology.^[383] However, if an assignment is achieved, these topological characteristics are automatically obtained from the visible interactions. Direct correlation of imino to aromatic protons in the same nucleobase can be achieved by long range HMBC at natural abundance.^[384] Since NOESY assignment of G4s is tedious but they are easily available from solid-phase synthesis it became a popular strategy to prepare sparsely isotopically labeled samples, which allows the unambiguous assignment of residues.^[385] Structure calculation is performed as in other nucleic acids, after data collection from NOESY spectra.

7. RNA and NMR in drug development

Drug development often involves screening of a large libraries of drug candidates against a certain target, to identify possible binders, which are subsequently developed into a drug. The success of these approaches in finding an actual drug largely depends on the choice of target, the library and the screening method. In this chapter some aspects of drug development with RNA targets, drug screening by NMR and fragment-based drug discovery are discussed.

7.1. Targeting RNA with small molecules

Up to now, most drug screening projects focus on protein targets, thus protein targeted drug screening and drug development are highly exhibited fields in which many medicinal advances are made. However, below 1000 proteins are currently drugged which amounts for only 0.05% of the human genome.^[386] Many pathologically relevant proteins are deemed undruggable, as they lack binding sites that can be bound specifically and with high affinity.^[387] However, while the proteins themselves are undruggable, they could potentially be indirectly targeted by drugs with a related regulatory RNA as a target. As discussed in Chapter 3 and 4, RNA molecules exhibit a vast structural space and are involved in numerous cellular processes especially in gene regulation. Additionally, some cellular processes are mediated only by RNA in bacteria, making RNA a preferable target in these cases.

RNA targeting with drug development approaches is a rising field, even though actual RNA-targeting drugs are still extremely rare, with the most prominent example being linezolid antibiotics. This compound interacts with the bacterial 50S subunit of the ribosome in order to block tRNA interaction.^[388]

Another example of bacterial RNA targets are riboswitches, which are discussed in Chapter 4.2. As they intrinsically contain a ligand binding pocket as part of their aptamer domain, they are able to recognize small molecules by default. Therefore, often analogues of cognate ligands are researched, mostly on highly prevalent riboswitches. For the TPP riboswitch the analogue pyrithiamine (PT) showed toxic effects on bacteria where TPP riboswitches were prevalent, proving the applicability of this approach.^[389] Roseoflavin, an FMN analogue has antimicrobial effects in bacteria with FMN riboswitches as well^[390] and for the lysine riboswitch, several analogues were identified which can inhibit bacterial growth.^[152]

Retroviruses, like the human immune deficiency virus (HIV) or hepatitis C virus have RNA genomes, lead to severe diseases and are therefore also frequently targeted. In HIV, the Tat/TAR complex (trans-activator of transcription/trans-activating response element), an RNA protein complex (RNP), is crucial for viral replication. The RNA TAR is targeted by drugs in order to disturb complex interaction.^[391,392] In the hepatitis C virus, an internal ribosomal entry site (IRES) is targeted, which plays a key role in cap dependent initiation.^[393]

Pathologically relevant RNA is also found in humans, particularly in genetic diseases which arise from pathogenic nucleotide repeats in the genome. In myotonic dystrophy, CUG-repeat containing RNA transcripts form pathologically relevant small hairpins which are sensed by so-called muscleblind-like proteins (MNBL), lead to disease. This interaction could be destabilized by a

small molecule compound.^[394] Interference pathways based on microRNA or lncRNA are described in Chapter 5.1 as a target for RNA-based drugs. As research on these RNAs is currently an emerging field, small molecule drugs targeting them are developed as well^[395], with small molecules having the advantages of smaller size, less charge and general better applicability than RNA-based drugs.

Endeavors are made trying to find characteristics of drugs, which specifically bind RNA over proteins in order to establish means for better design of screening libraries, since drugs for RNA differ in binding mode, size and charge from protein targeted small molecules.^[396] Interaction modes of small molecules with RNA were recently extensively reviewed and were shown to be highly versatile.^[397] In a different study, Weeks et al. argue, that many RNA based drugs show high affinity but low specificity as they mostly act by intercalation and are often highly polar. It is shown that these ligands are not sufficiently drug like e.g. suited for oral application in regard of size, charge and hydrophilicity. To facilitate target choice, they define the “information content”, a diversification factor for RNA to identify targets which allow higher possible affinity and selectivity.^[398] Modern computational methods like pattern recognition employed on screening data, additionally allow categorization and identification of popular interaction patterns of small molecules with RNA.^[399] These data provide valuable starting points for future drug screening projects.

7.2. Drug discovery by NMR

Modern drug discovery can be performed utilizing various biophysical methods, with the prerequisite that they are able to report on molecular recognition events or phenomenologically on drug interference with the precision needed to identify potential drug candidates. The principle relies on screening of a library of chemical compounds against a selected target, identifying possible candidates, so-called hits, and further development of a drug from the most promising hits. In this chapter, NMR as a method for hit identification is discussed.

NMR in drug discovery is, compared to other methods, characterized by its ability to sense small-molecule interaction with the biological target on a single atom level, allowing versatile experiments for detection and even structural elucidation of small molecule interaction.

The simplest method is the observation of chemical shift perturbation (CSP) often described as chemical-shift mapping. CSP is observed upon alteration of the chemical environment of the observed nuclei and therefore sensitive to ligand interaction. In this approach, the target is commonly isotopically labeled, and CSPs are monitored in hetero-correlated experiments (see **Chapter 6.1.4**) on the target.^[400] If resonance assignment is available the method allows localization of the binding site. Since the experiments are considerably fast, affinity constants can be determined from titration data of promising ligands. However, these experiments are acquired of the target molecule and are therefore ill-suited for ligand mixtures, since the interacting ligand cannot be easily identified. Since mixtures are beneficial to increase the screening throughput, other experiments are often employed, where ligand signals are detected in response to target interaction.

Saturation transfer difference (STD) experiments^[401] allow exactly this type of observations. In

these experiments the target molecule is selectively irradiated and upon ligand interaction, magnetization is partially transferred to the ligand via NOE. Signals of ligands unbound in solution, will display small linewidths due to their short correlation times and longer T_2 times. Based on the residence time of the ligand at the binding site, the intensity of the hit signals is attenuated. Since these differences can be subtle, the spectrum is usually subtracted from a conventional NMR spectrum leaving only signals of interacting ligands.

A comparable experiment, which adds an additional layer of complexation, is the WaterLOGSY experiment (water-ligand observed via gradient spectroscopy). In this experiment selective irradiation of the target is omitted in favor of irradiation of the bulk water (solvent) signal. Magnetization is transferred to the target in three major pathways: By NOE from water molecules in the binding site, by chemical exchange of labile protons and by NOE from water on the target surface, with the first pathway being major for magnetization passed on to the ligand.^[402,403] Consequently the transfer of magnetization is dependent on residence in the binding pocket and therefore ligand affinity, resulting in differential intensity between bound and unbound compounds, which can then be visualized in a difference experiment. WaterLOGSY experiments are described as more sensitive compared to STD experiments.^[402]

An even more sensitive method relies on the differential relaxation properties of bound compounds compared to unbound compounds. Since biomolecular targets are large molecules, they show longer rotational correlation times and therefore faster T_2 relaxation compared to small molecule ligands. Upon interaction though, ligands experience the tumbling of the biomolecule and therefore display altered relaxation behavior. These effects can be monitored by relaxation-edited NMR experiments^[404] for example employing a Carr-Purcell-Meiboom-Gill (CPMG) sequence^[405]. Differences in signal attenuation between samples containing the target molecule and pure fragment samples are then observed. Binders will show a strong difference, while non-binders are not significantly affected by presence of the macromolecule.

Most NMR based screenings rely on the observation of proton- or hetero-correlated proton resonances. While these experiments are highly sensitive, the abundance of these atoms can lead to strong convolution of spectra, especially if compound mixtures are screened in order to increase throughput. A convenient way to circumvent these limits is the introduction of fluorine atoms to either the target or the library compounds, where they act as probes in NMR applications. The NMR active isotope ^{19}F has a natural abundance of 100% and fluorine experiments show a sensitivity of 83% compared to proton NMR given their comparable gyromagnetic ratio^[406]. Additionally, ^{19}F spectra display a large chemical shift dispersion, beneficial for high-throughput studies with compound mixtures. The methods for incorporation of fluorine atoms into biological targets are developed on proteins^[407], but they can likewise be introduced into nucleosides to allow fluorination of RNA^[408]. ^{19}F -CSP is then observed upon ligand binding leading to hit identification. In general, though, it is more beneficial to use ^{19}F -containing ligands instead of biomolecules, to be able to benefit from the ability to use ligand mixtures. These libraries can be designed with the same approaches as conventional libraries, with the addition that ^{19}F must be incorporated. With the exception of WaterLOGSY the same experiments as in proton NMR can be applied^[409] bearing the advantage of showing only a single signal per

compound. Additionally, when binders which include ^{19}F are found, competition experiments with non-fluorinated compounds can be carried out in an experiment called FAXS (fluorine chemical shift anisotropy and exchange for screening) using ^{19}F as a probe while still being able to screen non fluorinated compounds.^[410]

7.3. Fragment-based drug discovery

Most drug development approaches today employ either development from an existing drug or natural compound^[411] or rely on high-throughput methods^[412], where large numbers of compounds are screened against the target. Still, ligands determined as binders in large screenings are not directly considered drugs as they usually do not inherit beneficial properties for drugs in terms of medical absorption, distribution and metabolism profile.^[413] They are referred to as lead compounds which are chemically optimized subsequent to identification. Lipinski et al. defined a set of rules for the size, charge and lipophilicity which characterizes drugs by their ability to be orally absorbed and therefore bioavailable, known as the “rule of five”.^[414] Lead compounds from conventional screening methods are in many cases already at the limits of these rules and optimization can be complicated. It would be advantageous to use smaller scaffolds which show binding but still allow room for optimization. These considerations are followed in fragment-based drug discovery attempts.

Fragment-based screening relies on the utilization of a comparatively smaller library of smaller compounds, so-called fragments, screened against the target, generating hits which enable strong modification. Sometimes, libraries are designed with the intent to comply to even stricter Lipinski rules described as a “rule of three”.^[415] Compared to hits from conventional methods, fragment hits generally cannot directly be considered lead compounds since they tend to bind with low affinities, often in the micromolar or even millimolar range.^[413] Consequently, screening methods must be able to detect these low affinity interactions.^[416] Common methods include X-ray crystallography^[417] which provides very detailed insights into the interaction, NMR with its highly versatile toolset described in Chapter 7.2, but also less sophisticated biophysical methods like surface plasmon resonance (SPR)^[418] or Microscale thermophoresis (MST)^[419], which allow higher throughput.

Hitting compounds are subsequently developed into leads, a process often described as structure-activity relationships (SAR), which commonly involves three different strategies. Growing is based on the addition of functional groups, beneficial to affinity, to the molecule scaffold. The additional groups are attached to formerly defined positions described as growth vectors.^[420] Structural information on the binding mode of the fragment aids this process strongly, allowing for a rational design approach and being commonly obtained from X-ray crystallography. However, X-ray structures are not always needed as other methods can yield the required information, or can even be completely omitted in favor of screening through several chemical derivatives.^[421] Linking is a technique which allows, if successful, rapid advance regarding affinity. Two adjacently binding fragments are connected with a linker, which has to be carefully chosen to accommodate to the geometry of binding of the two fragments. Since both fragments now contribute to the binding process with favorable interactions, substantial increases in affinity are observed.^[422] Merging is the third option and is applied when two fragments show differential interactions but

overlapping molecular scaffolds. They can then be merged into a single lead compound joining the binding interactions of both fragments comparable to a linked lead.^[416]

Fragment-based drug discovery can be considered a highly flexible method, where emphasis is put on careful lead development rather than relying on the screening process. Modern methods for structure elucidation are highly beneficial to this process and contribute largely to the success of this rising technique. Research Article V shows the results of a fragment-based screening study carried out on a pool of RNA targets employing ¹⁹F-NMR (see **Chapter 13**).

8. References

- [1] K. V. Morris, J. S. Mattick, *Nat. Rev. Genet.* **2014**, *15*, 423–437.
- [2] M. Corley, M. C. Burns, G. W. Yeo, *Mol. Cell* **2020**, *78*, 9–29.
- [3] A. Serganov, E. Nudler, *Cell* **2013**, *152*, 17–24.
- [4] S. Bajan, G. Hutvagner, *Cells* **2020**, *9*, DOI 10.3390/cells9010137.
- [5] T. R. Damase, P. B. Allen, *Bioconjug. Chem.* **2019**, *30*, 2–12.
- [6] P. Hunter, *EMBO Rep.* **2018**, *19*, 13–17.
- [7] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem* **2003**, *4*, 936–962.
- [8] M. Webba da Silva, *Methods* **2007**, *43*, 264–277.
- [9] T. Sugiki, N. Kobayashi, T. Fujiwara, *Comput. Struct. Biotechnol. J.* **2017**, *15*, 328–339.
- [10] T. Sugiki, K. Furuita, T. Fujiwara, C. Kojima, *Molecules* **2018**, *23*, 148.
- [11] A. Ren, K. R. Rajashankar, D. J. Patel, *Structure* **2015**, *23*, 1375–1381.
- [12] R. R. Breaker, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a003566–a003566.
- [13] J. Braun, S. Fischer, Z. Z. Xu, H. Sun, D. H. Ghoneim, A. T. Gimbel, U. Plessmann, H. Urlaub, D. H. Mathews, J. E. Weigand, *Nucleic Acids Res.* **2018**, *46*, 12109–12125.
- [14] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chemie Int. Ed.* **2011**, *50*, 10645–10648.
- [15] R. Norton, E. Leung, I. Chandrashekar, C. MacRaild, *Molecules* **2016**, *21*, 860.
- [16] W. G. Scott, A. Klug, *Trends Biochem. Sci.* **1996**, *21*, 220–4.
- [17] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
- [18] F. Crick, *Nature* **1970**, *227*, 561–3.
- [19] G. W. Beadle, E. L. Tatum, *Proc. Natl. Acad. Sci. U. S. A.* **1941**, *27*, 499–506.
- [20] R. A. Weinberg, S. Penman, *J. Mol. Biol.* **1968**, *38*, 289–304.
- [21] A. J. Berk, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 801–5.
- [22] Z. Miao, E. Westhof, *Annu. Rev. Biophys.* **2017**, *46*, 483–503.
- [23] W. Saenger, *Principles of Nucleic Acid Structure.*, **1986**.
- [24] E. Stofer, C. Chipot, R. Lavery, *J. Am. Chem. Soc.* **1999**, *121*, 9503–9508.
- [25] J. Stombaugh, C. L. Zirbel, E. Westhof, N. B. Leontis, *Nucleic Acids Res.* **2009**, *37*, 2294–312.
- [26] J. Šponer, K. E. Riley, P. Hobza, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595.
- [27] M. Szabat, R. Kierzek, *FEBS J.* **2017**, *284*, 3986–3998.
- [28] R. E. Dickerson, H. L. Ng, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 6986–8.
- [29] A. Rich, A. Nordheim, A. H. J. Wang, *Annu. Rev. Biochem.* **1984**, *53*, 791–846.
- [30] M. Aldwairi, R. Duwairi, W. Alqarqaz, in *2009 Int. Jt. Conf. Bioinformatics, Syst. Biol. Intell. Comput.*, IEEE, **2009**, pp. 109–115.
- [31] D. R. Groebe, O. C. Uhlenbeck, *Nucleic Acids Res.* **1988**, *16*, 11725–11735.
- [32] C. R. Woese, S. Winker, R. R. Gutell, *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8467–71.

- [33] C. Tuerk, P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. D'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1364–8.
- [34] K. S. Keating, N. Toor, A. M. Pyle, *J. Mol. Biol.* **2008**, *383*, 475–81.
- [35] S. E. Butcher, T. Dieckmann, J. Feigon, *J. Mol. Biol.* **1997**, *268*, 348–58.
- [36] C. R. Woese, R. Gutell, R. Gupta, H. F. Noller, *Microbiol. Rev.* **1983**, *47*, 621–69.
- [37] C. C. Correll, K. Swinger, *RNA* **2003**, *9*, 355–63.
- [38] L. D'Ascenzo, F. Leonarski, Q. Vicens, P. Auffinger, *RNA* **2017**, *23*, 259–269.
- [39] C. V Crowther, L. E. Jones, J. N. Morelli, E. M. Mastrogiacomo, C. Porterfield, J. L. Kent, M. J. Serra, *RNA* **2017**, *23*, 217–228.
- [40] P. N. Borer, Y. Lin, S. Wang, M. W. Roggenbuck, J. M. Gott, O. C. Uhlenbeck, I. Pelczer, *Biochemistry* **1995**, *34*, 6488–6503.
- [41] J. Karn, *J. Mol. Biol.* **1999**, *293*, 235–254.
- [42] T. Hermann, E. Westhof, *Structure* **1998**, *6*, 1303–1314.
- [43] J. L. Battiste, H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel, J. R. Williamson, *Science* . **1996**, *273*, 1547–1551.
- [44] J. L. Diener, P. B. Moore, *Mol. Cell* **1998**, *1*, 883–894.
- [45] T. Hermann, D. J. Patel, *Structure* **2000**, *8*, R47–R54.
- [46] S. J. Schroeder, M. E. Burkard, D. H. Turner, *Biopolymers* **1999**, *52*, 157–167.
- [47] G. Chen, B. M. Znosko, S. D. Kennedy, T. R. Krugh, D. H. Turner, *Biochemistry* **2005**, *44*, 2845–2856.
- [48] D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, *EMBO J.* **2001**, *20*, 4214–21.
- [49] L. Aldaz-Carroll, B. Tallet, E. Dausse, L. Yurchenko, J.-J. Toulmé, *Biochemistry* **2002**, *41*, 5883–5893.
- [50] G. Durand, E. Dausse, E. Goux, E. Fiore, E. Peyrin, C. Ravelet, J.-J. Toulmé, *Nucleic Acids Res.* **2016**, *44*, 4450–4459.
- [51] D. Dufour, M. de la Peña, S. Gago, R. Flores, J. Gallego, *Nucleic Acids Res.* **2009**, *37*, 368–381.
- [52] C. Laing, D. Wen, J. T. L. Wang, T. Schlick, *Nucleic Acids Res.* **2012**, *40*, 487–498.
- [53] P. L. ADAMS, *RNA* **2004**, *10*, 1867–1887.
- [54] A. Lescoute, E. Westhof, *RNA* **2006**, *12*, 83–93.
- [55] C. Laing, T. Schlick, *J. Mol. Biol.* **2009**, *390*, 547–59.
- [56] T. L. Beattie, J. E. Olive, R. A. Collins, *Proc. Natl. Acad. Sci.* **1995**, *92*, 4686–4690.
- [57] R. T. Batey, *Q. Rev. Biophys.* **2012**, *45*, 345–381.
- [58] S. Kolesnikova, M. Hubálek, L. Bednárová, J. Cvacka, E. A. Curtis, *Nucleic Acids Res.* **2017**, *45*, 8684–8696.
- [59] T. Ilc, P. Šket, J. Plavec, M. Webba da Silva, I. Drevenšek-Olenik, L. Spindler, *J. Phys. Chem. C* **2013**, *117*, 23208–23215.
- [60] Y. Chen, D. Yang, *Curr. Protoc. nucleic acid Chem.* **2012**, *Chapter 17*, Unit17.5.
- [61] M. Webba Da Silva, *Chem. - A Eur. J.* **2007**, *13*, 9738–9745.
- [62] D. Mohanty, M. Bansal, *Nucleic Acids Res.* **1993**, *21*, 1767–1774.

- [63] C.-D. Xiao, T. Shibata, Y. Yamamoto, Y. Xu, *Chem. Commun.* **2018**, *54*, 3944–3946.
- [64] J. S. Paige, K. Y. Wu, S. R. Jaffrey, *Science* . **2011**, *333*, 642–646.
- [65] C.-D. Xiao, T. Ishizuka, Y. Xu, *Sci. Rep.* **2017**, *7*, 6695.
- [66] H. Abou Assi, M. Garavís, C. González, M. J. Damha, *Nucleic Acids Res.* **2018**, *46*, 8038–8056.
- [67] E. P. Wright, J. L. Huppert, Z. A. E. Waller, *Nucleic Acids Res.* **2017**, *45*, 2951–2959.
- [68] A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, J. Zhu, A. D. Burton, C. B. Carlisle, C. J. Burrows, *J. Am. Chem. Soc.* **2017**, *139*, 4682–4689.
- [69] S. Dzatko, M. Krafčíková, R. Hänsel-Hertsch, T. Fessler, R. Fiala, T. Loja, D. Krafčík, J.-L. Mergny, S. Foldynova-Trantírková, L. Trantírek, *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 2165–2169.
- [70] G. Felsenfeld, D. R. Davies, A. Rich, *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.
- [71] P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, S. A. Strobel, *Nature* **2004**, *430*, 45–50.
- [72] S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, *Nat. Struct. Mol. Biol.* **2008**, *15*, 177–82.
- [73] S. E. Butcher, A. M. Pyle, *Acc. Chem. Res.* **2011**, *44*, 1302–1311.
- [74] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Müller, D. H. Mathews, M. Zuker, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9218–22.
- [75] G. Quigley, A. Rich, *Science* . **1976**, *194*, 796–806.
- [76] A. Lussier, L. Bastet, A. Chauvier, D. A. Lafontaine, *J. Biol. Chem.* **2015**, *290*, 26739–51.
- [77] P. Bouchard, P. Legault, *RNA* **2014**, *20*, 1451–64.
- [78] E. R. Gamache, J. H. Doh, J. Ritz, A. Laederach, S. Bellaousov, D. H. Mathews, M. J. Curcio, *Viruses* **2017**, *9*, DOI 10.3390/v9050093.
- [79] H. L. Townsend, B. K. Jha, R. H. Silverman, D. J. Barton, *RNA Biol.* **2008**, *5*, 263–272.
- [80] E. Ennifar, P. Walter, B. Ehresmann, C. Ehresmann, P. Dumas, *Nat. Struct. Biol.* **2001**, *8*, 1064–8.
- [81] S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, A. Rich, *Science* . **1973**, *179*, 285–288.
- [82] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* **2012**, *337*, 816–21.
- [83] M. Cobb, *Curr. Biol.* **2015**, *25*, R526–R532.
- [84] A.-L. Todeschini, A. Georges, R. A. Veitia, *Trends Genet.* **2014**, *30*, 211–219.
- [85] G. M. Cooper, in *Cell A Mol. Approach. 2nd Ed.*, **2000**.
- [86] N. Sonenberg, A. G. Hinnebusch, *Cell* **2009**, *136*, 731–45.
- [87] J. W. B. Hershey, N. Sonenberg, M. B. Mathews, *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, DOI 10.1101/cshperspect.a011528.
- [88] A. Łabno, R. Tomecki, A. Dziembowski, *Biochim. Biophys. Acta - Mol. Cell Res.* **2016**, *1863*, 3125–3147.
- [89] P. S. Lovett, E. J. Rogers, *Microbiol. Rev.* **1996**, *60*, 366–85.
- [90] K. Wethmar, J. J. Smink, A. Leutz, *Bioessays* **2010**, *32*, 885–93.
- [91] C. Vilela, J. E. G. McCarthy, *Mol. Microbiol.* **2003**, *49*, 859–867.

- [92] C. Barreau, L. Paillard, H. B. Osborne, *Nucleic Acids Res.* **2005**, *33*, 7138–50.
- [93] E. Van Assche, S. Van Puyvelde, J. Vanderleyden, H. P. Steenackers, *Front. Microbiol.* **2015**, *6*, 141.
- [94] G.-X. Ren, X.-P. Guo, Y.-C. Sun, *Front. Microbiol.* **2017**, *8*, 1276.
- [95] A. K. L. Leung, *Trends Cell Biol.* **2015**, *25*, 601–610.
- [96] J. Höck, G. Meister, *Genome Biol.* **2008**, *9*, 210.
- [97] L. F. R. Gebert, I. J. MacRae, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 21–37.
- [98] D. P. Bartel, *Cell* **2009**, *136*, 215–33.
- [99] V. Agarwal, G. W. Bell, J.-W. Nam, D. P. Bartel, *Elife* **2015**, *4*, DOI 10.7554/eLife.05005.
- [100] H. Guo, N. T. Ingolia, J. S. Weissman, D. P. Bartel, *Nature* **2010**, *466*, 835–40.
- [101] T. Fukaya, H.-O. Iwakawa, Y. Tomari, *Mol. Cell* **2014**, *56*, 67–78.
- [102] A. Rich, *Trends Biochem. Sci.* **1978**, *3*, 34–37.
- [103] S. M. Lyons, M. M. Fay, P. Ivanov, *FEBS Lett.* **2018**, *592*, 2828–2844.
- [104] F. J. Grundy, T. M. Henkin, *Cell* **1993**, *74*, 475–82.
- [105] F. J. Grundy, S. M. Rollins, T. M. Henkin, *J. Bacteriol.* **1994**, *176*, 4518–26.
- [106] W. Ross, C. E. Vrentas, P. Sanchez-Vazquez, T. Gaal, R. L. Gourse, *Mol. Cell* **2013**, *50*, 420–9.
- [107] I. V Novikova, S. P. Hennelly, K. Y. Sanbonmatsu, *Bioarchitecture n.d.*, *2*, 189–99.
- [108] R. R. Pandey, T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo, C. Kanduri, *Mol. Cell* **2008**, *32*, 232–46.
- [109] M. T. Y. Lam, W. Li, M. G. Rosenfeld, C. K. Glass, *Trends Biochem. Sci.* **2014**, *39*, 170–82.
- [110] N. Romero-Barrios, M. F. Legascue, M. Benhamed, F. Ariel, M. Crespi, *Nucleic Acids Res.* **2018**, *46*, 2169–2184.
- [111] J. Wang, X. Liu, H. Wu, P. Ni, Z. Gu, Y. Qiao, N. Chen, F. Sun, Q. Fan, *Nucleic Acids Res.* **2010**, *38*, 5366–83.
- [112] J. Carlevaro-Fita, A. Rahim, R. Guigó, L. A. Vardy, R. Johnson, *RNA* **2016**, *22*, 867–82.
- [113] A. D. Garst, A. L. Edwards, R. T. Batey, *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, DOI 10.1101/cshperspect.a003533.
- [114] K. S. Wilson, P. H. von Hippel, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8793–7.
- [115] P. J. Farnham, T. Platt, *Nucleic Acids Res.* **1981**, *9*, 563–577.
- [116] R. T. Batey, *Q. Rev. Biophys.* **2012**, *45*, 345–81.
- [117] C. Helmling, D.-P. Klötzner, F. Sochor, R. A. Mooney, A. Wacker, R. Landick, B. Fürtig, A. Heckel, H. Schwalbe, *Nat. Commun.* **2018**, *9*, 944.
- [118] J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **2005**, *18*, 49–60.
- [119] J. Shine, L. Dalgarno, *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 1342–6.
- [120] D. Omotajo, T. Tate, H. Cho, M. Choudhary, *BMC Genomics* **2015**, *16*, 604.
- [121] J. E. Johnson Jr, F. E. Reyes, J. T. Polaski, R. T. Batey, *Nature* **2012**, *492*, 133–137.
- [122] R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBioChem* **2007**, *8*, 896–902.
- [123] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, R. R. Breaker, *Nat. Struct. Biol.* **2003**,

- 10, 701–7.
- [124] A. Rentmeister, G. Mayer, N. Kuhn, M. Famulok, *Nucleic Acids Res.* **2007**, *35*, 3713–22.
- [125] A. Reining, S. Nozinovic, K. Schlepckow, F. Buhr, B. Fürtig, H. Schwalbe, *Nature* **2013**, *499*, 355–359.
- [126] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, *J. Biol. Chem.* **2002**, *277*, 48949–59.
- [127] N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **2003**, *9*, 644–7.
- [128] A. Haller, R. B. Altman, M. F. Soulière, S. C. Blanchard, R. Micura, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 4188–93.
- [129] A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **2006**, *441*, 1167–1171.
- [130] S. Li, R. R. Breaker, *Nucleic Acids Res.* **2013**, *41*, 3022–31.
- [131] M. Fontecave, M. Atta, E. Mulliez, *Trends Biochem. Sci.* **2004**, *29*, 243–9.
- [132] R. T. Batey, *Wiley Interdiscip. Rev. RNA n.d.*, *2*, 299–311.
- [133] C. Lu, F. Ding, A. Chowdhury, V. Pradhan, J. Tomsic, W. M. Holmes, T. M. Henkin, A. Ke, *J. Mol. Biol.* **2010**, *404*, 803–18.
- [134] C. Lu, A. M. Smith, R. T. Fuchs, F. Ding, K. Rajashankar, T. M. Henkin, A. Ke, *Nat. Struct. Mol. Biol.* **2008**, *15*, 1076–83.
- [135] K. Zhang, S. Li, K. Kappel, G. Pintilie, Z. Su, T.-C. Mou, M. F. Schmid, R. Das, W. Chiu, *Nat. Commun.* **2019**, *10*, 5511.
- [136] A. Sun, C. Gasser, F. Li, H. Chen, S. Mair, O. Krasheninina, R. Micura, A. Ren, *Nat. Commun.* **2019**, *10*, 5728.
- [137] Z. Weinberg, E. E. Regulski, M. C. Hammond, J. E. Barrick, Z. Yao, W. L. Ruzzo, R. R. Breaker, *RNA* **2008**, *14*, 822–8.
- [138] J. J. Trausch, Z. Xu, A. L. Edwards, F. E. Reyes, P. E. Ross, R. Knight, R. T. Batey, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6624–9.
- [139] L. Huang, D. M. J. Lilley, *Nucleic Acids Res.* **2018**, *46*, 6869–6879.
- [140] A. K. Weickhmann, H. Keller, J. P. Wurm, E. Strebiter, M. A. Juen, J. Kremser, Z. Weinberg, C. Kreutz, E. Duchardt-Ferner, J. Wöhnert, *Nucleic Acids Res.* **2019**, *47*, 2654–2665.
- [141] H. Weissbach, N. Brot, *Mol. Microbiol.* **1991**, *5*, 1593–7.
- [142] R. T. Fuchs, F. J. Grundy, T. M. Henkin, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4876–80.
- [143] F. J. Grundy, T. M. Henkin, *Mol. Microbiol.* **1998**, *30*, 737–749.
- [144] D.-J. Tang, X. Du, Q. Shi, J.-L. Zhang, Y.-P. He, Y.-M. Chen, Z. Ming, D. Wang, W.-Y. Zhong, Y.-W. Liang, J.-Y. Liu, J.-M. Huang, Y.-S. Zhong, S.-Q. An, H. Gu, J.-L. Tang, *Nat. Commun.* **2020**, *11*, 2794.
- [145] J. T. Polaski, S. M. Webster, J. E. Johnson, R. T. Batey, *J. Biol. Chem.* **2017**, *292*, 11650–11658.
- [146] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *RNA* **2003**, *9*, 1084–97.
- [147] P. K. Choudhary, R. K. O. Sigel, *RNA* **2014**, *20*, 36–45.
- [148] S. DebRoy, M. Gebbie, A. Ramesh, J. R. Goodson, M. R. Cruz, A. van Hoof, W. C. Winkler, D. A. Garsin, *Science* **2014**, *345*, 937–40.
- [149] T. L. Born, J. S. Blanchard, *Curr. Opin. Chem. Biol.* **1999**, *3*, 607–13.

- [150] J. Flores-Kim, A. J. Darwin, *Virulence* **2014**, *5*, 835–51.
- [151] A. D. Garst, A. Héroux, R. P. Rambo, R. T. Batey, *J. Biol. Chem.* **2008**, *283*, 22347–51.
- [152] K. F. Blount, J. X. Wang, J. Lim, N. Sudarsan, R. R. Breaker, *Nat. Chem. Biol.* **2007**, *3*, 44–9.
- [153] A. D. Garst, E. B. Porter, R. T. Batey, *J. Mol. Biol.* **2012**, *423*, 17–33.
- [154] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, *Nucleic Acids Res.* **2003**, *31*, 6748–57.
- [155] Y. Liu, R. H. White, W. B. Whitman, *J. Bacteriol.* **2010**, *192*, 3304–10.
- [156] S. Mukherjee, D. Barash, S. Sengupta, *PLoS One* **2017**, *12*, e0184314.
- [157] L. I. PIZER, *J. Bacteriol.* **1965**, *89*, 1145–50.
- [158] C. D. Torgerson, D. A. Hiller, S. Stav, S. A. Strobel, *RNA* **2018**, *24*, 1813–1827.
- [159] E. B. Butler, Y. Xiong, J. Wang, S. A. Strobel, *Chem. Biol.* **2011**, *18*, 293–8.
- [160] K. M. Ruff, A. Muhammad, P. J. McCown, R. R. Breaker, S. A. Strobel, *RNA* **2016**, *22*, 1728–1738.
- [161] T. V Erion, S. A. Strobel, *RNA* **2011**, *17*, 74–84.
- [162] M. Crum, N. Ram-Mohan, M. M. Meyer, *PLoS Comput. Biol.* **2019**, *15*, e1007564.
- [163] V. A. García-Angulo, *Crit. Rev. Microbiol.* **2017**, *43*, 196–209.
- [164] A. Serganov, L. Huang, D. J. Patel, *Nature* **2009**, *458*, 233–237.
- [165] Q. Vicens, E. Mondragón, R. T. Batey, *Nucleic Acids Res.* **2011**, *39*, 8586–98.
- [166] D. Pedrolli, S. Langer, B. Hobl, J. Schwarz, M. Hashimoto, M. Mack, *FEBS J.* **2015**, *282*, 3230–42.
- [167] J. E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6421–6.
- [168] M. Dambach, M. Sandoval, T. B. Updegrave, V. Anantharaman, L. Aravind, L. S. Waters, G. Storz, *Mol. Cell* **2015**, *57*, 1099–1109.
- [169] I. R. Price, A. Gaballa, F. Ding, J. D. Helmann, A. Ke, *Mol. Cell* **2015**, *57*, 1110–1123.
- [170] L. S. Waters, M. Sandoval, G. Storz, *J. Bacteriol.* **2011**, *193*, 5887–97.
- [171] A. Haller, M. F. Soulière, R. Micura, *Acc. Chem. Res.* **2011**, *44*, 1339–48.
- [172] L. C. Christiansen, S. Schou, P. Nygaard, H. H. Saxild, *J. Bacteriol.* **1997**, *179*, 2540–50.
- [173] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577–586.
- [174] M. Mandal, R. R. Breaker, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451–463.
- [175] A. Serganov, Y.-R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, D. J. Patel, *Chem. Biol.* **2004**, *11*, 1729–1741.
- [176] M. Mandal, R. R. Breaker, *Nat. Struct. Mol. Biol.* **2004**, *11*, 29–35.
- [177] S. D. Gilbert, C. D. Stoddard, S. J. Wise, R. T. Batey, *J. Mol. Biol.* **2006**, *359*, 754–68.
- [178] J. N. Kim, A. Roth, R. R. Breaker, *Proc. Natl. Acad. Sci.* **2007**, *104*, 16092–16097.
- [179] A. L. Edwards, R. T. Batey, *J. Mol. Biol.* **2009**, *385*, 938–48.
- [180] R. C. Morris, M. S. Elliott, *Mol. Genet. Metab.* **n.d.**, *74*, 147–59.

- [181] R. C. Spitale, A. T. Torelli, J. Krucinska, V. Bandarian, J. E. Wedekind, *J. Biol. Chem.* **2009**, *284*, 11012–6.
- [182] C. Pesavento, R. Hengge, *Curr. Opin. Microbiol.* **2009**, *12*, 170–6.
- [183] J. L. Botsford, J. G. Harman, *Microbiol. Rev.* **1992**, *56*, 100–22.
- [184] M. Valentini, A. Filloux, *J. Biol. Chem.* **2016**, *291*, 12547–55.
- [185] D. Cohen, S. Melamed, A. Millman, G. Shulman, Y. Oppenheimer-Shaanan, A. Kacen, S. Doron, G. Amitai, R. Sorek, *Nature* **2019**, *574*, 691–695.
- [186] 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.
- [187] K. D. Smith, S. V Lipchock, T. D. Ames, J. Wang, R. R. Breaker, S. A. Strobel, *Nat. Struct. Mol. Biol.* **2009**, *16*, 1218–1223.
- [188] 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.
- [189] K. D. Smith, C. a Shanahan, E. L. Moore, A. C. Simon, S. a Strobel, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7757–7762.
- [190] C. P. Jones, A. R. Ferré-D'Amaré, *EMBO J.* **2014**, *33*, 2692–703.
- [191] A. Gao, A. Serganov, *Nat. Chem. Biol.* **2014**, *10*, 787–92.
- [192] K. F. Block, M. C. Hammond, R. R. Breaker, *J. Bacteriol.* **2010**, *192*, 3983–9.
- [193] M. A. Kohanski, D. J. Dwyer, J. J. Collins, *Nat. Rev. Microbiol.* **2010**, *8*, 423–35.
- [194] G. S. Longo-Sorbello, J. R. Bertino, *Haematologica* **2001**, *86*, 121–7.
- [195] A. Bermingham, J. P. Derrick, *Bioessays* **2002**, *24*, 637–48.
- [196] B. R. Bochner, B. N. Ames, *Cell* **1982**, *29*, 929–37.
- [197] P. B. Kim, J. W. Nelson, R. R. Breaker, *Mol. Cell* **2015**, *57*, 317–328.
- [198] A. Becker, K. Fritz-Wolf, W. Kabsch, J. Knappe, S. Schultz, A. F. Volker Wagner, *Nat. Struct. Biol.* **1999**, *6*, 969–75.
- [199] Z. Weinberg, J. X. Wang, J. Bogue, J. Yang, K. Corbino, R. H. Moy, R. R. Breaker, *Genome Biol.* **2010**, *11*, R31.
- [200] A. Ren, K. R. Rajashankar, D. J. Patel, *Structure* **2015**, *23*, 1375–81.
- [201] T. D. Ames, D. A. Rodionov, Z. Weinberg, R. R. Breaker, *Chem. Biol.* **2010**, *17*, 681–5.
- [202] J. J. Trausch, P. Ceres, F. E. Reyes, R. T. Batey, *Structure* **2011**, *19*, 1413–23.
- [203] G. Shaw, R. Kamen, *Cell* **1986**, *46*, 659–67.
- [204] B. Conne, A. Stutz, J. D. Vassalli, *Nat. Med.* **2000**, *6*, 637–41.
- [205] Y. Audic, R. S. Hartley, *Biol. cell* **2004**, *96*, 479–98.
- [206] M.-J. Chae, H. Y. Sung, E.-H. Kim, M. Lee, H. Kwak, C. H. Chae, S. Kim, W.-Y. Park, *Exp. Mol. Med.* **2009**, *41*, 824–31.
- [207] E. Hitti, T. Bakheet, N. Al-Souhibani, W. Moghrabi, S. Al-Yahya, M. Al-Ghamdi, M. Al-Saif, M. M. Shoukri, A. Lánczky, R. Grépin, B. Györfy, G. Pagès, K. S. A. Khabar, *Cancer Res.* **2016**, *76*, 4068–80.
- [208] S. L. Schuster, A. C. Hsieh, *Trends in cancer* **2019**, *5*, 245–262.
- [209] C. A. Lagnado, C. Y. Brown, G. J. Goodall, *Mol. Cell. Biol.* **1994**, *14*, 7984–95.

- [210] W. S. Lai, D. M. Carrick, P. J. Blackshear, *J. Biol. Chem.* **2005**, *280*, 34365–77.
- [211] A. M. Zubiaga, J. G. Belasco, M. E. Greenberg, *Mol. Cell. Biol.* **1995**, *15*, 2219–30.
- [212] S. S. Peng, C. Y. Chen, A. B. Shyu, *Mol. Cell. Biol.* **1996**, *16*, 1490–9.
- [213] G. Brewer, *Mol. Cell. Biol.* **1991**, *11*, 2460–6.
- [214] Y. J. Choi, J.-H. Yoon, J. H. Chang, *Biomed Res. Int.* **2016**, *2016*, 3286191.
- [215] K. S. Sinsimer, F. M. Gratacós, A. M. Knapinska, J. Lu, C. D. Krause, A. V Wierzbowski, L. R. Maher, S. Scrudato, Y. M. Rivera, S. Gupta, D. K. Turrin, M. P. De La Cruz, S. Pestka, G. Brewer, *Mol. Cell. Biol.* **2008**, *28*, 5223–37.
- [216] N. Xu, C. Y. Chen, A. B. Shyu, *Mol. Cell. Biol.* **2001**, *21*, 6960–71.
- [217] W. S. Lai, M. L. Wells, L. Perera, P. J. Blackshear, *Wiley Interdiscip. Rev. RNA* **2019**, *10*, e1531.
- [218] M. R. Fabian, F. Frank, C. Rouya, N. Siddiqui, W. S. Lai, A. Karetnikov, P. J. Blackshear, B. Nagar, N. Sonenberg, *Nat. Struct. Mol. Biol.* **2013**, *20*, 735–9.
- [219] J. Lykke-Andersen, E. Wagner, *Genes Dev.* **2005**, *19*, 351–61.
- [220] R. Fu, M. T. Olsen, K. Webb, E. J. Bennett, J. Lykke-Andersen, *RNA* **2016**, *22*, 373–82.
- [221] R. Gherzi, K.-Y. Lee, P. Briata, D. Wegmüller, C. Moroni, M. Karin, C.-Y. Chen, *Mol. Cell* **2004**, *14*, 571–83.
- [222] C.-F. Chou, A. Mulky, S. Maitra, W.-J. Lin, R. Gherzi, J. Kappes, C.-Y. Chen, *Mol. Cell. Biol.* **2006**, *26*, 3695–706.
- [223] C. Y. Chen, R. Gherzi, S. E. Ong, E. L. Chan, R. Raijmakers, G. J. Pruijn, G. Stoecklin, C. Moroni, M. Mann, M. Karin, *Cell* **2001**, *107*, 451–64.
- [224] W. J. Ma, S. Cheng, C. Campbell, A. Wright, H. Furneaux, *J. Biol. Chem.* **1996**, *271*, 8144–51.
- [225] W. J. Ma, S. Chung, H. Furneaux, *Nucleic Acids Res.* **1997**, *25*, 3564–9.
- [226] N. Ripin, J. Boudet, M. M. Duszczczyk, A. Hinniger, M. Faller, M. Krepl, A. Gadi, R. J. Schneider, J. Šponer, N. C. Meisner-Kober, F. H.-T. Allain, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 2935–2944.
- [227] V. Katsanou, O. Papadaki, S. Milatos, P. J. Blackshear, P. Anderson, G. Kollias, D. L. Kontoyiannis, *Mol. Cell* **2005**, *19*, 777–89.
- [228] A. Fukao, Y. Sasano, H. Imataka, K. Inoue, H. Sakamoto, N. Sonenberg, C. Thoma, T. Fujiwara, *Mol. Cell* **2009**, *36*, 1007–17.
- [229] T. Fujiwara, A. Fukao, Y. Sasano, H. Matsuzaki, U. Kikkawa, H. Imataka, K. Inoue, S. Endo, N. Sonenberg, C. Thoma, H. Sakamoto, *Nucleic Acids Res.* **2012**, *40*, 1944–53.
- [230] Q. Jing, S. Huang, S. Guth, T. Zarubin, A. Motoyama, J. Chen, F. Di Padova, S.-C. Lin, H. Gram, J. Han, *Cell* **2005**, *120*, 623–34.
- [231] E. J. Fialcowitz, B. Y. Brewer, B. P. Keenan, G. M. Wilson, *J. Biol. Chem.* **2005**, *280*, 22406–17.
- [232] G. Stoecklin, M. Lu, B. Rattenbacher, C. Moroni, *Mol. Cell. Biol.* **2003**, *23*, 3506–15.
- [233] K. Leppek, J. Schott, S. Reitter, F. Poetz, M. C. Hammond, G. Stoecklin, *Cell* **2013**, *153*, 869–81.
- [234] T. Mino, Y. Murakawa, A. Fukao, A. Vandenberg, H.-H. Wessels, D. Ori, T. Uehata, S. Tartey, S. Akira, Y. Suzuki, C. G. Vinuesa, U. Ohler, D. M. Standley, M. Landthaler, T. Fujiwara, O. Takeuchi, *Cell* **2015**, *161*, 1058–1073.
- [235] A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res.* **2005**, *33*, 2901–7.

- [236] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–16.
- [237] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2007**, *35*, 406–13.
- [238] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.
- [239] F. Kouzine, D. Wojtowicz, L. Baranello, A. Yamane, S. Nelson, W. Resch, K.-R. Kieffer-Kwon, C. J. Benham, R. Casellas, T. M. Przytycka, D. Levens, *Cell Syst.* **2017**, *4*, 344–356.e7.
- [240] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* **1998**, *26*, 1167–1172.
- [241] S. Cogoi, L. E. Xodo, *Nucleic Acids Res.* **2006**, *34*, 2536–49.
- [242] H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle, S. Balasubramanian, *Biochemistry* **2006**, *45*, 7854–60.
- [243] N. Kim, *Curr. Med. Chem.* **2019**, *26*, 2898–2917.
- [244] S. Xiao, J.-Y. Zhang, K.-W. Zheng, Y.-H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 10379–90.
- [245] K. Zheng, S. Xiao, J. Liu, J. Zhang, Y. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 5533–41.
- [246] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 75–80.
- [247] J. S. Smith, Q. Chen, L. A. Yatsunyk, J. M. Nicoludis, M. S. Garcia, R. Kranaster, S. Balasubramanian, D. Monchaud, M.-P. Teulade-Fichou, L. Abramowitz, D. C. Schultz, F. B. Johnson, *Nat. Struct. Mol. Biol.* **2011**, *18*, 478–85.
- [248] K. Takahama, A. Takada, S. Tada, M. Shimizu, K. Sayama, R. Kurokawa, T. Oyoshi, *Chem. Biol.* **2013**, *20*, 341–50.
- [249] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle, L. H. Hurley, *J. Med. Chem.* **1997**, *40*, 2113–6.
- [250] S. M. Gowan, R. Heald, M. F. Stevens, L. R. Kelland, *Mol. Pharmacol.* **2001**, *60*, 981–8.
- [251] A. Joachimi, A. Benz, J. S. Hartig, *Bioorganic Med. Chem.* **2009**, *17*, 6811–6815.
- [252] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, **2014**, DOI 10.1038/NCHEM.1805.
- [253] C. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers, S. Balasubramanian, *Nat. Methods* **2016**, *13*, 841–844.
- [254] Y. Ding, Y. Tang, C. K. Kwok, Y. Zhang, P. C. Bevilacqua, S. M. Assmann, *Nature* **2014**, *505*, 696–700.
- [255] C. K. Kwok, A. B. Sahakyan, S. Balasubramanian, *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 8958–61.
- [256] D. Loughrey, K. E. Watters, A. H. Settle, J. B. Lucks, *Nucleic Acids Res.* **2014**, *42*, DOI 10.1093/nar/gku909.
- [257] J. U. Guo, D. P. Bartel, *Science* . **2016**, *353*.
- [258] M. M. Fay, S. M. Lyons, P. Ivanov, *J. Mol. Biol.* **2017**, *429*, 2127–2147.
- [259] C. Schaeffer, B. Bardoni, J. L. Mandel, B. Ehresmann, C. Ehresmann, H. Moine, *EMBO J.* **2001**, *20*, 4803–13.
- [260] S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nat. Chem. Biol.* **2007**, *3*, 218–221.
- [261] M.-C. Didiot, Z. Tian, C. Schaeffer, M. Subramanian, J.-L. Mandel, H. Moine, *Nucleic Acids Res.* **2008**, *36*, 4902–12.
- [262] D. Gomez, T. Lamarteleur, L. Lacroix, P. Mailliet, J. L. Mergny, J. F. Riou, *Nucleic Acids*

- Res. **2004**, 32, 371–379.
- [263] S. Khateb, P. Weisman-Shomer, I. Hershco-Shani, A. L. Ludwig, M. Fry, *Nucleic Acids Res.* **2007**, 35, 5775–88.
- [264] J. R. Babendure, J. L. Babendure, J.-H. Ding, R. Y. Tsien, *RNA* **2006**, 12, 851–61.
- [265] S. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nat. Chem. Biol.* **2007**, 3, 218–221.
- [266] A. L. Wolfe, K. Singh, Y. Zhong, P. Drewe, V. K. Rajasekhar, V. R. Sanghvi, K. J. Mavrikakis, M. Jiang, J. E. Roderick, J. Van der Meulen, J. H. Schatz, C. M. Rodrigo, C. Zhao, P. Rondou, E. de Stanchina, J. Teruya-Feldstein, M. a. Kelliher, F. Speleman, J. a. Porco, J. Pelletier, G. Rättsch, H.-G. Wendel, *Nature* **2014**, DOI 10.1038/nature13485.
- [267] S. C. Lee, J. Zhang, J. Strom, D. Yang, T. N. Dinh, K. Kappeler, Q. M. Chen, *Mol. Cell. Biol.* **2017**, 37, DOI 10.1128/MCB.00122-16.
- [268] E. Crenshaw, B. P. Leung, C. K. Kwok, M. Sharoni, K. Olson, N. P. Sebastian, S. Ansaloni, R. Schweitzer-Stenner, M. R. Akins, P. C. Bevilacqua, A. J. Saunders, *PLoS One* **2015**, 10, e0143160.
- [269] T. Endoh, Y. Kawasaki, N. Sugimoto, *Angew. Chem. Int. Ed. Engl.* **2013**, 52, 5522–6.
- [270] C.-H. Yu, M.-P. Teulade-Fichou, R. C. L. Olsthoorn, *Nucleic Acids Res.* **2014**, 42, 1887–92.
- [271] P. S. Bagga, L. P. Ford, F. Chen, J. Wilusz, *Nucleic Acids Res.* **1995**, 23, 1625–31.
- [272] M. Subramanian, F. Rage, R. Tabet, E. Flatter, J.-L. Mandel, H. Moine, *EMBO Rep.* **2011**, 12, 697–704.
- [273] K. Matsumura, Y. Kawasaki, M. Miyamoto, Y. Kamoshida, J. Nakamura, L. Negishi, S. Suda, T. Akiyama, *Oncogene* **2017**, 36, 1191–1199.
- [274] G. Mirihana Arachchilage, A. C. Dassanayake, S. Basu, *Chem. Biol.* **2015**, 22, 262–272.
- [275] S. Pandey, P. Agarwala, G. G. Jayaraj, R. Gargallo, S. Maiti, *Biochemistry* **2015**, 54, 7067–78.
- [276] R. C. Friedman, K. K.-H. Farh, C. B. Burge, D. P. Bartel, *Genome Res.* **2009**, 19, 92–105.
- [277] A. F. Christopher, R. P. Kaur, G. Kaur, A. Kaur, V. Gupta, P. Bansal, *Perspect. Clin. Res. n.d.*, 7, 68–74.
- [278] J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, T. R. Golub, *Nature* **2005**, 435, 834–8.
- [279] E. Callegari, L. Gramantieri, M. Domenicali, L. D'Abundo, S. Sabbioni, M. Negrini, *Cell Death Differ.* **2015**, 22, 46–57.
- [280] J. Soutschek, A. Akinc, B. Bramlage, K. Charisse, R. Constien, M. Donoghue, S. Elbashir, A. Geick, P. Hadwiger, J. Harborth, M. John, V. Kesavan, G. Lavine, R. K. Pandey, T. Racie, K. G. Rajeev, I. Röhl, I. Toudjarska, G. Wang, S. Wuschko, D. Bumcrot, V. Kotliansky, S. Limmer, M. Manoharan, H.-P. Vornlocher, *Nature* **2004**, 432, 173–8.
- [281] Y. Zeng, X. Cai, B. R. Cullen, *Methods Enzymol.* **2005**, 392, 371–80.
- [282] S. Q. Harper, P. D. Staber, X. He, S. L. Eliason, I. H. Martins, Q. Mao, L. Yang, R. M. Kotin, H. L. Paulson, B. L. Davidson, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 5820–5.
- [283] M. Amarguoui, J. J. Rossi, *Methods Mol. Biol.* **2008**, 442, 3–10.
- [284] S. Shukla, C. S. Sumaria, P. I. Pradeepkumar, *ChemMedChem* **2010**, 5, 328–49.
- [285] Y.-L. Chiu, T. M. Rana, *RNA* **2003**, 9, 1034–48.
- [286] P. Li, Z. A. Sergueeva, M. Dobrikov, B. R. Shaw, *Chem. Rev.* **2007**, 107, 4746–96.

- [287] F. Czauderna, *Nucleic Acids Res.* **2003**, *31*, 2705–2716.
- [288] A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen, J. Wengel, *Tetrahedron* **1998**, *54*, 3607–3630.
- [289] E. Mercuri, B. T. Darras, C. A. Chiriboga, J. W. Day, C. Campbell, A. M. Connolly, S. T. Iannaccone, J. Kirschner, N. L. Kuntz, K. Saito, P. B. Shieh, M. Tulinius, E. S. Mazzone, J. Montes, K. M. Bishop, Q. Yang, R. Foster, S. Gheuens, C. F. Bennett, W. Farwell, E. Schneider, D. C. De Vivo, R. S. Finkel, CHERISH Study Group, *N. Engl. J. Med.* **2018**, *378*, 625–635.
- [290] R. D. Santos, F. J. Raal, A. L. Catapano, J. L. Witztum, E. Steinhagen-Thiessen, S. Tsimikas, *Arterioscler. Thromb. Vasc. Biol.* **2015**, *35*, 689–99.
- [291] J. R. Mendell, L. R. Rodino-Klapac, Z. Sahenk, K. Roush, L. Bird, L. P. Lowes, L. Alfano, A. M. Gomez, S. Lewis, J. Kota, V. Malik, K. Shontz, C. M. Walker, K. M. Flanigan, M. Corridore, J. R. Kean, H. D. Allen, C. Shilling, K. R. Melia, P. Sazani, J. B. Saoud, E. M. Kaye, Eteplirsen Study Group, *Ann. Neurol.* **2013**, *74*, 637–47.
- [292] S. T. Croke, J. L. Witztum, C. F. Bennett, B. F. Baker, *Cell Metab.* **2018**, *27*, 714–739.
- [293] Z. Liang, H. Wu, S. Reddy, A. Zhu, S. Wang, D. Blevins, Y. Yoon, Y. Zhang, H. Shim, *Biochem. Biophys. Res. Commun.* **2007**, *363*, 542–6.
- [294] A. K. Wheatley, M. Kramski, M. R. Alexander, J. G. Toe, R. J. Center, D. F. J. Purcell, *PLoS One* **2011**, *6*, e18225.
- [295] C. Berens, F. Groher, B. Suess, *Biotechnol. J.* **2015**, *10*, 246–57.
- [296] M. Wieland, J. S. Hartig, *ChemBioChem* **2008**, *9*, 1873–1878.
- [297] S. V Harbaugh, J. A. Martin, J. Weinstein, G. Ingram, N. Kelley-Loughnane, *Methods* **2018**, *143*, 77–89.
- [298] J. Tang, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5784–9.
- [299] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–10.
- [300] J. M. Carothers, S. C. Oestreich, J. W. Szostak, *J. Am. Chem. Soc.* **2006**, *128*, 7929–7937.
- [301] J. E. Weigand, M. Sanchez, E.-B. Gunnesch, S. Zeiher, R. Schroeder, B. Suess, *RNA* **2008**, *14*, 89–97.
- [302] T. S. Bayer, C. D. Smolke, *Nat. Biotechnol.* **2005**, *23*, 337–43.
- [303] A. Ogawa, *RNA* **2011**, *17*, 478–88.
- [304] J. Sinha, S. J. Reyes, J. P. Gallivan, *Nat. Chem. Biol.* **2010**, *6*, 464–70.
- [305] Y. Y. Chen, M. C. Jensen, C. D. Smolke, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8531–6.
- [306] C. H. June, *J. Clin. Invest.* **2007**, *117*, 1204–12.
- [307] M. R. Jones, N. C. Seeman, C. A. Mirkin, *Science* . **2015**, *347*, 1260901–1260901.
- [308] N. C. Seeman, *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
- [309] P. W. K. Rothmund, *Nature* **2006**, *440*, 297–302.
- [310] O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman, I. Willner, *Nat. Nanotechnol.* **2009**, *4*, 249–254.
- [311] C. Zeng, N. Lu, Y. Wen, G. Liu, R. Zhang, J. Zhang, F. Wang, X. Liu, Q. Li, Z. Tang, M. Zhang, *ACS Appl. Mater. Interfaces* **2019**, *11*, 1790–1799.
- [312] D. Y. Zhang, G. Seelig, *Nat. Chem.* **2011**, *3*, 103–13.
- [313] L. Qian, E. Winfree, J. Bruck, *Nature* **2011**, *475*, 368–72.

- [314] K. Rinaudo, L. Bleris, R. Maddamsetti, S. Subramanian, R. Weiss, Y. Benenson, *Nat. Biotechnol.* **2007**, *25*, 795–801.
- [315] S. M. Douglas, I. Bachelet, G. M. Church, *Science* **2012**, *335*, 831–4.
- [316] R. Hu, X. Zhang, Z. Zhao, G. Zhu, T. Chen, T. Fu, W. Tan, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 5821–6.
- [317] R. Sha, L. Xiang, C. Liu, A. Balaeff, Y. Zhang, P. Zhang, Y. Li, D. N. Beratan, N. Tao, N. C. Seeman, *Nat. Nanotechnol.* **2018**, *13*, 316–321.
- [318] D. P. N. Gonçalves, T. L. Schmidt, M. B. Koeppel, A. Heckel, *Small* **2010**, *6*, 1347–52.
- [319] Y. Liu, C. Lin, H. Li, H. Yan, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 4333–8.
- [320] L. Olejko, P. J. Cywiński, I. Bald, *Nanoscale* **2016**, *8*, 10339–47.
- [321] K. Villa, M. Pumera, *Chem. Soc. Rev.* **2019**, *48*, 4966–4978.
- [322] M. Dračinský, P. Hodgkinson, *RSC Adv.* **2015**, *5*, 12300–12310.
- [323] P. Thaplyal, P. C. Bevilacqua, *Methods Enzymol.* **2014**, *549*, 189–219.
- [324] R. Brandes, A. Ehrenberg, *Nucleic Acids Res.* **1986**, *14*, 9491–508.
- [325] N. J. Anthis, G. M. Clore, *Q. Rev. Biophys.* **2015**, *48*, 35–116.
- [326] J. Buck, B. Fürtig, J. Noeske, J. Wönert, H. Schwalbe, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15699–15704.
- [327] A. Bugaut, P. Murat, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19953–19956.
- [328] N. K. Harika, A. Paul, E. Stroeve, Y. Chai, D. W. Boykin, M. W. Germann, W. D. Wilson, *Nucleic Acids Res.* **2016**, *44*, 4519–27.
- [329] K. Kloiber, R. Spitzer, M. Tollinger, R. Konrat, C. Kreutz, *Nucleic Acids Res.* **2011**, *39*, 4340–51.
- [330] M. R. Baranowski, M. Warminski, J. Jemielity, J. Kowalska, *Nucleic Acids Res.* **2020**, DOI 10.1093/nar/gkaa470.
- [331] L. Granqvist, P. Virta, *Chemistry* **2016**, *22*, 15360–15372.
- [332] F. Sochor, R. Silvers, D. Müller, C. Richter, B. Fürtig, H. Schwalbe, *J. Biomol. NMR* **2016**, *64*, 63–74.
- [333] A. Bahrami, L. J. Clos, J. L. Markley, S. E. Butcher, H. R. Eghbalnia, *J. Biomol. NMR* **2012**, *52*, 289–302.
- [334] A. W. Overhauser, *Phys. Rev.* **1953**, *92*, 411–415.
- [335] E. L. Hahn, D. E. Maxwell, *Phys. Rev.* **1952**, *88*, 1070–1084.
- [336] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185–189.
- [337] A. Bax, R. H. Griffey, B. L. Hawkins, *J. Magn. Reson.* **1983**, *55*, 301–315.
- [338] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci.* **1997**, *94*, 12366–12371.
- [339] P. Schanda, E. Kupce, B. Brutscher, *J. Biomol. NMR* **2005**, *33*, 199–211.
- [340] A. G. Palmer, J. Cavanagh, P. E. Wright, M. Rance, *J. Magn. Reson.* **1991**, *93*, 151–170.
- [341] R. Schnieders, A. C. Wolter, C. Richter, J. Wöhnert, H. Schwalbe, B. Fürtig, *Angew. Chemie* **2019**, *131*, 9238–9242.
- [342] R. G. Brinson, A. L. Szakal, J. P. Marino, *J. Virol.* **2011**, *85*, 13398–13408.
- [343] A. Wacker, J. Buck, D. Mathieu, C. Richter, J. Wöhnert, H. Schwalbe, *Nucleic Acids Res.*

2011, 39, 6802–12.

- [344] T. Carlomagno, *J. Magn. Reson.* **2014**, *241*, 126–136.
- [345] J. Buck, Y.-L. Li, C. Richter, J. Vergne, M.-C. Maurel, H. Schwalbe, *ChemBioChem* **2009**, *10*, 2100–2110.
- [346] K.-I. Oh, J. Kim, C.-J. Park, J.-H. Lee, *Int. J. Mol. Sci.* **2020**, *21*, DOI 10.3390/ijms21082673.
- [347] O. Becette, L. T. Olenginski, T. K. Dayie, *Molecules* **2019**, *24*, DOI 10.3390/molecules24193476.
- [348] K. Lu, Y. Miyazaki, M. F. Summers, *J. Biomol. NMR* **2010**, *46*, 113–25.
- [349] J. Xu, J. Lapham, D. M. Crothers, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 44–8.
- [350] I. Kawahara, K. Haruta, Y. Ashihara, D. Yamanaka, M. Kuriyama, N. Toki, Y. Kondo, K. Teruya, J. Ishikawa, H. Furuta, Y. Ikawa, C. Kojima, Y. Tanaka, *Nucleic Acids Res.* **2012**, *40*, 1–8.
- [351] S. Keyhani, T. Goldau, A. Blümmler, A. Heckel, H. Schwalbe, *Angew. Chemie Int. Ed.* **2018**, *57*, 12017–12021.
- [352] M. H. Levitt, R. Freeman, T. Frenkiel, *J. Magn. Reson.* **1982**, *47*, 328–330.
- [353] G. W. Vuister, A. Bax, *J. Magn. Reson.* **1992**, *98*, 428–435.
- [354] G. A. Morris, R. Freeman, *J. Am. Chem. Soc.* **1979**, *101*, 760–762.
- [355] L. Braunschweiler, R. . Ernst, *J. Magn. Reson.* **1983**, *53*, 521–528.
- [356] A. Bax, D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355–360.
- [357] J. P. Marino, D. M. Crothers, H. Schwalbe, C. Anklin, W. Bermel, C. Griesinger, *J. Am. Chem. Soc.* **1994**, *116*, 6472–6473.
- [358] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Biomol. NMR* **1995**, *5*, 87–92.
- [359] V. Sklenář, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1993**, *3*, 721–727.
- [360] L. E. Kay, G. Y. Xu, A. U. Singer, D. R. Muhandiram, J. D. Formankay, *J. Magn. Reson. Ser. B* **1993**, *101*, 333–337.
- [361] V. Sklenář, T. Dieckmann, S. E. Butcher, J. Feigon, *J. Biomol. NMR* **1996**, *7*, 83–87.
- [362] W. Bermel, K. Wagner, C. Griesinger, *J. Magn. Reson.* **1989**, *83*, 223–232.
- [363] G. E. Martin, **2002**, pp. 37–100.
- [364] C. Griesinger, O. . Sørensen, R. . Ernst, *J. Magn. Reson.* **1987**, *75*, 474–492.
- [365] B. Brutscher, I. D. B. Structurale, J. E. C. N. R. S. E. A, J. Horowitz, G. Cedex, **n.d.**
- [366] C. Richter, B. Reif, K. Wörner, S. Quant, J. P. Marino, W. Engels, C. Griesinger, H. Schwalbe, **1998**, *2*, 223–230.
- [367] S. Nozinovic, C. Richter, J. Rinnenthal, B. Fürtig, E. Duchardt-Ferner, J. E. Weigand, H. Schwalbe, *J. Am. Chem. Soc.* **2010**, *132*, 10318–10329.
- [368] H. Schwalbe, W. Samstag, J. W. Engels, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1993**, *3*, 479–486.
- [369] E. Duchardt, C. Richter, B. Reif, S. J. Glaser, J. W. Engels, C. Griesinger, H. Schwalbe, *J. Biomol. NMR* **2001**, *21*, 117–26.
- [370] I. C. Felli, C. Richter, C. Griesinger, H. Schwalbe, *J. Am. Chem. Soc.* **1999**, *121*, 1956–1957.

- [371] E. Duchardt, C. Richter, O. Ohlenschläger, M. Görlach, J. Wöhnert, H. Schwalbe, *J. Am. Chem. Soc.* **2004**, *126*, 1962–70.
- [372] S. Nozinovic, B. Fürtig, H. R. A. Jonker, C. Richter, H. Schwalbe, *Nucleic Acids Res.* **2010**, *38*, 683–94.
- [373] R. S. Lipsitz, N. Tjandra, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 387–413.
- [374] M. P. Williamson, T. F. Havel, K. Wüthrich, *J. Mol. Biol.* **1985**, *182*, 295–315.
- [375] D. J. Patel, L. Shapiro, D. Hare, *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 423–454.
- [376] H. Berman, K. Henrick, H. Nakamura, *Nat. Struct. Mol. Biol.* **2003**, *10*, 980–980.
- [377] R. P. Barnwal, F. Yang, G. Varani, *Arch. Biochem. Biophys.* **2017**, *628*, 42–56.
- [378] G. M. Clore, A. M. Gronenborn, *Prog. Nucl. Magn. Reson. Spectrosc.* **1991**, *23*, 43–92.
- [379] B. R. Reid, K. Banks, P. Flynn, W. Nerdal, *Biochemistry* **1989**, *28*, 10001–10007.
- [380] W. Rieping, B. Bardiaux, A. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381–382.
- [381] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. D. Biol. Crystallogr.* **1998**, *54*, 905–21.
- [382] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. Marius Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.
- [383] M. Adrian, B. Heddi, A. T. Phan, *Methods* **2012**, *57*, 11–24.
- [384] A. T. Phan, *J. Biomol. NMR* **2000**, *16*, 175–178.
- [385] A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
- [386] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, J. P. Overington, *Nat. Rev. Drug Discov.* **2017**, *16*, 19–34.
- [387] S. J. Dixon, B. R. Stockwell, *Curr. Opin. Chem. Biol.* **2009**, *13*, 549–55.
- [388] C. Roger, J. A. Roberts, L. Muller, *Clin. Pharmacokinet.* **2018**, *57*, 559–575.
- [389] N. Sudarsan, S. Cohen-Chalamish, S. Nakamura, G. M. Emilsson, R. R. Breaker, *Chem. Biol.* **2005**, *12*, 1325–1335.
- [390] E. R. Lee, K. F. Blount, R. R. Breaker, *RNA Biol.* **2009**, *6*, 187–194.
- [391] H. Y. Mei, D. P. Mack, A. A. Galan, N. S. Halim, A. Heldsinger, J. A. Loo, D. W. Moreland, K. A. Sannes-Lowery, L. Sharmeen, H. N. Truong, A. W. Czarnik, *Bioorg. Med. Chem.* **1997**, *5*, 1173–84.
- [392] A. Davidson, T. C. Leeper, Z. Athanassiou, K. Patora-Komisarska, J. Karn, J. A. Robinson, G. Varani, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931–6.
- [393] P. P. Seth, A. Miyaji, E. A. Jefferson, K. A. Sannes-Lowery, S. A. Osgood, S. S. Propp, R. Ranken, C. Massire, R. Sampath, D. J. Ecker, E. E. Swayze, R. H. Griffey, *J. Med. Chem.* **2005**, *48*, 7099–102.
- [394] J. F. Arambula, S. R. Ramisetty, A. M. Baranger, S. C. Zimmerman, *Proc. Natl. Acad. Sci.* **2009**, *106*, 16068–16073.
- [395] M. Matsui, D. R. Corey, *Nat. Rev. Drug Discov.* **2017**, *16*, 167–179.
- [396] B. S. Morgan, J. E. Forte, R. N. Culver, Y. Zhang, A. E. Hargrove, *Angew. Chemie Int. Ed.* **2017**, *56*, 13498–13502.
- [397] G. Padroni, N. N. Patwardhan, M. Schapira, A. E. Hargrove, *RSC Med. Chem.* **2020**, DOI 10.1039/D0MD00167H.

- [398] K. D. Warner, C. E. Hajdin, K. M. Weeks, *Nat. Rev. Drug Discov.* **2018**, *17*, 547–558.
- [399] C. S. Eubanks, A. E. Hargrove, *Biochemistry* **2019**, *58*, 199–213.
- [400] M. Pellecchia, D. S. Sem, K. Wüthrich, *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219.
- [401] B. Meyer, J. Klein, M. Mayer, R. Meinecke, H. Möller, A. Neffe, O. Schuster, J. Wülken, Y. Ding, O. Knaie, J. Labbe, M. M. Palcic, O. Hindsgaul, B. Wagner, B. Ernst, in *Leucoc. Traffick.*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2004**, pp. 149–167.
- [402] C. Raingeval, O. Cala, B. Brion, M. Le Borgne, R. E. Hubbard, I. Krimm, *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 1218–1225.
- [403] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349–359.
- [404] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, *J. Am. Chem. Soc.* **1997**, *119*, 12257–12261.
- [405] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **1954**, *94*, 630–638.
- [406] J. T. Gerig, **1989**, pp. 3–23.
- [407] C. T. Gee, K. E. Arntson, A. K. Urick, N. K. Mishra, L. M. L. Hawk, A. J. Wisniewski, W. C. K. Pomerantz, *Nat. Protoc.* **2016**, *11*, 1414–1427.
- [408] P. Liu, A. Sharon, C. K. Chu, *J. Fluor. Chem.* **2008**, *129*, 743–766.
- [409] C. Dalvit, P. E. Fagerness, D. T. A. Hadden, R. W. Sarver, B. J. Stockman, *J. Am. Chem. Soc.* **2003**, *125*, 7696–7703.
- [410] C. Dalvit, A. Vulpetti, *J. Med. Chem.* **2019**, *62*, 2218–2244.
- [411] A. HARVEY, *Drug Discov. Today* **2008**, *13*, 894–901.
- [412] D. A. Pereira, J. A. Williams, *Br. J. Pharmacol.* **2007**, *152*, 53–61.
- [413] C. W. Murray, D. C. Rees, *Nat. Chem.* **2009**, *1*, 187–192.
- [414] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.
- [415] H. Jhoti, G. Williams, D. C. Rees, C. W. Murray, *Nat. Rev. Drug Discov.* **2013**, *12*, 644–5.
- [416] P. Kirsch, A. M. Hartman, A. K. H. Hirsch, M. Empting, *Molecules* **2019**, *24*, DOI 10.3390/molecules24234309.
- [417] Z. Chilingaryan, Z. Yin, A. J. Oakley, *Int. J. Mol. Sci.* **2012**, *13*, 12857–79.
- [418] I. Navratilova, A. L. Hopkins, *ACS Med. Chem. Lett.* **2010**, *1*, 44–8.
- [419] M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, *J. Mol. Struct.* **2014**, *1077*, 101–113.
- [420] C. W. Murray, D. C. Rees, *Angew. Chemie Int. Ed.* **2016**, *55*, 488–492.
- [421] D. A. Erlanson, B. J. Davis, W. Jahnke, *Cell Chem. Biol.* **2019**, *26*, 9–15.
- [422] A. Bancet, C. Raingeval, T. Lomberget, M. Le Borgne, J.-F. Guichou, I. Krimm, *J. Med. Chem.* **2020**, acs.jmedchem.0c00242.
- [423] E. J. Strobel, L. Cheng, K. E. Berman, P. D. Carlson, J. B. Lucks, *Nat. Chem. Biol.* **2019**, *15*, 1067–1076.
- [424] C. P. Jones, A. R. Ferré-D'Amaré, *suppli_Nature Struct. Mol. Biol.* **2015**, *22*, 679–685.
- [425] G. Pagès, V. Gilard, R. Martino, M. Malet-Martino, *Analyst* **2017**, *142*, 3771–3796.
- [426] N. K. Tanner, *FEMS Microbiol. Rev.* **1999**, *23*, 257–275.

9. Research article I: The conformational landscape of transcription intermediates involved in the regulation of the ZMP-sensing riboswitch from *Thermosinus carboxydivorans*

Oliver Binas, Tatjana Schamber, Harald Schwalbe

Nucleic Acids Research, 2020, 48, 6970 – 6979

In this article, structural properties of transcriptional intermediates of a ZMP-sensing riboswitch are investigated. In *t. carboxydivorans*, a pfl-type transcriptional on-switch which senses ZMP, controls genes associated with purine synthesis and one-carbon metabolism. While the aptamer structure of these riboswitches and the general principles of transcriptional regulation were already studied, details of the structures attained during progression of transcription as well as kinetics were largely unknown at the beginning of the study. A study by Lucks et al., which was published shortly prior to publication shows results in the same field achieved by SHAPE (selective 2' hydroxyl acylation analyzed by primer extension) analysis and was discussed in the manuscript.^[423] In our approach, NMR spectroscopy was used to elucidate the underlying structural and functional properties. Resonance assignment was established by 2D-NOESY and a reporter peak for formation of the important pseudoknot structure was established. Structural characteristics of pseudoknot formation were investigated in dependence of ligand and magnesium concentration. Pseudoknot formation prior to ligand interaction could be confirmed unambiguously, inspired by an approach introduced by Ferré-D'Amaré et al. who confirmed riboswitch action independent of the linking sequence between the pseudoknot forming domains.^[424] Ten RNA sequences resembling structural intermediates of the riboswitch were prepared and investigated by NMR spectroscopy. Three RNAs showing ZMP interaction were further investigated by isothermal titration calorimetry (ITC) and NMR titration, in order to determine affinity and binding competent fractions.

All experiments including RNA preparation data acquisition and evaluation were performed by the author of the thesis, with the exception of the preparation and NMR screening of the ten transcriptional intermediates, which were performed by T. Schamber under supervision of the author. The author of the thesis wrote the manuscript together with H. Schwalbe.

The conformational landscape of transcription intermediates involved in the regulation of the ZMP-sensing riboswitch from *Thermosinus carboxydivorans*

Oliver Binas , Tatjana Schamber and Harald Schwalbe *

Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt/Main, Germany

Received April 07, 2020; Revised May 03, 2020; Editorial Decision May 06, 2020; Accepted May 29, 2020

ABSTRACT

Recently, prokaryotic riboswitches have been identified that regulate transcription in response to change of the concentration of secondary messengers. The ZMP (5-Aminoimidazole-4-carboxamide ribonucleotide (AICAR))-sensing riboswitch from *Thermosinus carboxydivorans* is a transcriptional ON-switch that is involved in purine and carbon-1 metabolic cycles. Its aptamer domain includes the *pfl* motif, which features a pseudoknot, impeding rho-independent terminator formation upon stabilization by ZMP interaction. We herein investigate the conformational landscape of transcriptional intermediates including the expression platform of this riboswitch and characterize the formation and unfolding of the important pseudoknot structure in the context of increasing length of RNA transcripts. NMR spectroscopic data show that even surprisingly short pre-terminator stems are able to disrupt ligand binding and thus metabolite sensing. We further show that the pseudoknot structure, a prerequisite for ligand binding, is preformed in transcription intermediates up to a certain length. Our results describe the conformational changes of 13 transcription intermediates of increasing length to delineate the change in structure as mRNA is elongated during transcription. We thus determine the length of the key transcription intermediate to which addition of a single nucleotide leads to a drastic drop in ZMP affinity.

INTRODUCTION

Riboswitches are cis-acting regulatory RNA elements, which sense the concentration variation of metabolites of low molecular weight. They are found in the 5'-untranslated

region of mRNA (1), primarily within prokaryotes (2). In general, riboswitches consist of a phylogenetically conserved aptamer domain that undergoes conformational change upon binding a specific ligand, which induces a further allosteric conformational change in a downstream expression platform (3). More than 20 different classes of riboswitches have been identified and have been found to sense metabolites including thiamin pyrophosphate (4,5), adenosylcobalamin (6,7), S-adenosylmethionine (8,9) and flavin mononucleotide (10). Typically, ligands sensed by riboswitches are cell metabolites but exceptions including riboswitches sensing fluoride (11) or Mg²⁺ (12) are known. Various riboswitches bind nucleobases such as adenine (13) and guanine (14) or the nucleoside deoxyguanosine (15). Often, riboswitches are found upstream of genes coding for proteins involved in the metabolism or catabolism of the ligand resulting in a feedback loop (16).

Transcriptional riboswitches modulate gene expression during transcription by controlling the formation of rho-independent terminator structures (17). Structural rearrangement of the aptamer domain upon binding leads to formation or destabilization of an antiterminator structure, competing with the terminator structure.

Despite the abundance of crystal structures of metabolite-bound aptamer domains, high-resolution structural information about the full range of transcriptional intermediates and their potentially heterogeneous conformations in the context of co-transcriptional folding are sparse. Switching efficiency between functional ON- and OFF-states in transcriptional riboswitches strongly depends on the time window during which the switch adopts a binding-competent form. In the *Bacillus subtilis* FMN riboswitch, transcriptional pausing is an important mechanism to achieve kinetic control (18). Time-resolved NMR experiments in our laboratory on two different purine-sensing riboswitches, the guanine-sensing riboswitch from *B. subtilis* and the 2'-deoxyguanosine-sensing riboswitch from *Mesoplasma florum* showed that these riboswitches

*To whom correspondence should be addressed. Tel: +49 69 798 29737; Fax: +49 69 798 29515; Email: schwalbe@nmr.uni-frankfurt.de

exhibit kinetic control of regulation and the required structural transition to ensure both, functional ON- and OFF-states, are matched to the time window available during transcript elongation (18,19).

The function of second messengers in cells is long-known (20), but their role in riboswitch-based regulation was not identified until the late 2000s when Breaker *et al.* showed that gene expression is regulated by riboswitches that bind cyclic di-GMP (21). Several other conserved second messenger-sensing motifs were identified in the following years, including the *pfl* motif which binds the cellular alarmone ZMP (5-Aminoimidazole-4-carboxamide ribonucleotide (AICAR)) (22).

The ZMP-sensing riboswitch was identified in 2015 and its aptamer fold was named *pfl* motif, as it is frequently associated with a gene coding for pyruvate-formate-lyase (23). As a precursor of inosine, ZMP is involved in the purine and folate biocycle where it is converted into FAICAR by addition of one carbon, provided by 10f-THF. In case of increasing folate stress, ZMP acts as an alarmone, upregulating the expression of associated proteins (24). Regulation is achieved by riboswitches containing the *pfl* motif, which specifically recognize ZMP and ZTP. High ZMP concentrations induce ZMP-binding to the riboswitch, which leads to antiterminator formation and subsequent activation of gene expression. The ZMP-switch is thus considered a transcriptional ON-Switch.

The *pfl* motif consists of three helical stems with P2 located in the loop of P1 (Figure 1A). Junction J1/2, located between those helices, forms a pseudoknot with the loop of P3, which features a complementary sequence. Upon ligand binding, the pseudoknot is further stabilized by stacking interactions with ZMP. From the crystal structures it is known that the aromatic ring of ZMP stacks right below the pseudoknot, where it interacts with a U residue of the loop capping P3 (25). Without ZMP, nucleotides within the P3 stem base-pair to downstream residues as transcription proceeds, forming a rho-independent terminator hairpin. The stabilized pseudoknot thus acts as antiterminator, preventing formation of the terminator hairpin (24).

The linker between P1 and P3 is not conserved in length and cannot be resolved in crystal structures due to structural heterogeneity. Studies by Ferré-d'Amaré *et al.* showed that removal of the linker does not abrogate ZMP binding, though a 10-fold lowered affinity is observed (26).

In a recent study, Lucks *et al.* applied SHAPE-seq (27) to the ZMP-sensing riboswitch from *Clostridium beijerinckii* and described its folding landscape in presence and absence of ZMP (28). They found that riboswitch function is ensured by a labile balance between the conserved short P3 stem, terminator hairpin and pseudoknot stability, which can be strongly perturbed by alteration of a single base-pair in one of those key-players.

In the present study, we characterize the conformations of transcriptional intermediates of the ZMP-sensing riboswitch from *T. carboxydivorans* under equilibrium conditions. These intermediates include the 48 nt containing construct Zsw⁴⁸ encompassing stems P1 and P2, the 77 nt construct Zsw⁷⁷ including stems P1, P2, and P3, ten transcriptional intermediates (Zsw⁸¹-Zsw⁹¹, Figure 1A) capable to switch between antiterminator and terminator conformations

as well as the full-length riboswitch Zsw¹⁰¹. We use NMR spectroscopy as outlined in Fuertig *et al.* (29) to determine the conformations of transcriptional intermediates under varying conditions, employing rapid sample preparation methodology developed by Helmling *et al.* (30) ensuring homogeneity in transcript length. Further, an *in-trans* RNA construct consisting of isolated P1–P2 (Zsw⁴⁸) and P3 (Zsw¹⁵) was investigated to monitor pseudoknot formation in *trans* (Figure 1B), utilizing differential isotopic labeling for the two RNA constructs for isotope-filtered NMR experiments.

We provide a detailed conformational characterization of the states that are relevant for regulation and that are sufficiently stable to be populated during transcription in order to extend simple two state models describing function of transcriptional ON-switches.

MATERIALS AND METHODS

Transcription templates

Transcription templates were generated by PCR from pUC57 vector that contains the full-length native sequence of the ZMP-sensing riboswitch from *T. carboxydivorans* or from previously PCR-amplified DNA. Reverse primers used to obtain templates for Zsw⁴⁸, Zsw⁷⁷ and Zsw⁸¹⁻⁹¹ are summarized in Table S1 (see Supplementary Material). Primers with high annealing temperatures exceeding 60 °C were chosen for Zsw¹⁰¹ as they led to significantly better PCR results. Primers for Zsw⁸¹⁻⁹¹ contained 2'-methoxy modifications to ensure 3'-end homogeneous transcription products (30). PCR was performed according to the standard protocol by New England Biolabs® (0.5 mM of each primer, 200 mM dNTPs) using home-made Phusion polymerase. As DNA template, we used either 0.1 µl of 1 µM plasmid solution containing the full-length riboswitch sequence or 0.1 µl of PCR-amplified template. Sample integrity was confirmed by native PAGE or agarose gel electrophoresis.

RNA preparation

All RNAs except Zsw¹⁵ were prepared by *in vitro* transcription with T7 RNA polymerase (RNAP). PCR-reactions were directly used for transcription without purification. 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, 12.5 mM Mg(OAc)₂, 0.2 u/ml yeast inorganic pyrophosphatase (New England Biolabs®) and 144 nM home-made T7 RNAP. Unlabeled NTPs were purchased from Carl Roth GmbH + Co. KG (Germany). ¹³C, ¹⁵N labeled NTPs were purchased from Silantes GmbH (Germany).

The transcripts Zsw⁸¹⁻⁹¹ were purified according to the protocol developed by Helmling *et al.* (30). After transcription in 10 ml scale, the RNA was washed on 5000 MWCO centrifugal concentrators (Vivaspin 20® from Sartorius AG, Germany). After loading the RNA on the centrifugal concentrator, the first two washing steps were performed with transcription buffer, to remove excess phosphate. Afterwards, the RNA was concentrated to 1 ml and washed at

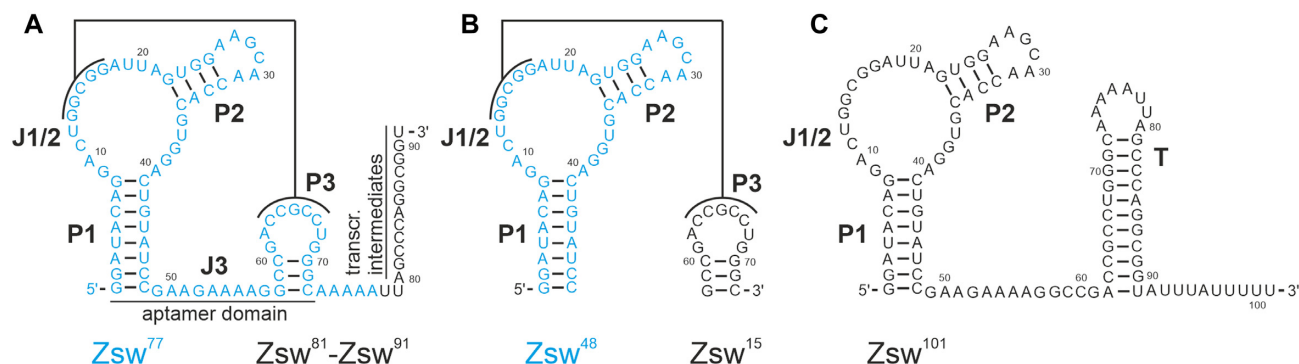


Figure 1. Overview of the secondary structures of the ZMP-sensing riboswitch from *T. carboxydivorans*. (A) Secondary structure of the ZMP-sensing riboswitch in pseudoknot form with the construct Zsw^{77} highlighted in blue. Investigated transcriptional intermediates are annotated accordingly. (B) Secondary structures of constructs Zsw^{48} and Zsw^{15} used for investigation of pseudoknot formation. (C) Secondary structure of Zsw^{101} , terminator fold of the ZMP-sensing riboswitch. Stems P1, P2, P3, the Junction J1/2 and the terminator hairpin T are annotated.

least 10 times with NMR buffer (25 mM potassium phosphate buffer, pH 6.2, 50 mM KCl) in 5 ml steps. Final NMR samples were concentrated to a volume of 250 μ l and concentrations ranging from 80 to 120 μ M. The RNA was thermally refolded before sample preparation by heating at 60°C for 5 min, subsequent addition of desired amount of magnesium chloride, heating at 40°C for 5 min and cooling on ice for 30 min.

The RNA constructs Zsw^{77} and Zsw^{48} were purified by denaturing PAGE. The transcription mixture was loaded on a 3000 MWCO centrifugal concentrator (Vivaspin 20®) from Sartorius AG, Germany) and washed repeatedly with water until the elute did not show a significant signal at 180–220 nm in the UV-VIS spectrum (NanoDrop One, ThermoFischer Scientific). The solution was concentrated to 1 ml, mixed with 30 % glycerol and loaded onto a 15% polyacrylamide (PAA) gel (7 M urea). The RNA band was excised from the gel and eluted in 0.6 M NaOAc, pH 5.5. The RNA was precipitated once with cold EtOH and twice with 2 % (w/v) LiClO₄ in acetone. After resuspending in water, buffer exchange to NMR buffer (as described above) was performed on a centrifugal concentrator.

Zsw^{15} RNA was bought from Dharmacon GmbH. The RNA was deprotected according to the supplier protocol. Following HPLC purification, the HPLC buffer was removed on a 1000 MWCO centrifugal concentrator. After one EtOH and two LiClO₄ precipitations (see above), the RNA was resuspended in water and the buffer exchanged to NMR buffer. The NMR buffer contained 25 mM potassium phosphate buffer adjusted to a pH of 6.2 and 50 mM KCl. The RNA was thermally refolded before sample preparation.

NMR spectroscopy

NMR samples were prepared by adding 10 % D₂O and 7.5 nmol DSS as internal reference to RNA stock solutions in NMR buffer (see above). All spectra were recorded of 280 μ l samples in Shigemi NMR tubes (Shigemi Inc.).

NMR experiments were conducted on Bruker AV600, AV700 and AV800 spectrometers, equipped with cryogenic probes. Data were processed with Bruker Topspin 3.5 (Bruker Biospin) and sparky 3.14 (31). Water suppres-

sion was achieved using WATERGATE (32) or jump-and-return echo (33) water suppression pulse schemes. ¹⁵N-editing in ¹H,¹H-NOESY was achieved by implementing X-filter schemes before and after t₁ chemical shift evolution (34). Analysis of dissociation constants and binding competent fraction were carried out by measuring intensities of the imino proton signal of U42 in ¹H-1D spectra, normalization of values and subsequent fitting while leaving the RNA concentration variable.

ITC measurements

RNA samples for ITC measurements were essentially prepared as for NMR spectroscopy. ITC measurements were performed on a Malvern® MicroCal iTC200 instrument with RNA sample concentrations of 60 μ M and ligand concentrations of 600 μ M (for Zsw^{77} and Zsw^{91}) or 300 μ M (for constructs Zsw^{81-83}). Raw data were exported and analyzed using NITPIC (35) and SEDPHAT (36). Fractions bound were obtained from leaving the n-value variable in the fitting procedure.

RESULTS

NMR resonance assignment

NMR signals of longer RNA constructs can often be dissected to arise from structural elements already present in shorter constructs following a divide-and-conquer-strategy (29). Since stems P1 and P2 are present in Zsw^{48} and in Zsw^{77} (see Figure 1A and b), comparable sets of imino signals are expected for these constructs and could in fact be observed in ¹H-1D as well as ¹H,¹H-2D NOESY experiments (Figure 2). By contrast, NMR signals at 10.5, 11.6 and 12.8 ppm, observable in Zsw^{15} , were also observed in Zsw^{77} but not in Zsw^{48} .

For ¹H proton chemical shift assignment, ¹H,¹H-NOESY experiments were conducted mapping out the interaction of imino protons with adjacent protons less than 5 Å apart from each other and led to the sequential assignment of G and U residues in P1 and P2 stems. Distinction between signals arising from G and U residues, respectively, was achieved by analysis of the cross-signals to the oppos-

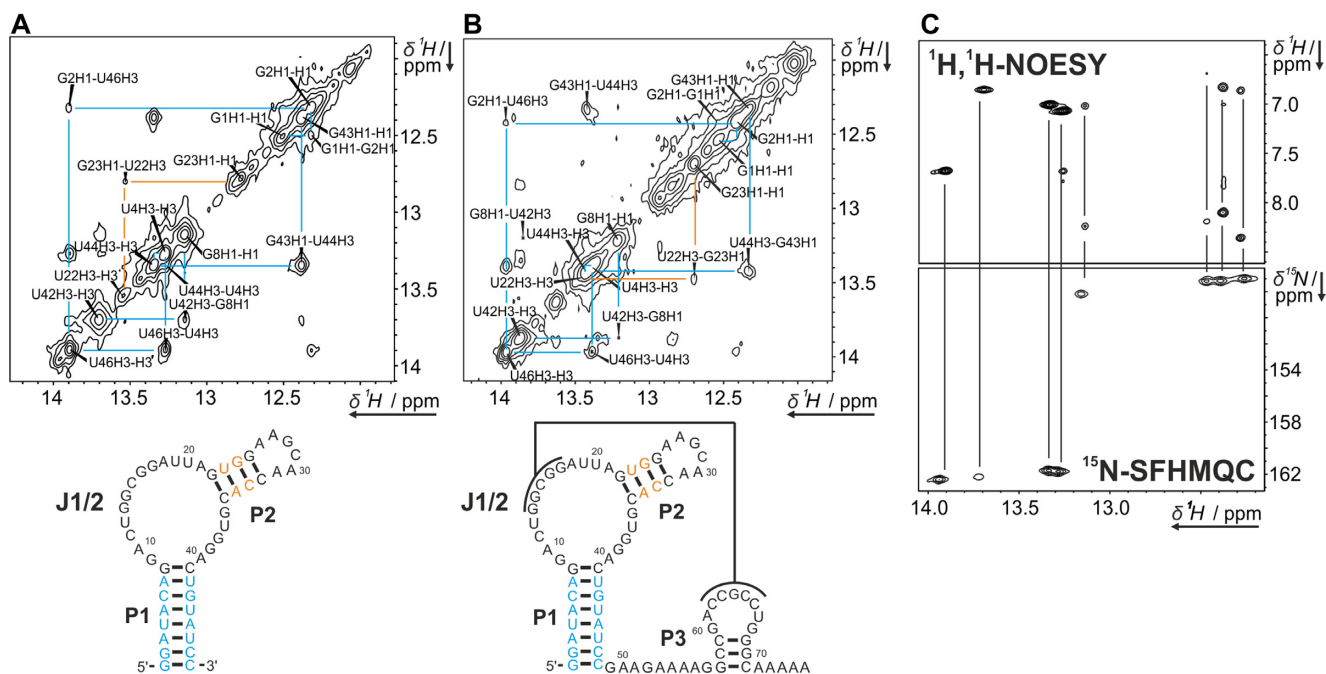


Figure 2. Imino regions of $^1\text{H}, ^1\text{H}$ -NOESY spectra of Zsw^{48} (A) and Zsw^{77} (B) annotated with assignment. Color-coding indicates resonances from either P1 (cyan) or P2 (orange). Data were measured at 600 MHz, 2048×640 points and 184 scans and 800 MHz, 2048×512 and 256 points for Zsw^{48} and Zsw^{77} , respectively. Samples contained $800 \mu\text{M}$ (Zsw^{48}) or $400 \mu\text{M}$ (Zsw^{77}) RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl and 10 % D_2O . Larger representations of the spectra are shown in Supplementary Figure S1 (C) Imino to aromatic region of the $^1\text{H}, ^1\text{H}$ -NOESY of Zsw^{48} (see A) and ^{15}N -SOFAST-HMQC (^{15}N -SFHMQC) of Zsw^{48} , showing ^{15}N chemical shift differences for G and U residues. ^{15}N -SOFAST-HMQC was measured at 700 MHz, room temperature, 2048×128 points and 16 scans on a sample containing $400 \mu\text{M}$ ^{13}C , ^{15}N -labeled RNA in 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl and 10 % D_2O . Subtle differences in chemical shift are a result of slight temperature differences between spectrometers.

ing base, which are two C amino protons for G-C base-pairs, but only one aromatic adenine H2 proton for A-U base pairs; and confirmed by $^1\text{H}, ^{15}\text{N}$ -correlation spectra on an isotopically labeled sample of Zsw^{48} (Figure 2C). Imino protons of the terminal helical base-pairs in P1 and P2 could not be observed due to solvent exchange. Since the P2 helix is not involved in ligand interaction or pseudoknot formation, we omitted further characterization of this stem. For helix P3 in Zsw^{77} and Zsw^{15} (Figure 1B), no sequential imino proton NOEs were observed, most likely due to the rapid solvent exchange in this very short stem. Subsequently, the assignment of Zsw^{48} was transferred to Zsw^{77} (Figure 2A and B). In summary, we established the conformations for both Zsw^{77} and Zsw^{48} . Further imino proton signals were observed at 12.1 and 12.7 ppm for Zsw^{48} and an additional signal at 12.3 ppm for Zsw^{77} . These signals likely arise from residues located in the non-helical regions between P1 and P2, since they show no NOESY cross peaks to other imino proton signals.

Investigation of pseudoknot interaction via *in-trans* construct

To investigate the formation of the pseudoknot between J1/2 and P3, which is the essential interaction stabilizing the antiterminator conformation, we designed two RNA constructs Zsw^{48} and Zsw^{15} . Zsw^{48} consisted of the P1 and P2 stem including the junction J1/2, and Zsw^{15} consisted of the P3 helix with the corresponding loop. To assess whether these two RNA strands reconstruct the aptamer domain

in-trans, we added one equivalent of Zsw^{15} to Zsw^{48} and monitored chemical shifts of imino proton signals in the presence of 10 mM Mg^{2+} and 2 eq. ZMP (Figure 3A). Imino proton spectra of the *in-trans* construct $Zsw^{48} + Zsw^{15}$ showed no difference to spectra of Zsw^{77} , confirming that the pseudoknot formation does not require the linker J3. However, spectra also showed that ZMP affinity is higher for the *in-cis* construct Zsw^{77} than for the *in-trans* construct $Zsw^{48} + Zsw^{15}$. The linewidth of reporter signal U42 was larger for $Zsw^{48} + Zsw^{15}$, indicating conformational heterogeneity involving a residual unbound conformation, while U42 in Zsw^{77} showed a significant chemical shift perturbation (CSP) of 0.06 ppm. Strikingly, however, $Zsw^{48} + Zsw^{15}$ showed clear ligand binding (Figure 3A).

We prepared $^{13}\text{C}, ^{15}\text{N}$ -isotope labeled Zsw^{48} and added unlabeled Zsw^{15} to further characterize the pseudoknot interaction. We recorded ^{15}N -edited ^1H -1D spectra as well as conventional ^1H -1D spectra of all constructs (Figure 3B upper and middle spectra, respectively). In ^{15}N -edited spectra, peaks of only those ^1H can be observed that are covalently bound to ^{15}N , which includes only signals of residues in Zsw^{48} . In the ^1H -1D spectrum, several signals were observed which could not be attributed to Zsw^{48} since they did appear only in conventional ^1H -1D-, but not in ^{15}N -edited spectra. Of these, signals at 11.7 ppm and 12.7 ppm (Figure 3B, black arrows) were assigned to the stem imino protons of Zsw^{15} , as they closely resemble the signals observed in pure Zsw^{15} (Figure 3B lower spectrum). Signals at 12 ppm and 12.8 ppm belonged to Watson-Crick type

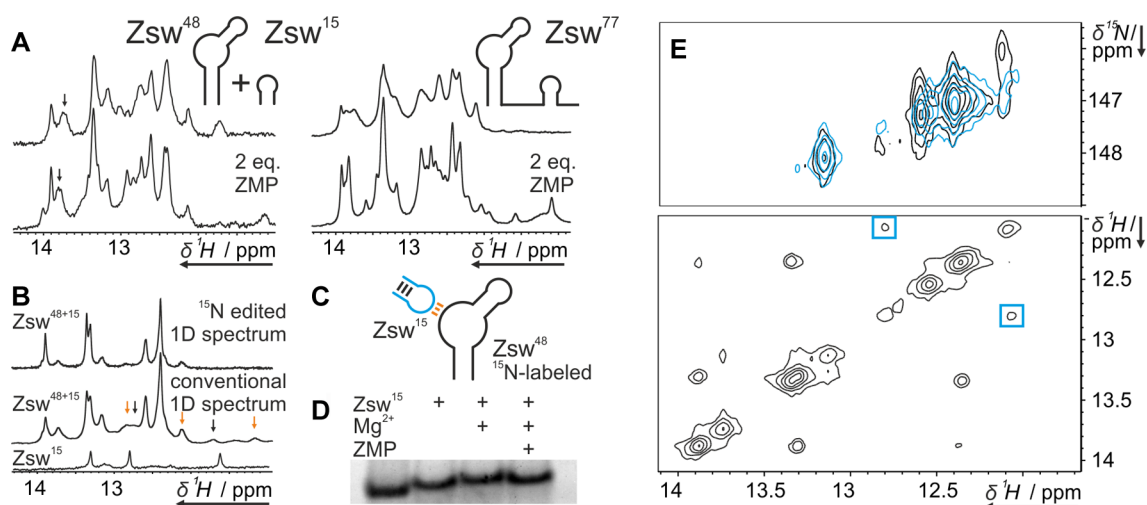


Figure 3. (A) ^1H -1D NMR spectra of the *in-trans* construct $\text{Zsw}^{48}+\text{Zsw}^{15}$ and Zsw^{77} before and after addition of 2 eq. ZMP in presence of 10 mM Mg^{2+} . (B) ^{15}N -edited and conventional ^1H -1D NMR spectra of ^{15}N - $\text{Zsw}^{48}+^{14}\text{N}$ - Zsw^{15} and conventional ^1H -1D of Zsw^{15} in the presence of 10 mM Mg^{2+} . Peaks that arise only in the conventional spectrum marked with arrows. Samples contained 100 μM RNA. Stacked and enlarged representations of those spectra are shown in Supplementary Figure S2. (C) Schematic representation of the labelling scheme of ^{15}N - $\text{Zsw}^{48}+^{14}\text{N}$ - Zsw^{15} . Colors of base-pairs according to arrow colors in (B). (D) 12% native PAGE of Zsw^{48} in the presence of 1 eq Zsw^{15} , 10 mM Mg^{2+} and 1 eq. ZMP. Pockets were loaded with 200 nmol Zsw^{48} and additions as indicated. The gel is displayed enlarged in Supplementary Figure S3 (E) Top: ^{15}N -SOFAST-HMQC spectrum (37) of Zsw^{48} before (cyan) and after (black) addition of Zsw^{15} in the presence of 10 mM Mg^{2+} . Spectra were recorded at 800 MHz at room temperature with 2048×128 points and 8 scans. Bottom: ^{15}N -edited ^1H , ^1H -NOESY of $\text{Zsw}^{48}+\text{Zsw}^{15}$ with the cross peaks that arise from addition of Zsw^{15} marked blue. Spectra were recorded at 800 MHz with 2048×432 points and 256 scans.

base-pairs between Zsw^{48} and Zsw^{15} indicative of pseudoknot formation (Figure 3B, orange arrows). Following up on this observation, we compared ^1H , ^{15}N -correlation spectra before and after addition of Zsw^{15} and observed an additional imino resonance at 12 ppm obscured by another ^1H resonance, which had not been detected in 1D experiments and only appears upon addition of Zsw^{15} and therefore pseudoknot formation. A cross peak from this resonance to 12.8 ppm was observed in the ^{15}N -edited ^1H , ^1H -NOESY of ^{15}N - $\text{Zsw}^{48}+^{14}\text{N}$ - Zsw^{15} (Figure 3E) proving the formation of two consecutive Watson-crick type G-C base-pairs of the pseudoknot.

These NMR results were supported by native PAGE. By native PAGE (Figure 3D), we detected a slower migration speed for $\text{Zsw}^{48}+\text{Zsw}^{15}$ in comparison to pure Zsw^{48} even without addition of Mg^{2+} . Adding Mg^{2+} slightly reduced migration speed, while addition of ligand slightly enhanced migration speed. We attribute the slower migration speed with Mg^{2+} to alteration of ion strength and the slight enhancement upon ligand interaction to a slight compaction of the structure upon pseudoknot stabilization.

Probing ZMP-binding to transcriptional intermediates of increasing length

We investigated binding of ZMP to potentially critical transcriptional intermediates of the *pfl* riboswitch by NMR spectroscopy. In the crystal structure of the riboswitch (25), a Mg^{2+} -ion is detected in close proximity to the binding pocket and ITC studies by Patel *et al.* suggested strong dependence of ZMP affinity on Mg^{2+} concentration (25). Thus, we assumed structural changes to take place upon addition not only of ZMP, but also of Mg^{2+} , which were detectable as CSP on imino proton resonances. For two sam-

ples of Zsw^{77} , we added either 3 mM or 10 mM of MgCl_2 ($[\text{Mg}^{2+}]:[\text{RNA}] = 15/50:1$), respectively, before adding two equivalents ($[\text{RNA}]:[\text{ZMP}] = 1:2$) of ligand. In both cases, CSPs to the downfield region were observed for all proton signals after addition of Mg^{2+} . However, these shifts were accompanied by signal broadening at 10 mM Mg^{2+} . After adding two equivalents of ligand, signal line widths decreased and the signals shifted non-uniformly for the 10 mM Mg^{2+} sample. In contrast, the 3 mM Mg^{2+} sample showed no spectral changes upon ZMP addition (Figure 4A). This confirms the observations previously reported (26). The Mg^{2+} -induced increase in signal line widths was reversed by the addition of ZMP. The imino proton resonance of U42 represents a potent probe for ligand binding, as it is well separated from the other signals in the spectrum and close enough to the ligand binding pocket to show significant CSPs upon interaction. In the pure RNA sample, it was observed as a strong signal at 13.77 ppm (Figure 4B black spectra). However, upon addition of Mg^{2+} , an additional signal appeared at 13.9 ppm with the two populations in an estimated 1:1 ratio (Figure 4B orange spectrum). Addition of ligand then led to loss of the signal at 13.77 ppm and increase of the signal at 13.9 ppm accompanied by a small downfield shift (Figure 4B cyan spectra). These effects might be attributed to the Mg^{2+} -dependent pre-formation of a conformationally flexible binding-competent conformational state, which is subsequently stabilized by ligand interaction. When Mg^{2+} was added to Zsw^{48} no second conformation or line broadening was observed (Supplementary Figure S4). For Zsw^{101} , no response to ligand or magnesium could be detected (Supplementary Figure S5).

We investigated transcriptional intermediates between the fully binding-competent Zsw^{77} and the binding-incompetent Zsw^{101} RNA for ligand binding. Ligand bind-

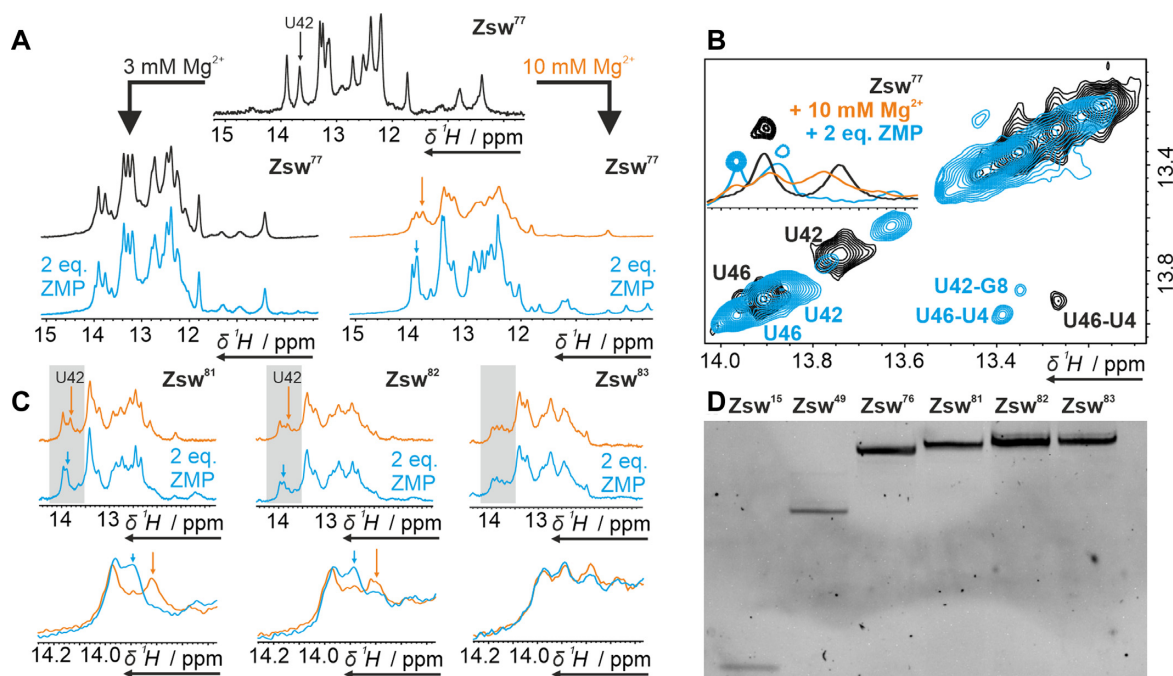


Figure 4. (A) ^1H -1D NMR spectra of Zsw^{77} before and after addition of Mg^{2+} and subsequent addition of 2 eq. ZMP at room temperature. Samples contained $100\ \mu\text{M}$ RNA. (B) NOESY and ^1H -1D data showing an overlay of the Ligand and Mg^{2+} free (black), Mg^{2+} bound (orange, only 1D) and ligand bound (cyan) state of Zsw^{77} at 288K. (C) ^1H -1D NMR spectra of transcripts $\text{Zsw}^{81,82,83}$ under influence of $10\ \text{mM}$ Mg^{2+} before and after addition of 2 eq. ZMP. U42, which is indicative for binding, is marked with an arrow. ZMP binding is visible for Zsw^{81} and Zsw^{82} while no binding is observed for Zsw^{83} . Samples contained $120\ \mu\text{M}$ RNA. Zooms of the regions marked grey in top spectra are displayed in bottom spectra. (D) Denaturing PAGE of RNAs Zsw^{15} , Zsw^{49} , Zsw^{76} and Zsw^{81-83} . Gel contained 12% PAA and 7 M urea. The full gel is shown in Supplementary Figure S8.

ing is linked to pseudoknot formation, which competes with terminator stem formation at increasing transcript length. From investigation of the secondary structure we could deduce that the shortest potential construct, to form a closing base-pair of a terminator stem-loop is Zsw^{81} , and no additional stabilization for the terminator conformation can be expected beyond Zsw^{91} . Accordingly, titration experiments were performed on constructs of lengths between 81 nt and 91 nt, at $10\ \text{mM}$ Mg^{2+} and up to 2 eq. ZMP. The constructs were prepared by shifting the reverse primer 3'-end from 5' to 3' of the template, achieving different lengths of template DNA for *in vitro* transcription by PCR. *In vitro* transcriptions according to the protocol of Helmling *et al.* (30) allowed quick and homogeneous RNA preparation as shown by denaturing PAGE (Figure 4D and Supplementary Figure S6).

NMR titration (Figure 4C) of Zsw^{81} showed similar CSPs as for Zsw^{77} . In the ZMP-free form, under influence of Mg^{2+} , a structural equilibrium was observed with U42 signals at 13.77 and 13.9 ppm as in spectra of Zsw^{77} . As opposed to the 1:1 ratio observed in Zsw^{77} , the binding-competent state was less populated in Zsw^{81} , observed from the lower intensity of the U42 signal at 13.9 ppm. Lucks *et al.* showed that the addition of a nucleotide enabling the formation of a one base-pair terminator stem partially destabilizes the antiterminator conformation (28). This effect would explain the spectral alterations observed in the ZMP-free samples of Zsw^{77} and Zsw^{81} . However, ligand binding was clearly detectable by chemical shift change of U42. In contrast to the spectrum of ligand-bound

Zsw^{77} , considerable line-broadening was observed even in the bound state, which suggests population of two or multiple states with interconversion kinetics in the intermediate time regime. For Zsw^{82} , the effect of ligand addition was already minor compared to Zsw^{81} , observed as $\sim 50\%$ lower differences in peak intensities. Alongside, we observed increased spectral complexity and line-broadening. Zsw^{83} , able to form 3 base-pairs of the terminator hairpin, showed a further shift of the conformational ensemble towards the terminator conformation. Adding ZMP did not lead to any significant structural alteration anymore, which indicates that the terminator conformation was already stable enough to fully outcompete the ligand-binding competent antiterminator conformation. This was also observed for ZMP titration to constructs Zsw^{84-90} , which did not show any spectral changes upon ligand addition (Supplementary Figure S7).

We performed ITC measurements (Figure 5A) on the RNA constructs which showed ligand binding according to NMR titration experiments to determine the equilibrium binding constants. In most cases, riboswitch aptamers exhibit a 1:1 stoichiometry for ligand binding. However, conformational equilibria between binding competent and binding incompetent folds, as observed by NMR, decrease the effective concentration of ligand-binding competent RNA conformations. Accordingly, the data were fitted to accommodate to possible binding incompetent fractions by leaving the stoichiometric parameter variable. Thus, not only the dissociation constants, but also the relative binding competent fractions of the RNA could be obtained.

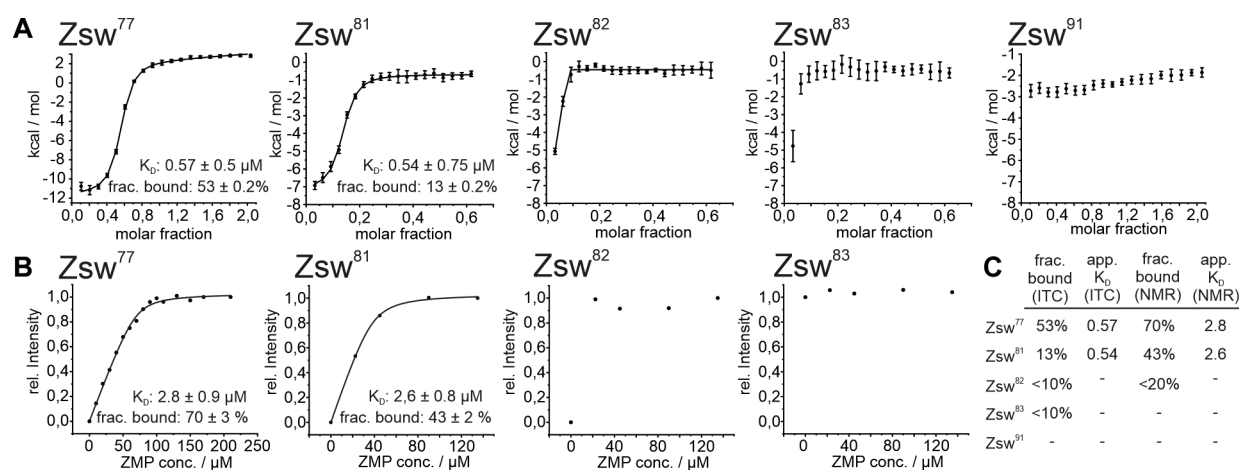


Figure 5. (A) ITC titration curves of ZMP to constructs Zsw⁷⁷, Zsw^{81–83} and Zsw⁹¹. Dissociation constants determined by the fit are annotated as applicable. (B) NMR titration curves of ZMP to constructs Zsw⁷⁷, Zsw^{81–83} and Zsw⁹¹. (C) Table summarizing the results of the titration experiments. Values for Zsw⁸² and Zsw⁸³ were estimated from the curves displayed. Errors, obtained from the fitting procedures are displayed with the respective curve.

Dissociation constants for Zsw⁷⁷ and Zsw⁸¹ were 0.57 and 0.54 μM respectively, which is in good agreement with 1 μM obtained by Ferré-d'Amaré *et al.* for the *Fusobacterium ulcerans* ZMP-sensing riboswitch (26). Dissociation constants obtained from NMR measurements (Figure 5B) were slightly higher with 2.78 μM for Zsw⁷⁷ and 2.58 μM for Zsw⁸¹. While the affinity remains constant throughout the two different RNA constructs, the amount of binding competent RNA largely differs. By ITC, the binding-competent fractions are 53 % (Zsw⁷⁷) and 13% (Zsw⁸¹), while in NMR measurements, a decrease from 70% to 43% was observed (Figure 5C). However, the degree of decreased binding differs largely between the two methods ITC and NMR-spectroscopy with a relative decrease of roughly 75% in ITC and only 40% in NMR measurements from Zsw⁷⁷ to Zsw⁸¹. A possible explanation could be the slow time-scale of equilibration between measurements. While ZMP was added with few minutes of waiting time between injections in ITC, NMR experiments take around a factor of 10 longer. For Zsw⁸², only a very minor binding competent population was observable both in ITC and NMR titrations, which prevented curve fitting and thus the determination of a dissociation constant or the exact population of states. However, we estimated from curve shapes of constructs longer than 81 nt that the amount of binding competent RNA must be below 10 % in ITC measurements, up to Zsw⁹¹, which showed no interaction at all.

DISCUSSION

In this study, we explored the folding landscape of the ZMP-sensing riboswitch from *T. carboxydivorans* by NMR spectroscopy and ligand binding affinities by ITC. From ¹H,¹H-NOESY experiments imino proton assignment of Zsw⁴⁸ could be achieved and transferred to the longer construct Zsw⁷⁷, enabling the observation of ligand binding on the well separated resonance of U42. We additionally could confirm pre-formation of the pseudoknot and binding competence of an *in-trans* construct in which the linker between P1 and P3 was deleted. We observed only two G-residues of

the possible four G–C base-pairs in the pseudoknot, suggesting a lower stability at one edge of the pseudoknot helix. Taking the crystal structure (25) into account, it is highly likely that G16–C62 and G17–C61 are less stable than G14–C65 and G15–C64 base-pairs, since G14 sits directly adjacent to the highly structured binding-pocket, while G17–C61 constitutes the closing base-pair of the pseudoknot helix. As reported previously (25), binding was only achieved at a high Mg²⁺ concentration of 10 mM, where substantial line broadening of NMR signals was observed, while the initial linewidth was regained after ZMP addition. Upon addition of 10 mM Mg²⁺, we additionally observed a structural equilibrium between two states, in which one state showed reporter peak shifts similar to the ligand-bound and the other one similar to the free state. Upon ligand interaction the system collapsed into the ligand bound state with minimal chemical shift alteration compared to the Mg²⁺-bound state. We therefore propose an additional Mg²⁺-dependent pre-formation of the binding pocket, which goes beyond just pseudoknot-formation, since pseudoknot formation is already observed without Mg²⁺ addition. However, pseudoknot formation is required for this additional Mg²⁺-stabilized state, since Zsw⁴⁸ does not show any spectral alteration upon Mg²⁺ addition, in the absence of Zsw¹⁵. The Mg²⁺ binding step can thus be considered an intermediate step, taking place after pseudoknot formation and facilitating ligand interaction and therefore final stabilization of the binding pocket.

Longer constructs showed only slight alteration in the population of these states in the presence of Mg²⁺. However, the amount of RNA which refolded to the ligand-bound state upon ZMP addition plummeted between Zsw⁸¹ and Zsw⁸², visible from quantitative NMR and ITC data. This further indicated that the existence of a Mg²⁺-induced pre-bound state is decoupled from the pseudoknot-terminator equilibrium, but both must be present to allow ligand binding. Lucks *et al.* reported that already the addition of a single nucleotide, competing with P3 helix stability (i.e. Zsw⁸¹), leads to partial ZMP binding incompetence, which we also observe via ITC and NMR spectroscopy. However, for this

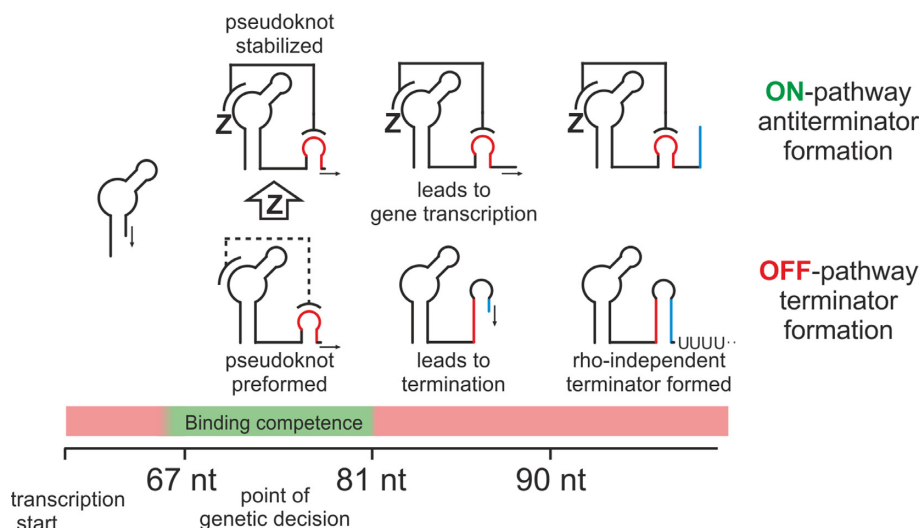


Figure 6. Schematic representation of the model for genetic control employed by the ZMP-sensing riboswitch from *T. carboxydivorans*. The point of genetic decision during transcription of nucleotides 67–81 is characterized by binding competence. At this stage, the pseudoknot is preformed. Binding of the cognate ligand ZMP leads to stabilization of the pseudoknot bearing antiterminator structure and subsequent gene transcription. No binding leads to formation of the terminator structure and rho-independent transcription termination.

construct, absolute values differed largely between ITC and much slower NMR measurements. We therefore suspect very slow folding kinetics on a timescale of minutes between a binding incompetent pre-terminator state back to a binding competent state if binding competent RNA is removed from the equilibrium by ligand interaction. This would render the pre-terminator state kinetically trapped in a biological context. Consequently, it is expected that RNA displaying formation of a single terminator base-pair could already be predominantly binding incompetent *in vivo*, where fast progress of transcription limits the refolding time available. In comparison, Zsw⁸² displays strongly reduced binding competence and an almost full population of the terminator state, leaving Zsw⁸¹ as the only construct with considerable population of both states. Therefore, in a biological context, the transition to the binding incompetent state is expected to be comparably sharp in contrast to other examples like the deoxyguanosine-sensing riboswitch, which displays a smoother transition (19). The high stability of short pre-terminator constructs leads to a very short window for ligand interaction during transcription of the riboswitch, which is only possible while RNA of lengths 74–76 to 81 nt is accessible to refolding and therefore ligand interaction. Linker length between P3 and terminator stem could therefore be an important feature that tunes riboswitch activity over genetic occurrences as in other prokaryotes, where this length strongly varies (25). It should be noted, that in terms of co-transcriptional folding, 14 nucleotides, starting from the active side are involved in the transcription elongation complex (TEC), with 9 nt hybridized to DNA and 5 nt located in the exit channel (38). In principle 2 to 3 more residues could be involved in RNA folding and removed from the DNA:RNA duplex without the TEC dissociating (39), leaving 7 to 8 residues in the exit channel with potential for interaction. While these residues are unable to interact in tertiary interactions with further upstream residues, as in the pseudoknot structure, RNA secondary structure forma-

tion inside the exit channel is observed in some SHAPE-seq (40) and smFRET (41) experiments for specific constructs. The assessment of the ability of the Zsw terminator structure to form in the exit channel of the *T. carboxydivorans* RNAP is beyond the scope of this project. Therefore, nucleotide numbers mentioned here, refer to residues accessible to folding in a manner comparable to *in vitro* folding, so most likely outside the RNAP exit channel. Consequently, during folding of the first pre-terminator base-pair at position 81, the RNAP active site will reside at position 91–93, the position of the terminator poly-U stretch. Therefore, transcriptional speed might be additionally modulated by transcriptional pausing, a well studied effect in transcriptional regulation and action of rho-independent terminators (18,42).

We propose a model (Figure 6) in which a pre-formed pseudoknot is further stabilized by ligand interaction, leading to anti-termination of the riboswitch, which is in agreement with the results Lucks *et al.* obtained using SHAPE-seq (28). The time-window of the binding competent conformation is, due to the very stable pre-terminator Zsw⁸², comparably short, featuring structural equilibria only at a transcript length of 81 nt. This implies that for biological function to be carried out correctly, the binding event must be considerably fast compared to transcript elongation (possibly modulated by trans acting factors such as Nus proteins), which could be an interesting premise for further kinetic experiments. With the terminator fold being thermodynamically more stable than the ligand-bound antiterminator fold, the stabilization of the pseudoknot established by ZMP addition can be considered temporary and therefore kinetically driven. This type of regulation is a common motif observed in riboswitches such as fluoride- (43), FMN- (44), deoxyguanosine- (19) or guanosine-sensing (45) riboswitches. Biological function of the riboswitch is therefore expected to be heavily tuned by terminator folding kinetics with and without the influence of ZMP.

To resolve these kinetics, real-time NMR measurements can be performed by rapid addition of interactors or by introduction and subsequent elimination of photolabile protecting groups. However, these studies require sufficient exploration of the underlying system in equilibrium state. We herein provided the NMR spectroscopic basis for such studies, showing important key players of the conformational landscape of the *pfl* riboswitch from *T. carboxydivorans* and their structural characteristics at atomic resolution.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank Dr B. Fürtig, Prof Dr J. Wöhnert, Dr A. Wacker, A. Völklein M.Sc., and T. Landgraf M.Sc. for fruitful discussion. They further wish to thank Dr C. Richter for help and maintenance of NMR spectrometers.

FUNDING

DFG priority programme [SPP 1879]; collaborative research center [SFB902]; B.M.R.Z. is supported by the state of Hesse. Funding for open access charge: institutional funding.

Conflict of interest statement. None declared.

REFERENCES

- Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
- McCown, P.J., Corbino, K.A., Stav, S., Sherlock, M.E. and Breaker, R.R. (2017) Riboswitch diversity and distribution. *RNA*, **23**, 995–1011.
- Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
- Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
- Edwards, T.E. and Ferré-D'Amaré, A.R. (2006) Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA–Small molecule recognition. *Structure*, **14**, 1459–1468.
- Nahvi, A., Barrick, J.E. and Breaker, R.R. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.*, **32**, 143–150.
- Peselis, A. and Serganov, A. (2012) Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat. Struct. Mol. Biol.*, **19**, 1182–1184.
- Grundy, F.J. and Henkin, T.M. (1998) The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.*, **30**, 737–749.
- Montange, R.K. and Batey, R.T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, **441**, 1172–1175.
- Serganov, A., Huang, L. and Patel, D.J. (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature*, **458**, 233–237.
- Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B. and Breaker, R.R. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*, **335**, 233–235.
- Ramesh, A. and Winkler, W.C. (2010) Magnesium-sensing riboswitches in bacteria. *RNA Biol.*, **7**, 77–83.
- Lemay, J.F. and LaFontaine, D.A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA*, **13**, 339–350.
- Batey, R.T., Gilbert, S.D. and Montange, R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
- Kim, J.N., Roth, A. and Breaker, R.R. (2007) Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 16092–16097.
- Sherwood, A. V. and Henkin, T.M. (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.*, **70**, 361–374.
- Farnham, P.J. and Platt, T. (1981) Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Res.*, **9**, 563–577.
- Steinert, H., Sochor, F., Wacker, A., Buck, J., Helmling, C., Hiller, F., Keyhani, S., Noeske, J., Grimm, S., Rudolph, M.M. *et al.* (2017) Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *Elife*, **6**, e21297.
- Helmling, C., Klötzner, D.-P., Sochor, F., Mooney, R.A., Wacker, A., Landick, R., Fürtig, B., Heckel, A. and Schwalbe, H. (2018) Life times of metastable states guide regulatory signaling in transcriptional riboswitches. *Nat. Commun.*, **9**, 944.
- Pollard, T.D., Thomas, D., Earnshaw, W.C., Lippincott-Schwartz, J. and Johnson, G.T. (2017) Second Messengers. *Cell biology*, Elsevier Inc. pp. 443–462.
- Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H. and Breaker, R.R. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science*, **321**, 411–413.
- 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.
- Atteia, A., Van Lis, R., Gelius-Dietrich, G., Adrait, A., Garin, J., Joyard, J., Rolland, N. and Martin, W. (2006) Pyruvate formate-lyase and a novel route of eukaryotic ATP synthesis in *Chlamydomonas* mitochondria. *J. Biol. Chem.*, **281**, 9909–9918.
- Biosynthesis, P., Metabolism, O., Kim, P.B., Nelson, J.W., Ronald, R., Kim, P.B., Nelson, J.W. and Breaker, R.R. (2015) An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell*, **57**, 317–328.
- Ren, A., Rajashankar, K.R. and Patel, D.J. (2015) Global RNA fold and molecular recognition for a *pfl* riboswitch bound to ZMP, a Master regulator of one-carbon metabolism. *Structure*, **23**, 1375–1381.
- Jones, C.P. and Ferré-D'Amaré, A.R. (2015) Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Suppl. Nat. Struct. Mol. Biol.*, **22**, 679–685.
- Watters, K.E. and Lucks, J.B. (2016) Mapping RNA structure in vitro with SHAPE chemistry and next-generation sequencing (SHAPE-Seq). In: *Methods in Molecular Biology*. Humana Press Inc., Vol. **1490**, pp. 135–162.
- Strobel, E.J., Cheng, L., Berman, K.E., Carlson, P.D. and Lucks, J.B. (2019) A ligand-gated strand displacement mechanism for ZTP riboswitch transcription control. *Nat. Chem. Biol.*, **15**, 1067–1076.
- Fürtig, B., Richter, C., Wöhnert, J. and Schwalbe, H. (2003) NMR spectroscopy of RNA. *ChemBioChem*, **4**, 936–962.
- Helmling, C., Keyhani, S., Sochor, F., Fürtig, B., Hengesbach, M. and Schwalbe, H. (2015) Rapid NMR screening of RNA secondary structure and binding. *J. Biomol. NMR*, **63**, 67–76.
- Lee, W., Tonelli, M. and Markley, J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
- Liu, M., Mao, X., Ye, C., Huang, H., Nicholson, J.K. and Lindon, J.C. (1998) Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J. Magn. Reson.*, **132**, 125–129.
- Marius Clore, G., Kimber, B.J. and Gronenborn, A.M. (1983) The 1-1 hard pulse: a simple and effective method of water resonance suppression in FT 1H NMR. *J. Magn. Reson.*, **54**, 170–173.
- Ikura, M. and Bax, A. (1992) Isotope-filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.*, **114**, 2433–2440.

35. Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C.A. and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, **84**, 5066–5073.
36. Zhao, H., Piszczek, G. and Schuck, P. (2015) SEDPHAT - a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods*, **76**, 137–148.
37. Schanda, P. and Brutscher, B. (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.*, **127**, 8014–8015.
38. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. and Darst, S.A. (2000) A structural model of transcription elongation. *Science*, **289**, 619–625.
39. Komissarova, N., Becker, J., Solter, S., Kireeva, M. and Kashlev, M. (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol. Cell*, **10**, 1151–1162.
40. Watters, K.E., Strobel, E.J., Yu, A.M., Lis, J.T. and Lucks, J.B. (2016) Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.*, **23**, 1124–1131.
41. Widom, J.R., Nedialkov, Y.A., Rai, V., Hayes, R.L., Brooks, C.L., Artsimovitch, I. and Walter, N.G. (2018) Ligand modulates cross-coupling between riboswitch folding and transcriptional pausing. *Mol. Cell*, **72**, 541–552.
42. Gromak, N., West, S. and Proudfoot, N.J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell Biol.*, **26**, 3986–3996.
43. Zhao, B., Guffy, S.L., Williams, B. and Zhang, Q. (2017) An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.*, **13**, 968–974.
44. Wickiser, J.K., Winkler, W.C., Breaker, R.R. and Crothers, D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell*, **18**, 49–60.
45. Buck, J., Fürtig, B., Noeske, J., Wönert, J. and Schwalbe, H. (2007) Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 15699–15704.

Supplementary data

The conformational landscape of the transcription intermediates of the ZMP-sensing riboswitch from *Thermosinus carboxydivorans*

Oliver Binas,^[a] Tatjana Schamber^[a] and Harald Schwalbe^{*[a]}

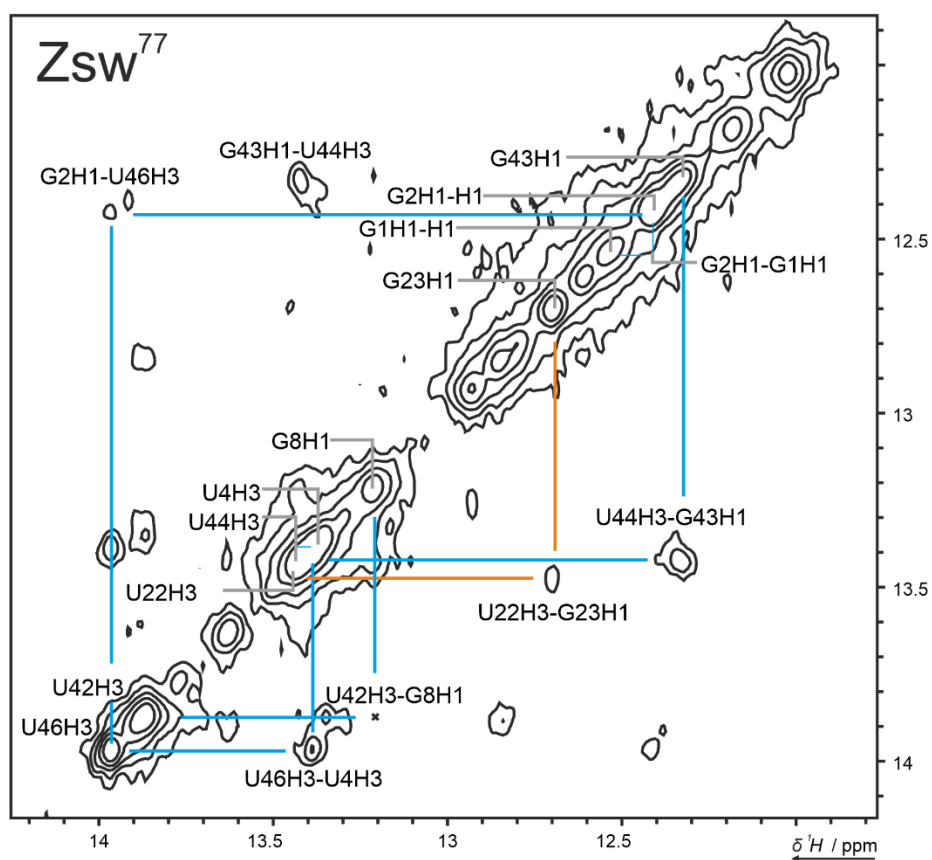
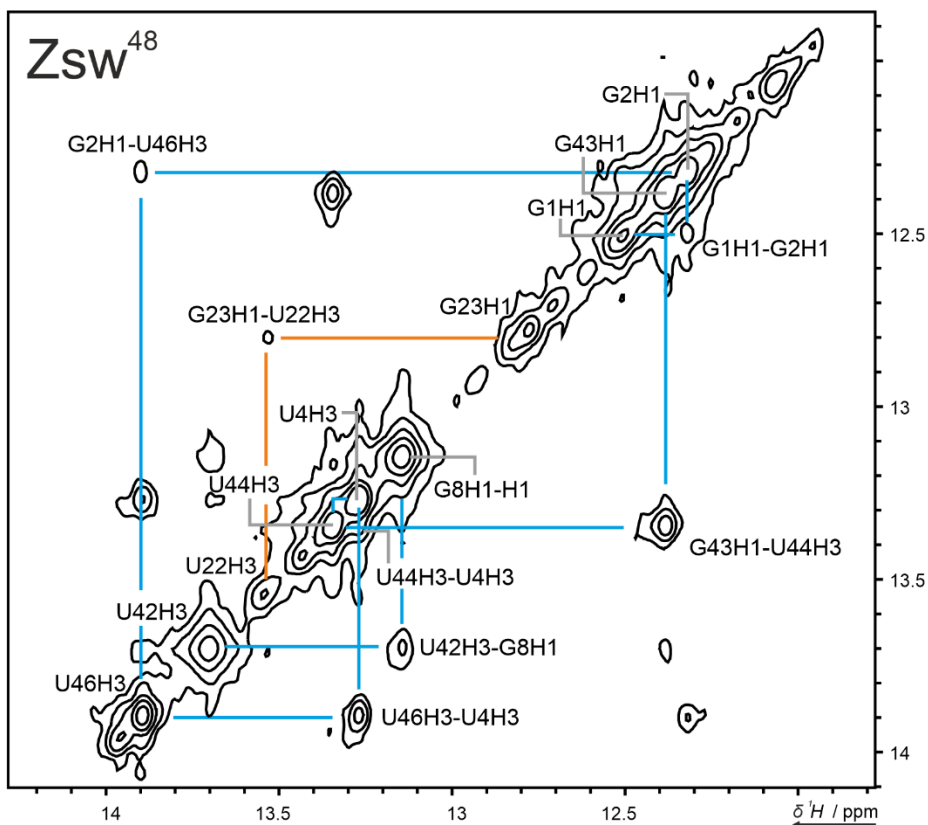
Table of Contents:

- **Supplementary Table S1.** Summary of all forward and reverse DNA primers used in this study.
- **Supplementary Figure S1.** Imino regions of ¹H,¹H-NOESY spectra of Zsw⁴⁸ and Zsw⁷⁷.
- **Supplementary Figure S2.** ¹⁵N-edited and conventional ¹H-1D NMR spectra of ¹⁵N-Zsw⁴⁸+¹⁴N-Zsw¹⁵ and conventional ¹H-1D of Zsw¹⁵ in the presence of 10 mM Mg²⁺.
- **Supplementary Figure S3.** 12 % native PAGE of Zsw⁴⁸ in the presence of 1 eq Zsw¹⁵, 10 mM Mg²⁺ and 1 eq. ZMP.
- **Supplementary Figure S4.** ¹H-1D spectra of Zsw⁴⁸ in presence and absence of 10 mM Mg²⁺ and 1 eq Zsw¹⁵.
- **Supplementary Figure S5.** ¹H-1D spectra of Zsw¹⁰¹ in presence and absence of 1 eq ZMP.
- **Supplementary Figure S6.** 12 % native PAGE of Zsw⁸¹⁻⁹¹.
- **Supplementary Figure S7.** ¹H-1D spectra of Zsw⁸¹⁻⁹¹ in presence and absence of 1 or 2 eq. ZMP.
- **Supplementary Figure S8.** 12% denaturing PAGE of Zsw¹⁵, Zsw⁴⁸, Zsw⁷⁷ and Zsw⁸¹⁻⁸³.

[a] O. Binas T. Schamber,, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Goethe University Frankfurt
Max-von-Laue Straße 7, 60438 Frankfurt
E-mail: schwalbe@em.uni-frankfurt.de

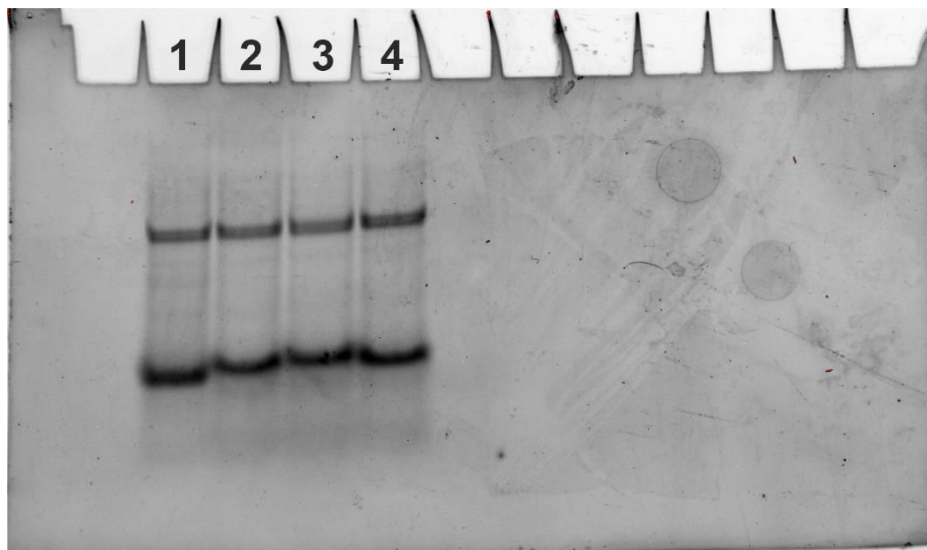
Supplementary Table S1. Summary of all forward and reverse DNA primers used in this study, as well as the RNA sequence of the full length ZMP riboswitch (Zsw¹⁰⁰ RNA).

Construct	Forward Primer	Reverse Primer
Zsw ⁴⁸	5'-TAATACGACTCACTATAGG-3'	5'-CGGATACAGTCCACGTGGTTGCTTC-3'
Zsw ⁷⁷		5'-TTTTGCCCAGGCGGTCGGCTTTTC-3'
Zsw ⁸¹		5'- [CT]AATTTTGGCCAGGCGGTCGGC-3'
Zsw ⁸²		5'- [GC]TAATTTTGGCCAGGCGGTCGG-3'
Zsw ⁸³		5'- [GG]CTAATTTTGGCCAGGCGGTCG-3'
Zsw ⁸⁴		5'- [GG]GCTAATTTTGGCCAGGCGGTC-3'
Zsw ⁸⁵		5'- [TG]GGCTAATTTTGGCCAGGCGGT-3'
Zsw ⁸⁶		5'- [CT]GGGCTAATTTTGGCCAGGCGG-3'
Zsw ⁸⁷		5'- [CC]TGGGCTAATTTTGGCCAGGCG-3'
Zsw ⁸⁹		5'- [CG]CCTGGGCTAATTTTGGCCAGG-3'
Zsw ⁹⁰		5'- [CC]GCCTGGGCTAATTTTGGCCAG-3'
Zsw ⁹¹		5'- [AC]CGCCTGGGCTAATTTTGGCCA-3'
Zsw ¹⁰¹	5'-TAATACGACTCACTATAGGATACA GGACTGGCG-3'	5'-AAAAATAAATACCGCCTGGGCTAATTT TTGCC-3'
Zsw ¹⁰⁰ (RNA)	5'-GGAUACAGGACUGGCGGAUUAGUGGAAGCAACCACGUGGACUGUAUCCGAAGAAAAGCCGACCGCCUGGGCAAAAUU AGCCAGGCGGUUUUAUUUUU-3'	

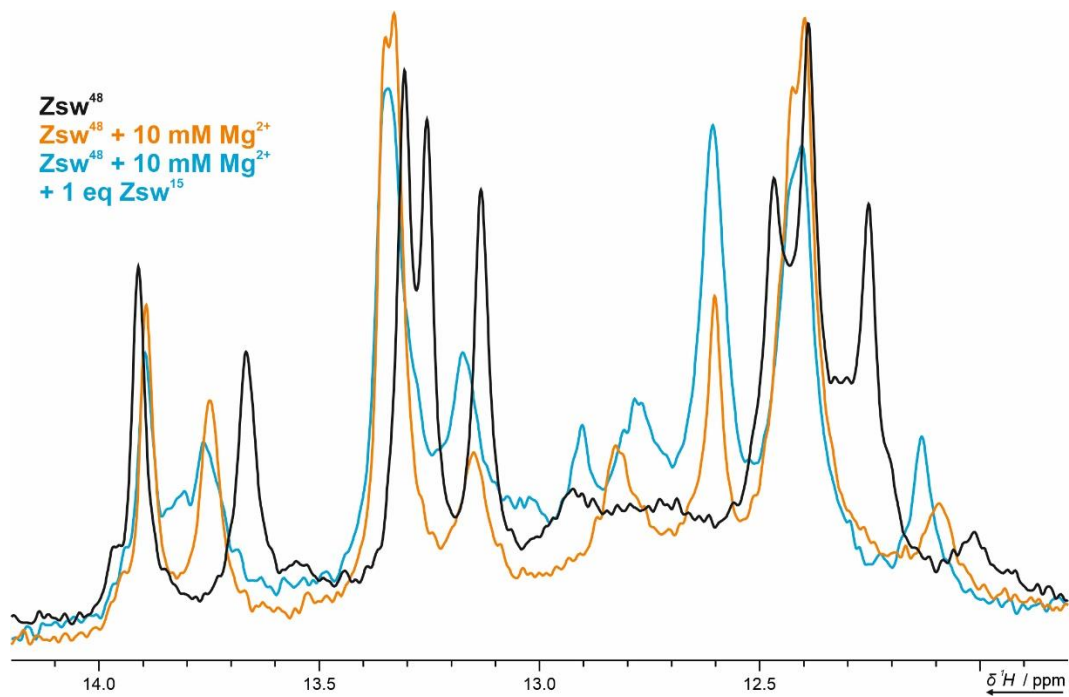


Supplementary Figure S1. Imino regions of ^1H , ^1H -NOESY spectra of Zsw^{48} (a) and Zsw^{77} (b) annotated with assignment. Color-coding indicates resonances from either P1 (cyan) or P2 (orange). Data were measured at 600 MHz, 2048 x 640 points and 184 scans and 800 MHz, 2048 x 512 and 256 points for Zsw^{48} and Zsw^{77} respectively. Samples contained 800 μM (Zsw^{48}) or 400 μM (Zsw^{77}) RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl and 10 % D_2O .

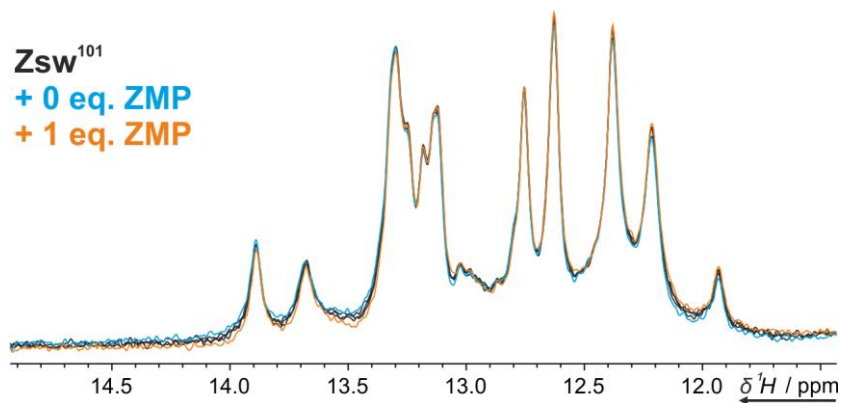
Supplementary Figure S2. ^{15}N -edited and conventional ^1H -1D NMR spectra of ^{15}N -Zsw 48 + ^{14}N -Zsw 15 and conventional ^1H -1D of Zsw 15 in the presence of 10 mM Mg^{2+} . Samples contained 100 μM RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl.



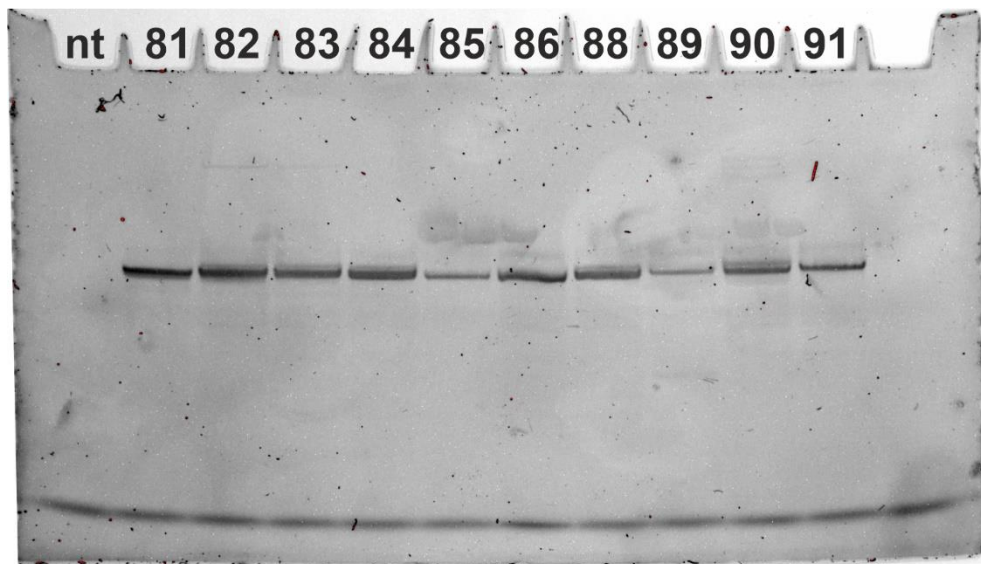
Supplementary Figure S3. 12 % native PAGE of Zsw 48 in the presence of 1eq Zsw 15 , 10 mM Mg^{2+} and 1 eq. ZMP. Pockets were loaded with 200 nmol Zsw 48 (lane 1) and additionally 1 eq. Zsw 15 (lane 2), 10 mM Mg^{2+} (lane 3) and 1 eq. ZMP (lane 4). Upper bands are residual double stranded DNA from the preparation of Zsw 48 .



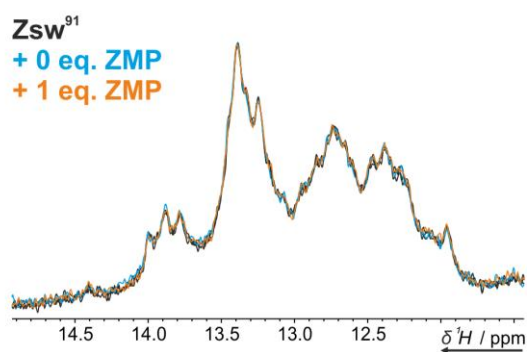
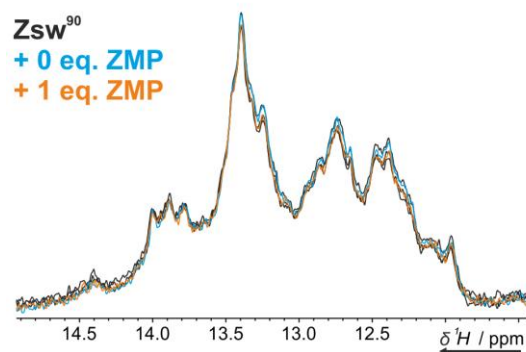
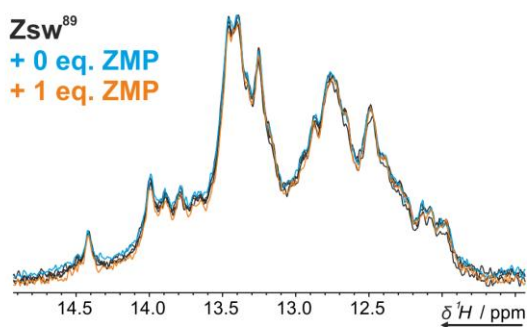
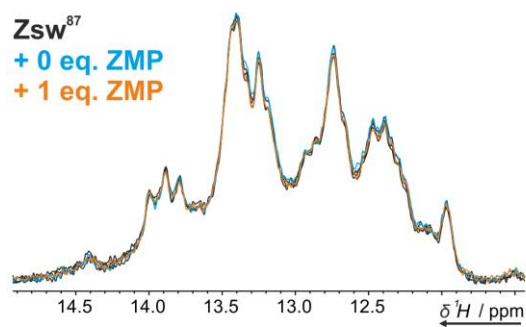
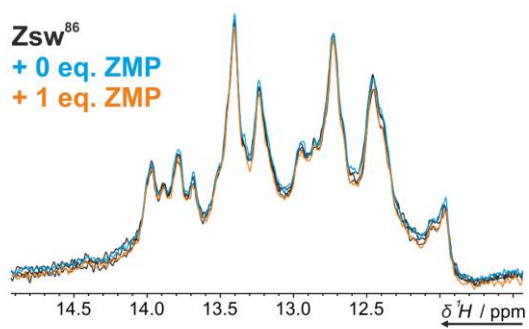
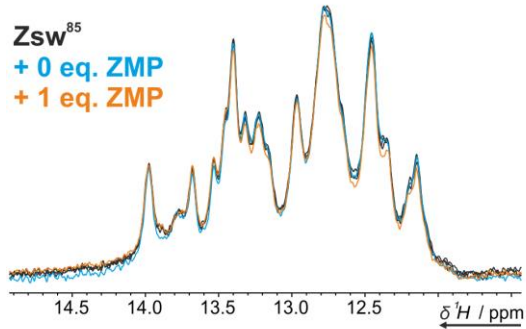
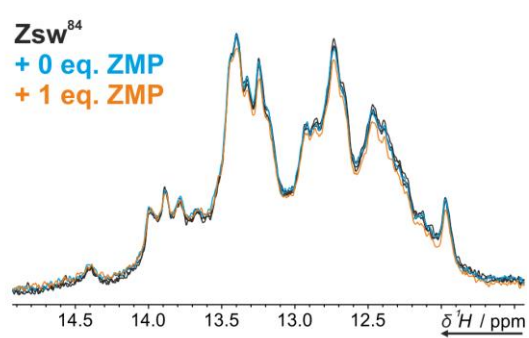
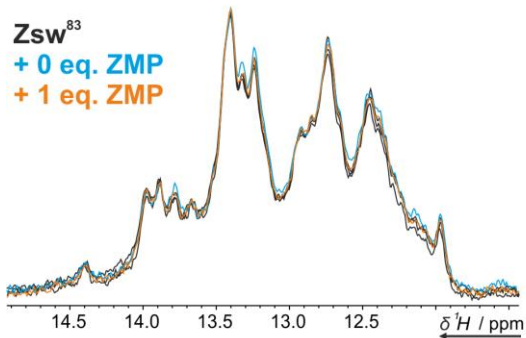
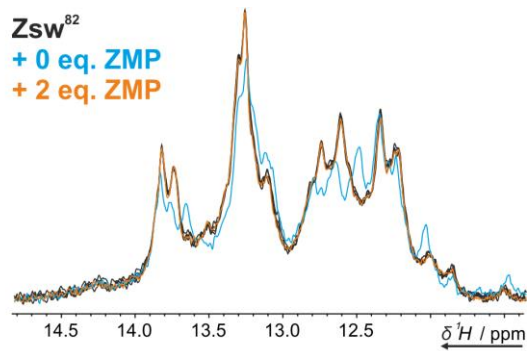
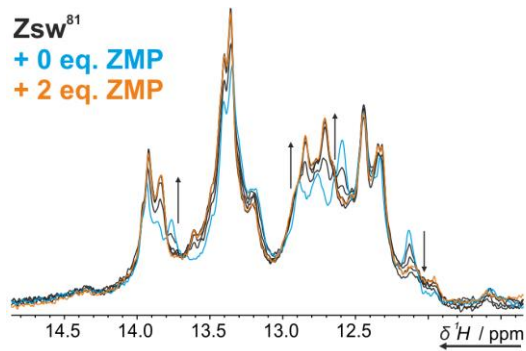
Supplementary Figure S4. 1H -1D spectra of Zsw^{48} in presence and absence of 10 mM Mg^{2+} and 1 eq Zsw^{15} . Samples contained 450 μM RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl 10 mM Mg^{2+} and 10% D_2O .



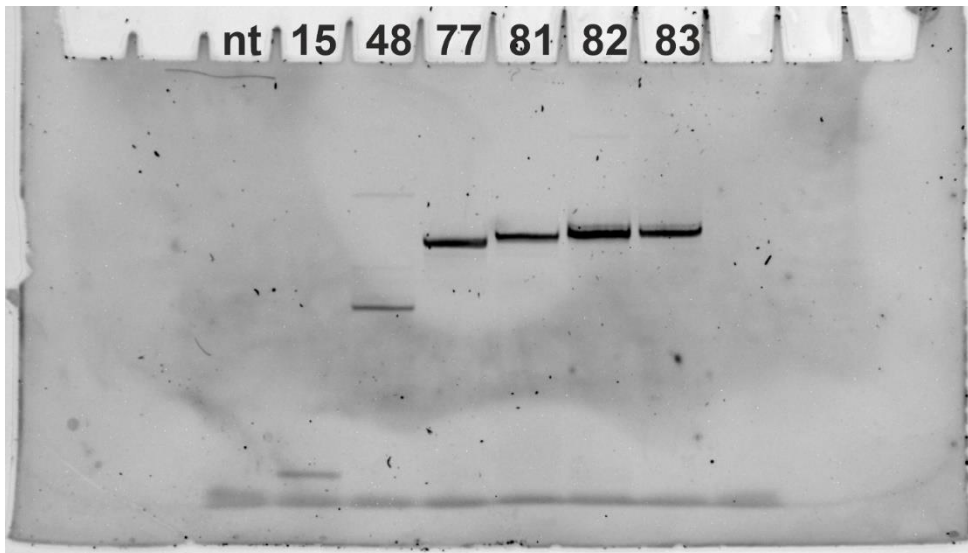
Supplementary Figure S5. 1H -1D spectra of Zsw^{101} in presence and absence of 1 eq ZMP. Samples contained 150 μM RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl 3 mM Mg^{2+} and 10% D_2O .



Supplementary Figure S6. 12 % native PAGE of Zsw⁸¹⁻⁹¹. Pockets were loaded with 0.15 μ l RNA transcription mix.



Supplementary Figure S7. 1H-1D spectra of Zsw⁸¹⁻⁹¹ in presence and absence of 1 or 2 eq. ZMP. Binding is observed only for Zsw⁸¹ and Zsw⁸². Samples contained 100-120 μM RNA, 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl 10 mM Mg²⁺ and 10% D₂O.



Supplementary Figure S8. 12% denaturing PAGE of Zsw¹⁵, Zsw⁴⁸, Zsw⁷⁷ and Zsw⁸¹⁻⁸³. Pockets were loaded with approx. 200 nmol of the respective RNA. Smaller constructs are naturally stained less by the visualization agent and therefore appear weaker.

10. Research article II: Structural basis for the recognition of transiently structured AU-rich elements by Roquin

Oliver Binas*, Jan-Niklas Tants*, Stephen A. Peter, Robert Janowski, Elena Davydova, Johannes Braun, Dierk Niessing, Harald Schwalbe, Julia E. Weigand and Andreas Schlundt

*authors contributed equally to the work

Nucleic Acids Research, **2020**, *48*, 7385 – 7403

AU-rich elements are well-known regulatory elements in eukaryotes, which were long thought to be unstructured. Recently structured AU-rich elements were bioinformatically discovered, which also act as a constitutive decay element (CDE) interacting with Roquin in a cellular context. However, Roquin interaction has only been observed on a phenomenological basis lacking structural evidence. In this work, two AU-rich CDEs were studied by NMR spectroscopy, X-ray crystallography and circular dichroism (CD) spectroscopy. The structural basis for CDE folding was established by NMR structure calculation, while the RNAP complex was investigated with X-ray crystallography. Additionally, conformational switching to an unstructured ARE and interaction with the common ARE-BP AUF1 was confirmed. RNA-protein interaction was monitored in NMR as well as CD experiments.

The author of the thesis contributed the structural characterization of unbound CDEs and CD measurements to the project. RNAs were prepared in the lab of Dr. Weigand by S. Peter. All data necessary for RNA structure calculation were obtained by the author including acquisition, evaluation and calculation of distal and angular restraint parameters. Structure calculation, evaluation, additional refinement and data deposition was performed by the author. Additionally, the author contributed CD melting data of the RNA and the Roquin complexes and prepared the according paragraphs of the manuscript.

Structural basis for the recognition of transiently structured AU-rich elements by Roquin

Oliver Binas^{1,†}, Jan-Niklas Tants^{2,†}, Stephen A. Peter³, Robert Janowski⁴,
Elena Davydova⁴, Johannes Braun³, Dierk Niessing^{4,5}, Harald Schwalbe¹,
Julia E. Weigand^{3,*} and Andreas Schlundt^{2,*}

¹Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany, ²Institute for Molecular Biosciences, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany, ³Department of Biology, Technical University of Darmstadt, Darmstadt 64287, Germany, ⁴Institute of Structural Biology, Helmholtz-Zentrum München, 85764 Neuherberg, Germany and ⁵Institute of Pharmaceutical Biotechnology, Ulm University, 89081 Ulm, Germany

Received March 24, 2020; Revised May 16, 2020; Editorial Decision May 19, 2020; Accepted May 20, 2020

ABSTRACT

Adenylate/uridylylate-rich elements (AREs) are the most common *cis*-regulatory elements in the 3'-untranslated region (UTR) of mRNAs, where they fine-tune turnover by mediating mRNA decay. They increase plasticity and efficacy of mRNA regulation and are recognized by several ARE-specific RNA-binding proteins (RBPs). Typically, AREs are short linear motifs with a high content of complementary A and U nucleotides and often occur in multiple copies. Although thermodynamically rather unstable, the high AU-content might enable transient secondary structure formation and modify mRNA regulation by RBPs. We have recently suggested that the immunoregulatory RBP Roquin recognizes folded AREs as constitutive decay elements (CDEs), resulting in shape-specific ARE-mediated mRNA degradation. However, the structural evidence for a CDE-like recognition of AREs by Roquin is still lacking. We here present structures of CDE-like folded AREs, both in their free and protein-bound form. Moreover, the AREs in the *UCP3* 3'-UTR are additionally bound by the canonical ARE-binding protein AUF1 in their linear form, adopting an alternative binding-interface compared to the recognition of their CDE structure by Roquin. Strikingly, our findings thus suggest that AREs can be recognized in multiple ways, allowing control over mRNA regulation by adapting distinct

conformational states, thus providing differential accessibility to regulatory RBPs.

INTRODUCTION

Precise control of gene expression is essential for every organism, in every cell type and tissue, but appears to be particularly crucial during development and in immune responses of higher eukaryotes. Consequently, dysregulated levels of gene products are a major cause of uncontrolled immune responses, autoimmune diseases and various types of cancer. A key point for regulation of gene expression is at the post-transcriptional level. Steady-state mRNA levels result from their balanced synthesis and decay. mRNA half-lives are primarily controlled by *cis*-regulatory elements within the 3'-UTR of the mRNA, which are recognized by *trans*-acting factors such as microRNAs, but even more often by RNA-binding proteins (RBPs). RBP interactions with target mRNAs trigger downstream effects including mRNA decay, translational inhibition or long-term protection from decay factors leading to an accumulation of certain mRNAs. Such post-transcriptional regulation is crucial to balance the mRNA levels of pro-inflammatory cytokine-encoding mRNAs. These are inherently unstable to promote the resolution of inflammation, preventing unintended tissue damage and autoimmune reactions. *cis*-regulatory elements mediating this instability are either miRNA binding sites, stem-loop (SL) elements or adenylate/uridylylate (AU)-rich elements (AREs) (1,2).

AREs are the currently best-studied *cis*-regulatory elements that control mRNA fate, owing to their early dis-

*To whom correspondence should be addressed. Tel: +49 69 798 29699; Fax: +49 69 798 29225; Email: schlundt@bio.uni-frankfurt.de
Correspondence may also be addressed to Julia E. Weigand. Tel: +49 6151 16 22005; Fax: +49 6151 16 22003; Email: julia.weigand@tu-darmstadt.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Johannes Braun, Center for Thrombosis and Homeostasis, Johannes Gutenberg University Medical Center, Mainz 55131, Germany.

covery and broad investigation over the past decades (3). AREs were initially identified in mRNAs encoding early response genes (4,5), but soon after that also identified in mRNAs of growth factors, pro-inflammatory cytokines and proto-oncogenes, where they likewise induce transcript degradation (6–9). In line with that, aberrantly expressed transcripts of tumor-promoting transcription factors like *c-fos* are characteristic for their lack of AREs (9). However, AREs are now considered a widespread and heterogeneous group of elements that occur in numerous transcripts. They are listed in databases such as AREsite2, which provide a bioinformatic starting point for the characterization of AU-rich elements in a given target gene (10).

AREs had first been subdivided in three major classes with different sequences, where the currently best-studied class comprises AREs with a core sequence of AUUUA, primarily embedded in an (A)U-rich context (3,11). However, various, less-defined but primarily U-rich sequences are able to mediate mRNA decay (9,11). Notably, AREs are most effective in their function when organized in repeats of their core motifs and a total sequence length of up to 150 nucleotides. In addition, they specially act as *cis*-element, as they appear to be both miRNA target elements and targets of regulatory RBPs in inducing mRNA decay, occasionally in a directly overlapping manner (12–14).

Prominent example proteins that promote target mRNA decay are Tristetraprolin (TTP or ZFP36), Butyrate response factor 1 (BRF1 or ZFP36L1) and K-homology splicing regulatory protein (KSRP), all of which recognize one or more linear ARE core motifs (3,15). Of note, such linear AREs are also targeted by mRNA-protecting RBPs, e.g. by the ELAV (embryonic lethal abnormal visual system)-like proteins, such as human antigen R (HuR or ELAVL1) ((16–18). Interestingly, the first protein found to recognize AREs (19), the ARE/poly(U)-binding factor 1 (AUF1 or hnRNP D) is capable of both stabilization and destabilization of its mRNA targets. While it is still enigmatic how AUF1 exerts its dual function, its (de-)stabilizing activity seems to be dependent on cell type, the specific AUF1 protein isoform as well as the target itself [reviewed in (20)]. AUF1 shows a relatively promiscuous target sequence specificity, recognizing not only canonical AREs, but more generally U-rich targets (21,22).

In general, these ARE-BPs harbor multiple RNA-binding domains (RBDs) and recognize clustered regions of short single-stranded RNA *cis*-elements, e.g. multiple copies of AREs organized in hubs. Here, affinity and specificity are increased through multivalent interactions using multiple copies of the same or sets of different RBDs, which is a unifying principle in driving affinity and specificity for RNA-recognition by multi-domain RBPs (18,23–27). While by default, AREs are regarded as linear single-stranded RNA motifs, earlier studies have suggested that the transient presence of expanded structures in ARE regions interferes with the recognition by ssRNA-specific RBPs (23,28), in particular for the canonical AUF1 RRM (RNA recognition motif) domains (29–31). This will *in vivo* lead to an intrinsic dynamic regulation of mRNA stability through the formation of transient RNA secondary structure (Figure 1A), and at the same time influence the potential presence of miRNA target sites.

Recently, small SL elements (constitutive/alternative decay elements, CDE/ADE) have been identified as an alternative way to promote mRNA decay through highly specific recognition by novel types of domains, most prominently shown for Roquin and Regnase. Both proteins are essential regulators of immune responses, preventing excessive cytokine production and autoimmune reactions by regulating a shared set of target mRNAs through a common SL element (32–37). Herein, Roquin uses its unique ROQ domain for specific binding to CDE and ADE SLs (38–42). The involvement of the ROQ domain in mRNA repression has long been assumed independent from the simultaneous presence of single-stranded AREs.

The strictness of the CDE sequences in functional targets for Roquin has been a matter of debate for years now, including contradictory findings (36,38,43,44). Certainly, a major requirement for a functional SL represents its accessibility *in vivo*, i.e. not being masked by other RBPs or unfolded because of low thermodynamic stability. An ongoing redefinition and expansion of the CDE ‘consensus’ has recently led to the identification of AREs that were suggested to fold into CDEs (44). Interestingly, in the *UCP3* mRNA 3′-UTR, such CDE-AREs appear as a tandem recognized by the Roquin ROQ domain (Figure 1B). Despite strong *in vitro* and cellular data underlining the regulation, this earlier study could not unambiguously prove these RNA moieties to be bound by Roquin in a CDE-like manner (i.e. comparable to the *Tnf* CDE (38)). The latter, however, needs to be taken into account considering the recently identified capabilities of the Roquin ROQ domain to bind single-stranded AU-containing sequences (45). Further, the high AU-content of CDE-AREs suggests that these elements exhibit a bi-functional nature, switching between a structured and linear form. As such, CDE-AREs would be hotspots of RBPs competing for these AREs (e.g. AUF1), independent of their preference in recognizing ssRNA or structured RNA. However, we to date lack any atom-resolved evidence of stem-looped AREs, a feature that had been suggested to affect the function of AUF1 before (30,31).

We here provide structural evidence for the existence of CDE-like folded AREs, both in an unbound as well as ROQ domain-bound form, by NMR and X-ray crystallography. By investigating the *UCP3* CDEs 1 and 2, we show that, despite a low intrinsic thermostability, both CDEs are locked in stable complexes with the ROQ domain and engage with affinities comparable to the most canonical CDE found in the 3′-UTR of *Tnf* mRNA (46). Besides, we show that the *UCP3* CDE1 is also a target of the AUF1 RRM1-2 domains in its unfolded form and prove a direct competition of the two proteins for the same CDE-ARE, underlining a central role for the equilibrium between the linear and structured forms of AREs.

MATERIALS AND METHODS

Plasmids

The Roquin ROQ domain used in this study was expressed from the pETTrx1a vector, obtained from Gunther Stier (EMBL); its cloning has been described before (45). For protein purification of the RNA recognition motifs of AUF1 (AUF1.RRM2 and AUF1.RRM1-2), cod-

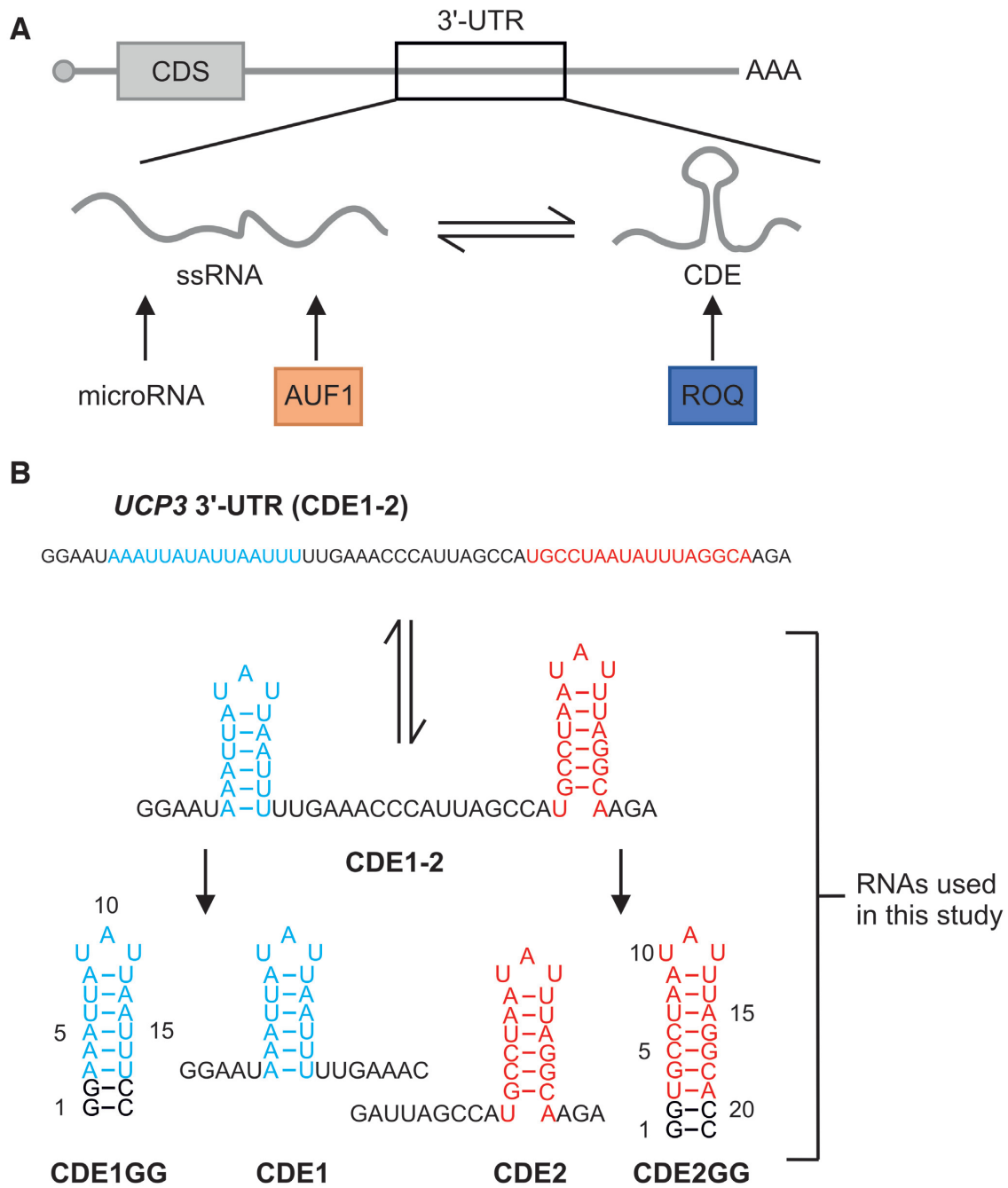


Figure 1. Introduction to AREs, their cognate RBPs and the RNA constructs used in the present study. (A) Scheme of AREs as integral part of an mRNA's 3'-UTR (upper panel). Lower panel, a potential ARE can exist in an equilibrium between its single-stranded form (ssRNA, e.g. recognized by the protein AUF1) and a structured form, the latter exemplified by a CDE-like shaped element bound by the highly specific Roquin ROQ domain. The target ARE conformers are mutually exclusive and can lead to redundant or competitive regulation of the ARE-containing mRNA through the respective cognate RBP. (B) The previously described regulatory ARE region in the *UCP3* mRNA 3'-UTR shown as linear sequence and as its structured conformer with the two CDEs bound by Roquin (44). Below that, the two CDEs are shown as isolated wild type species and as stabilized SLs with the underlying numbering of bases as used in this study. See also Supplementary Figure S1.

ing sequences were generated by PCR amplification using pCMV-hnRNP D37 as a template (47) and introduced into *Xba*I and *Xho*I sites of an His₆- and thioredoxin-tag encoding plasmid based on pETTrx1a, i.e. including a TEV (tobacco etch virus) site for proteolytic removal of all tags.

For RNA synthesis, CDE1, CDE2, CDE1GG, CDE2GG and CDE1-2 (tandem) sequences together with the T7 promoter were generated by hybridization of complementary oligonucleotides and introduced into the NcoI and HindIII sites of an HDV ribozyme encoding plasmid based on the pSP64 vector (Promega). RNAs were transcribed as HDV ribozyme fusions to obtain uniform 3' ends.

For the luciferase reporter system, 3'-UTR variants were generated by hybridization of complementary oligonucleotides and introduced downstream of the firefly luciferase open reading frame into the multiple cloning site of pDLP (48) using NotI and HindIII restriction sites.

All protein and RNA sequences are summarized in Supplementary Tables S1 and S2.

Protein production

The proteins were produced as His₆-Trx-fusion proteins containing a TEV cleavage site from a single freshly transformed clone grown in an overnight culture and used to inoculate an expression culture in either LB or M9 minimal media containing 100 µg/l kanamycin. Cells were grown to an OD₆₀₀ of 0.7–0.9, protein production was induced with 0.5 mM isopropyl-β-D-thiogalactopyranosid, and cultures were grown overnight at 37°C. Cell pellets were lysed by sonification in 150 mM sodium chloride, 50 mM Tris pH 8.3, 4 mM β-mercaptoethanol. For affinity purification via immobilized metal ion affinity chromatography (IMAC), manual Ni²⁺-NTA (nickel-nitrilotriacetic acid agarose) columns were used with the lysis buffer complemented with 50, 100 and 300 mM imidazole. The eluted protein was TEV-cleaved overnight during dialysis against 1 M sodium chloride, 50 mM Tris pH 8.0, 4 mM β-mercaptoethanol at 4°C. For the reverse IMAC the cleaved protein was loaded onto Ni²⁺-NTA columns and washed with the same buffer containing 50, 100 and 300 mM Imidazole. The flow-through containing the target protein was concentrated prior to size exclusion chromatography. For that, a Superdex S75 16/60 column was run in 1 M sodium chloride, 20 mM Tris pH 7.0 and 2 mM Tris-(2-carboxyethyl)-phosphin (TCEP). The protein was frozen in liquid nitrogen, stored at –80°C and buffer exchanged to 150 mM sodium chloride, 20 mM Tris pH 7.0, 2 mM TCEP using Amicon[®] centrifugal filter units prior to use. For NMR titration experiments the proteins were concentrated to 40–50 µM or 1–1.5 mM stock concentrations when used for competition experiments.

In vitro transcription of RNAs

pHDV constructs were linearized with HindIII and purified by phenol extraction. 2 mg of linearized DNA were transcribed overnight at 37°C using the following conditions: 200 mM Tris-HCl pH 8.0, 20 mM magnesium acetate, 50 mM dithiothreitol (DTT), 2 mM spermidine, 4 mM nucleoside triphosphates (NTP) (each) and 100 µg of in-

house-made T7 polymerase. After transcription, precipitated pyrophosphate was pelleted by centrifugation and 20% (v/v) EDTA (0.5 M, pH 8.0) was added to the supernatant. After ethanol precipitation, the RNA was purified using denaturing polyacrylamide gel electrophoresis. The RNA was detected via ultraviolet (UV) shadowing, excised and eluted from the gel in 0.3 M sodium acetate pH 6.5 at 4°C overnight. To remove the remaining gel slices, the supernatant was filtered and the RNA was again precipitated using ethanol. Finally, the RNA was resolved in double-distilled water and stored at –20°C. The purified RNA was folded by heating to 95°C for 5 min prior to injection into five equivalents of ice-cold water. For buffer exchange (NMR buffer: 25 mM potassium phosphate pH 6.2, 50 mM potassium chloride) and concentration, we used Vivaspin[®] concentrators (molecular weight cut-off 3.000 Da).

NMR spectroscopy

NMR experiments for RNA structure determination and RNA-protein interactions were performed at the BMRZ at the Goethe-University Frankfurt. Spectra were acquired on Bruker AV600, AV700, AV800 and AV950 spectrometers equipped with triple-resonance cryoprobes.

¹H–¹⁵N-HSQC of ROQ alone and with UCP3 CDE RNA variants were measured at 40 µM in 20 mM Tris, 150 mM sodium chloride, 2 mM TCEP, 0.02% NaN₃, pH 7.0 at 298 K. Spectra for AUF1 RRM2 were recorded at 100 µM and for tandem RRM1-2 at 500 µM in the same buffer. For titration and competition experiments AUF1 RRM1-2, concentrations were adjusted to 40 µM. 1D-¹H spectra for the imino proton region of the individual CDE versions were recorded at concentrations of 20 µM with 2k scans at 900 MHz proton frequency. Both CDE1 and CDE2 were titrated in molar ratios of 0.25, 0.5, 1.0, 1.5 and 2.0 to ¹⁵N-labeled ROQ. Tandem CDE1-2 was titrated in steps of 0.125, 0.25, 0.5, 1.0 and 1.5 molar ratios. To test competition for RNA binding, the complex of ¹⁵N-labeled ROQ and CDE1 was titrated with unlabeled AUF1 RRM1-2 in steps of 0.5, 1.0, 2.0 and 3.0 molar ratio. The reverse experiment was done with ¹⁵N-labeled AUF1 RRM1-2/CDE1 complex and unlabelled ROQ (1.0, 2.0, 4.0 molar ratio). To saturate the first binding protein (ROQ or AUF1, respectively) CDE1 was added in small excess (1:1.125). For each titration and competition step, separate imino proton spectra were recorded. To assess the stabilizing effect of Roquin binding, 1D imino spectra were recorded of 20 µM RNA samples titrated with 0.5 and 1.0 (all RNAs), additionally 2.0 (CDE1, CDE2, tandem CDE1-2), and additionally 4.0 (CDE1 and CDE2) molar ratios of ROQ. All protein and RNP spectra were recorded at 298 K.

All samples for the CDE1/2GG structure calculations contained 100 µM sodium tri-methyl-silyl-propane-sulfonate as an internal reference. CDE1GG samples contained 220 µM RNA ([H₂O]/[D₂O] = 9:1), 350 µM RNA (100% D₂O) and 300 µM RNA (100% D₂O) for unlabelled and uniformly ¹³C¹⁵N-labeled samples, respectively. CDE2GG samples contained 900 µM RNA ([H₂O]/[D₂O] = 9:1) and 500 µM RNA (100% D₂O) or 450 µM RNA ([H₂O]/[D₂O] = 9:1) and 1.3 mM RNA (100%

D₂O) for unlabeled and uniformly ¹³C,¹⁵N-labeled samples, respectively. Distance restraints were obtained from homonuclear 2D-¹H,¹H-NOESY spectroscopy on unlabeled RNA samples. NOESY spectra were acquired at 600 MHz (CDE1GG) / 950 MHz (CDE2GG) with 2048 × 656 (CDE1GG) or 4096 × 712 (CDE2GG) points. Angular restraints for angles β, ε and χ were obtained from J-modulated HCP (49), PFIDS (50) and Γ-HCN (51) experiments (see Supplementary Figures S3 and S4). T₁ and het-NOE rates were determined from ¹³C-correlated 2D sensitivity enhanced experiments. We used relaxation delays of 10, 50, 100, 200, 400, 700, 1000 and 1500 ms for T₁ determination. Temperature compensation for spin-lock heating was achieved by implementation of spin-lock pulses distributing identical heating during the relaxation delay of the experiments. HetNOE experiments were acquired in an interleaved manner. Fitting of relaxation curves was conducted using the software implemented in the Bruker Dynamic Center (Bruker Biospin).

Structure calculation of free CDEs

Structure calculation was performed with a modified version of CNS 1.2 (ARIA 1.3) (52). We included all distance restraints obtained from NOESY data (Table 1). Angular restraints for angles β, ε and χ were experimentally determined and incorporated with error margins of ±15°. Angles α, γ and ξ were taken from the work of Nozinovic *et al.* (53) and were incorporated as constraints only for stem residues with error margins of ±20°. Error margins of angular restraints of the loop-closing base pair were set to ±50° to allow backbone adaptation to loop geometry. No violations of angular restraints were observed. Data were corrected for spin diffusion with a rotational correlation time of 4.3 ns for both structures [calculated by HydroNMR (54)]. Refinement of structures was carried out with Xplor-NIH (55). All restraints applied in the original structure calculation were employed in addition to the statistical torsion potential torsionDB_{RNA} provided with the Xplor-NIH software package.

Circular dichroism (CD) spectroscopy

CD melting curves were acquired on a JASCO J-810 CD-spectrometer in a 0.2 μm quartz cuvette. Samples contained 10 μM RNA, 25 mM potassium phosphate and 50 mM potassium chloride. The wavelength showing maximal circular dichroism was determined from prior CD-spectra and was in the range from 260 to 265 nm. CD-melting curves were measured at a temperature-sampling rate of 1°C/min in a range from 5 to 95°C with a sampling rate of 10/min. Data was smoothed with a Savitzky-Golay filter (10 points). Melting curves were fit to a sigmoidal curve in the respective sigmoidal region with errors <2%. The melting temperature was obtained as the point with the highest slope.

Isothermal titration calorimetry (ITC)

ITC measurements were performed with either MicroCal PEAQ- or VP-ITC devices (Malvern, United Kingdom) in NMR buffer. In all experiments, ROQ domain protein was

Table 1. Statistics of CDE1GG and CDE2GG NMR structure calculations. Statistics and accession numbers derived from structures obtained with ARIA/CNS and when additionally refined with Xplor-NIH

NMR statistics	6XXB	6XXA
PDB ID (CNS)	6XWW	6XWJ
PDB ID (Xplor-NIH)		
RNA name	CDE1GG	CDE2GG
NOE restraints		
unambiguous:		
intraresidual	144	298
sequential	59	134
medium range	5	6
long range	14	28
ambiguous:		
intraresidual	0	6
sequential	0	9
medium range	0	1
long range	0	2
Base pair restraints		
Planarity	7	9
H-bond	32	46
Dihedral restraints		
sugar pucker	95	105
backbone	117	94
glycosidic bond	16	21
Violations (Xplor-NIH)		
Distance > 0.5 Å	1	2
Dihedral > 10°	0	1
RMSDs/Å (CNS)		
Stem (without terminal)	1.1	1.6
Loop	2.2	2.3
All atom	1.7	1.7
RMSDs/Å (Xplor-NIH)		
Stem (without terminal)	0.17	0.27
Loop	0.39	0.21
All atom	0.33	0.32

titrated from a stock of 10–20-fold excess to 20–40 μM of *UCP3* CDE1/2GG or *Tnf* CDE RNAs provided in the reaction cell. In individual runs, we used 19–25 injections of protein with 150–180 s of spacing at 283 K and a stirring speed of 750 rpm. Raw data were analyzed with the NIT-PIC software tool (56) and heat production was fitted to a one-site binding model. For all runs, we performed a buffer subtraction using NMR buffer instead of RNA.

Crystallization, diffraction data collection and processing

The crystallization experiments for ROQ-RNA complexes were performed at the X-ray Crystallography Platform at Helmholtz Zentrum München. The crystals of Roquin-1 ROQ (171–326) with *UCP3* CDE2GG were grown in 250 mM ammonium sulphate and 30% (v/w) PEG 4000. The crystals of Roquin-1 ROQ (171–326) with *UCP3* CDE1GG were grown in 90 mM Bis-tris propane pH 6.5, 290 mM sodium tartrate and 19% (v/w) PEG 3350. Crystallization was performed using the hanging-drop vapor-diffusion method at 292 K in 24-well plates and a protein concentration of 11.5 mg/ml. The crystals appeared after two days. For the X-ray diffraction experiments, the crystals were mounted in a nylon fiber loop and flash cooled to 100 K in liquid nitrogen. The cryoprotection was performed for five seconds in reservoir solution complemented with 25% (v/v) 2-methyl-2,4-pentanediol for the structure with CDE2GG, or 25% (v/v) ethane-1,2-diol for the structure with CDE1GG. X-ray diffraction data for the ROQ-RNA

Table 2. Data collection and refinement statistics for the ROQ domain in complex with *UCP3* CDEs

Data collection	6TQB CDE1GG	6TQA CDE2GG
PDB ID		
RNA name		
Beamline	SLS PXIII X06DA	SLS PXIII X06DA
Wavelength (Å)	1.0	1.0
Space group	<i>P4</i> ₂ <i>2</i> ₁ <i>2</i>	<i>P2</i> ₁
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	86.99, 86.99, 72.99	44.53, 160.66, 68.01
α , β , γ (°)	90, 90, 90	90, 107.96, 90
No. of molecules per asymmetric unit	2 (1 protein + 1 RNA)	8 (4 protein + 4 RNA)
Resolution (Å)	47–1.6 (1.64–1.60)*	50–2.4 (2.46–2.40)*
<i>R</i> _{merge}	4.7 (73.6)	10.0 (83.7)
<i>I</i> / σ <i>I</i>	23.16 (2.46)	14.05 (2.15)
CC (1/2)	100 (79.7)	100 (63.1)
Completeness (%)	96.9 (75.5)	99.6 (99.8)
Redundancy	8.4 (6.3)	5.2 (5.1)
Refinement		
Resolution (Å)	47–1.6 (1.642–1.600)	50–2.4 (2.462–2.400)
No. reflections	36 355 (2052)	33 655 (2618)
<i>R</i> _{work} / <i>R</i> _{free}	15.63/19.72 (23.80/24.10)	19.68/25.90 (29.40/34.60)
No. atoms	2195	6,942
Protein	1224	4847
RNA	400	1796
Water	241	295
Ions	7	4
<i>B</i> -factor overall	28.25	55.20
RMSDs		
Bond lengths (Å)	0.013	0.01
Bond angles (°)	1.836	1.796
Ramachandran plot		
Most favored (%)	100	96
Additional allowed (%)	0	4

*Values in parentheses are for highest-resolution shell.

complexes were collected using a Pilatus 2M-F detector at 1.0 Å wavelength on the PXIII X06DA beamline at SLS (Villigen, Switzerland). The best data set was indexed and integrated using *XDS* (57) and scaled using *SCALA* (58). Intensities were converted to structure-factor amplitudes using the program *TRUNCATE* (59). Table 2 summarizes data collection and processing statistics.

Structure determination and refinement

The crystal structures of ROQ-RNA were solved by molecular replacement using *PHASER* (60). The native Roquin-1 ROQ structure served as a search model (PDB: 4QI0 (38)) for solving the structure of Roquin-1 ROQ (aa 171–326) with CDE2GG and the Roquin-1 ROQ structure in complex with *Tnf* CDE served as a search model (PDB: 4QI2 (38)) for solving the structure of Roquin-1 ROQ (aa 171–326) with CDE1GG. Model building was performed in *COOT* (61). The RNA molecules were modeled manually. The refinement was performed in *REFMAC5* (62) using the maximum-likelihood target function including translation, liberation and screw-rotation displacements of a pseudo-rigid body (63). The final model of the structure with CDE2GG is characterized by *R* and *R*_{free} factors of 19.7 and 25.9%, the final model of the structure with

CDE1GG is characterized by *R* and *R*_{free} factors of 15.8 and 19.7%, respectively (Table 2). The stereochemical analysis of the final model was done in *PROCHECK* (64) and *MolProbity* (65).

Cell culture

HEK293 cells (Leibniz-Institute DSMZ, Germany, ACC 305) were cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS Superior, Biochrom), 1 mM sodium pyruvate (Thermo Fischer Scientific) and Pen Strep (Thermo Fisher Scientific) at 37°C in a 5% CO₂ humidified incubator.

Transient transfection and luciferase assay

For transfection of pDLP reporter plasmids, 100 000 HEK293 cells were seeded in 24-well plates. 24 h after seeding, cells were transfected with 100 ng reporter plasmid using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocol. Firefly and *Renilla* luciferase activity were measured 24 h post-transfection using the Dual Luciferase Reporter Assay System (Promega).

RNA affinity purification and Western blot

Purified RNA was dephosphorylated using calf intestine phosphatase (Roche). For 5' end biotinylation, 600 pmol RNA were incubated with 0.2 mM γ -S-ATP (Biomol) and T4 polynucleotide kinase (New England Biolabs) for 30 min at 37°C. Biotin-long-arm maleimide (Vector Laboratories) was added and incubated for 30 min at 65°C. Unincorporated label was depleted by lithium chloride precipitation. Biotinylated RNA (200 pmol) was conjugated to Dynabeads M-280 (Invitrogen) in incubation buffer (10 mM Tris-HCl pH 7.4, 150 mM potassium chloride, 0.5 mM DTT, 0.05% NP40, 100 U/ml RNasin) for 2 h at 4°C with continuous rotation. Whole cell protein lysate (1 mg) of HEK293 cells together with 200 μ g yeast tRNA (Sigma-Aldrich) and 5 mg Heparin (Sigma-Aldrich) was added to the beads and incubated 1 h at 4°C followed by 15 min at room temperature with continuous rotation. Beads were washed five times with incubation buffer, resuspended in 30 μ l protein loading dye and boiled at 95°C for 10 min. Eluted proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Bio-Rad). The primary antibody was anti-AUF1 (1:10 000; rabbit; Millipore; 07-260) and the secondary antibody was Horseradish-conjugated anti-rabbit (1:7000; goat; Jackson ImmunoResearch). Blots were developed using ECL Select (Life Technologies). Imaging was performed on a ChemiDoc Imaging system (Bio-Rad).

RESULTS

Recognition of *UCP3* AU-rich elements is CDE-like

Recognition of AREs by regulatory RBPs has mainly been described for binding of single-stranded (ss) RNAs to RBDs including ZnF and RRM domains, while the formation of ARE secondary structure had been suggested

to interfere with ARE-mediated regulation, e.g. in the protein AUF1 (30) (Figure 1A). This general hypothesis has recently been challenged by CDE-like AREs bound by the protein Roquin (44). The 3'-UTR of *UCP3* harbors two AU-rich elements with the potential to fold into trinucleotide hairpin loops, reminiscent of canonical CDEs (36). CDE1 contains only AU nucleotides and CDE2 harbors the canonical ARE motif 5'-AUUUA-3' preceded by four AU nucleotides (Figure 1B and Supplementary Figure S1). While regulation of the *UCP3* mRNA was shown to function through these sequences, we still lack an unambiguous evidence for whether the unstructured, single-stranded ARE RNA species or only the CDE-folded AREs are the responsible target of the ROQ domain within the functional *UCP3* mRNP.

Thus, we sought to investigate the recognition mode of those CDE-AREs by Roquin. Taking into account the high intrinsic lability of AU-rich sequences, we decided to use solution NMR spectroscopy, which is unique in monitoring also weak molecular interactions under near physiological environment and at atomic resolution. We recorded ^1H - ^{15}N -HSQC spectra comparing CDE1 and CDE2 individually in their free and ROQ domain-bound forms. As shown in Figure 2A and B, native individual CDE RNAs provoke strong chemical shift perturbations (CSPs) in the ROQ domain and the subsequent plot patterns are reminiscent of the CSPs that had been observed with the *bona fide* *Tnf* CDE before (38) (Supplementary Figure S2A). In contrast, titration of ROQ with ssRNAs was recently found to cause significantly smaller CSPs, in line with a 10–100-fold lower affinity (45), e.g. for the *Nfkbiz* 15-mer LBE RNA (Supplementary Figure S2A). This finding supports the assumption that *UCP3* AREs are bound by the ROQ domain in a CDE-like manner. Notably, HSQC spectra of ROQ with either CDE are well superimposable despite obvious differences in CDE sequences (Supplementary Figure S2B). This underlines a highly similar, CDE-shape recognition rather than recognition of a linear sequence. The presence of RNA structure in CDE1 and 2 upon binding to ROQ is also evidenced by the observation of imino proton signals indicating the formation of H-bonds between complementary bases in a stem (Supplementary Figure 2E). In contrast, we only detected imino signals for the free GC-containing CDE2, while the pure-AU CDE1 did not reveal the presence of long-lived, i.e. NMR-detectable base pairs in its free form (Supplementary Figures S1 and S2) underlining the low intrinsic stability of AU-rich base pairs, while the underlying sequences suggest to be present in an unfolded, linear form to certain extent of their life time.

To exclude that only a small fraction of RNA is bound as SL, we next constructed stabilized CDEs with an additional two GC base pairs at the lower stem (Figure 1B). 1D- ^1H NMR spectra highlighting the respective imino proton regions show that these CDE-ARE versions are thermodynamically more stable in their free forms (Supplementary Figure S1), thus increasing the overall likelihood to be engaged by the ROQ domain in a CDE-preformed shape. HSQC titrations show that the stabilized CDE-AREs provoke an identical pattern in CSP plots; both in significantly shifting residues and absolute intensities (Figure 2C and D). Thus, we conclude that both wild type (WT, non-stabilized)

and stabilized *UCP3* CDE-AREs are equally bound as structured, CDE-like RNAs.

Structures of folded AU-rich elements

We next investigated whether the *UCP3* CDE-AREs form the expected CDE-like SLs in the absence of Roquin, to test for their potency in acting as a pre-structured, unique *cis*-regulatory element, despite their AU-rich content. Given the apparent high intrinsic dynamics of the WT CDEs, yet the identical binding behavior towards ROQ, we determined the structures and dynamics of apo CDE1GG and CDE2GG by solution NMR (Figure 3 and Supplementary Figures S3 and S4). Spectral regions to demonstrate NOE data quality are displayed in Figure 3A and B for CDE1GG and CDE2GG, respectively, and in Supplementary Figures S3 and S4.

For CDE1GG, the ten lowest energy structures converged to an all-residue-all-atom RMSD of 1.7 Å (stem RMSD without terminal residues: 1.1 Å, loop RMSD: 2.2 Å) after structure calculation with ARIA/CNS (Table 1 and Supplementary Figure S3). The CDE1GG stem is composed of eight Watson-Crick base pairs spanning residues 1–8 and 12–19. While residues G1 and C19 show increased RMSD values, common for the termini in RNA SL structures, residues in the core of the stem converge well within the 10 best structures. In contrast, the loop residues are considerably less converged. For example, the loop residue U9 mostly stacks with the A8 residue in direction of the major groove of the stem, while in two structures of the bundle, U9 points to the solvent (Supplementary Figure S3G).

For CDE2GG, the ten lowest energy structures converged to an all-residue-all-atom RMSD of 1.7 Å (stem RMSD without terminal residues: 1.6 Å, loop RMSD: 2.3 Å) (Table 1 and Supplementary Figure 4). The stem consists of nine well-converged Watson-Crick base pairs. In contrast to the structure of CDE1GG, the central purine of the tri-loop (A11) does not point in the direction of the major groove, but in the opposite direction (Supplementary Figure S4G). In nine of the ten structures, the flanking loop residues U10 and U12 point towards the same direction as A11. In a single, deviating structure, we observed a flip of the first loop residue (U10). Notably, the analogous residue in the structure of CDE1GG (U9) showed a similar degree of structural heterogeneity as CDE2GG U10. Increased RMSD may partially be the result of lack of NOE data, but also a consequence of loop dynamics faster than NMR time-scale. As both the CDE1GG and the CDE2GG structures show increased RMSD values for loop nucleotides compared to the base-pairing stem (Table 1), we suggest their increased dynamics to facilitate protein recognition.

The increased dynamics of the two RNA SLs are also evident from heteronuclear relaxation data (Supplementary Figures S3F and S4F). We found around 40% increase of sugar C1' proton R_1 rates and over 15% increase in hetNOE intensity on aromatic protons for residues 8–12 (CDE1GG) and 9–13 (CDE2GG), respectively, compared to other residues. Thus, loops and loop flanking residues show increased dynamics compared to stem residues. In the context of molecular recognition by ROQ (Figure 2), the

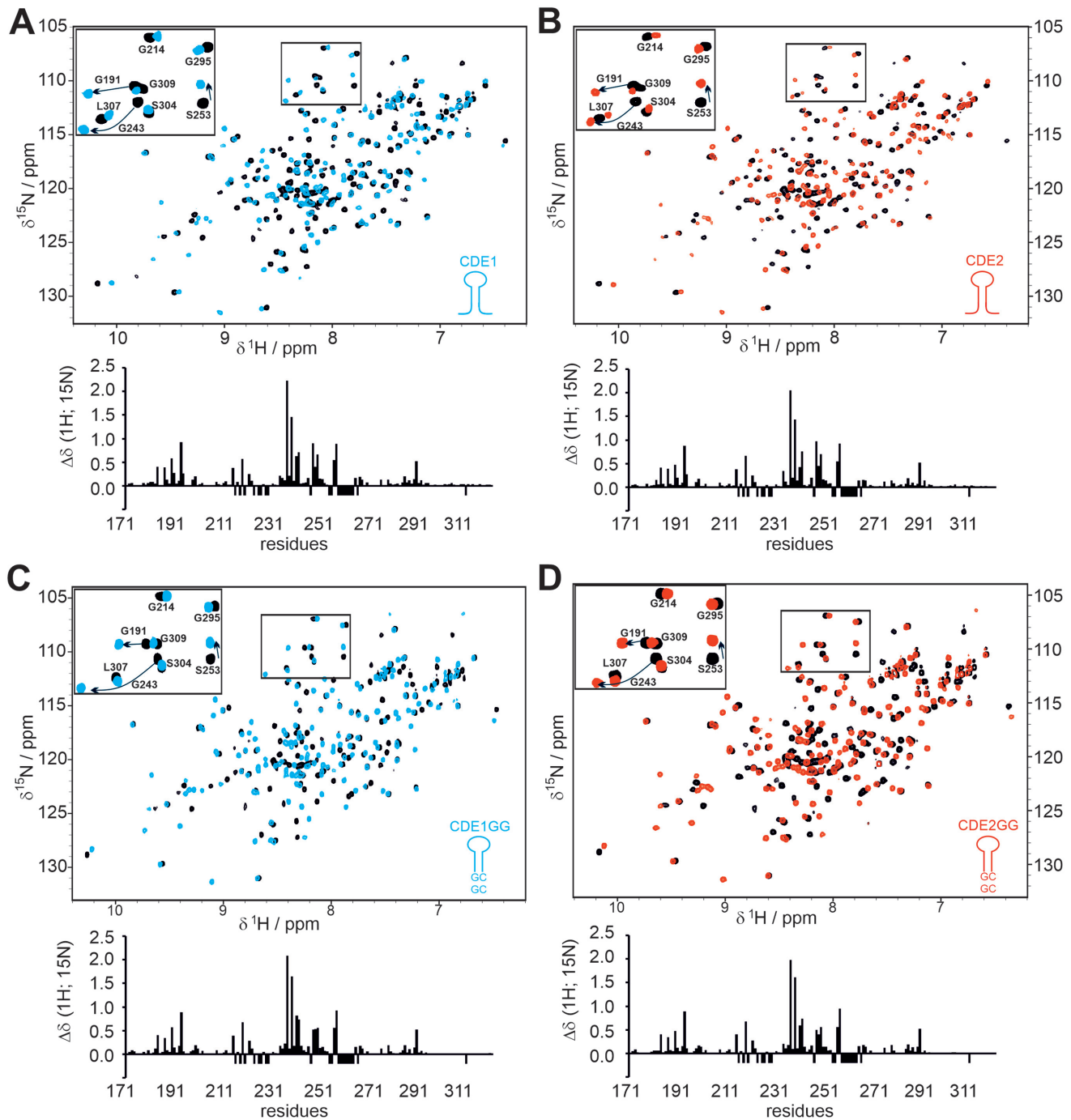


Figure 2. Roquin recognizes wild type and stabilized *UCP3* AREs in a CDE-like manner. (A and B) Overlay of ^1H - ^{15}N -HSQCs showing ROQ alone and in complex with *UCP3* wild type CDE1 (A) and CDE2 (B). The bar plots below the spectral overlays show the combined CSPs of amide resonances along with the primary ROQ domain sequence. (C and D) The same as in panels (A) and (B), but with CDE1GG (C) and CDE2GG (D) SLs from the *UCP3* 3'-UTR. The insets consistently show an equivalent zoom-in from all overlays highlighting a part of the glycine region. Selective amide assignments are given for orientation. Negative bars indicate residues with missing assignments in at least one of the two spectra or prolines. See also Supplementary Figure S2.

Figure 3. Solution structures of free CDE1GG and CDE2GG RNAs. (A and B) Zoom-in of the assigned aromatic H1' region of the $^1\text{H}, ^1\text{H}$ -NOESYs of CDE1GG (A) and CDE2GG (B). The intraresidual H5/H6 cross peak assignment is shown in light grey. (C and D) The three dimensional structures of CDE1GG (blue) and CDE2GG (red) after refinement with the force-field *RNA-ff1* (55) (see also Table 1). Two orientations for each CDE show the side-on view with the loops pointing into the direction of the viewer, and a top down view displaying the loop-orientation with respect to the helical main axis. Tri-loop residues are labeled in sequential order (orange, magenta and green respectively) to highlight differences in loop geometry. (E) NMR structure of the *Tnf* CDE (43). See also Supplementary Figures S3 and S4.

different loop geometries between the two CDEs suggest that the large dynamics facilitate folding of CDEs into a conformation, susceptible for engagement with the ROQ domain.

We additionally refined the best resulting structures with the RNA-specific force field *RNA-ff1* (55), to accommodate to the backbone angles found in X-ray structures of RNAs deposited in the PDB (Table 1 and Figure 3C and D). This procedure removes most of the angular deviations in the loop and therefore does not account for the specifics of loop flexibility. The resulting structure represents the lowest energy and thus most populated structure free in solution. The all-atom RMSD of the refined structures is 0.33 Å for CDE1GG and 0.32 Å for CDE2GG, respectively (Table 1). We attribute the strong tightening of the bundle to parametrization introduced with the torsion potential of the *RNA-ff1* force field. In line with that, the additional refinement led to improved backbone conformations reflected by an increased backbone suiteness from 0.19 to 0.69 for

CDE1GG and 0.14 to 0.75 for CDE2GG, a value measuring the overall agreement of backbone parameters with a test set of RNA structures (66).

In comparison, the solution structure of the apo *Tnf* CDE (43) shows a structural arrangement of the loop residues oriented towards the major groove form, i.e. comparable to CDE1GG (Figure 3E). Additionally, a widening of the major groove is observed in the *Tnf* CDE, attributed to a purine stack in the 3' side of the stem. As the *UCP3* CDE-AREs do not feature this type of purine stack, this effect is not visible in our RNA SL structures. It was further shown that inversion of the two closing base pairs below the tri-loop in the *Tnf* CDE, and thus distortion of the purine stack, induced large differences in the loop structure, accompanied by loss of Roquin binding capability (43). Interestingly, *UCP3* CDE1GG and CDE2GG differ from each other by an inversion of the penultimate loop-closing base pair (Figure 1B), which potentially accounts for the structural alteration observed in the loop, although their binding com-

petence for the ROQ domain appears unaffected (Figure 2 and Supplementary Figure S2). Altogether, we assume that the less rigid AU-rich loop-closing sequence in CDE1 and CDE2 allows for loop dynamics, beneficial for adoption of a ROQ-binding-competent structure.

Structures of folded AU-rich elements in complex with the ROQ domain

To finally prove that the *UCP3* CDE-AREs are tightly complexed by Roquin in the characteristic arrangement, we crystallized the ROQ domain with the two stabilized SLs. We obtained well-diffracting crystals with both SLs and found ROQ-CDE1GG present with one complex molecule per asymmetric unit at a resolution of 1.6 Å (Table 2, Figure 4A and Supplementary Figure S5). ROQ-CDE2GG crystallized with four complex molecules per unit (2.4 Å resolution) that show an overall RMSD of 0.35 Å, while the complex appears as monomer according to the observed line width in NMR spectra. This is in line with all previously published crystal structures of SL RNAs in complex with the core ROQ domain.

Importantly, in neither of the two complexes the additional, stabilizing GC base pairs participate in direct interactions with the ROQ domain, indicating that the wild type CDE SLs are sufficient for a canonical CDE-like recognition by Roquin. In case of CDE1GG, this also reflects the first CDE structure with a pure-AU interface with the ROQ domain, adding functional impact and relevance to the so far unlikely concept of structured AREs. In addition, the high resolution for CDE1GG allowed for the observation of two distinct RNA conformations at the 5' end of the stem, i.e. the first five nucleotides, while the complement strand appears as a single conformer (Supplementary Figure S5B). Although, one of the two conformations appears to be more compacted over the other, we find comparable and complete H-bonding patterns in respective Watson-Crick base pairs. In addition, both conformers share a highly conserved interaction of the Arginine 188 side chain with the backbone of the adenine nucleotide at position 3 (Figure 4B and Supplementary Figure S5C). It indicates that the two forms do not reflect two independent RNP folds, but are likely to be interchanging with each other in solution. This characteristic underlines the special feature of relatively labile CDE SL folds, which are likely to be stabilized by the ROQ domain.

For both complexes, we found a conserved network of interactions between protein and RNA, reflecting the previously described contacts between ROQ and the *Tnf* CDE (Figure 4A and B). Consequently, the overall fold of all three CDEs in the bound form is highly comparable underlining the earlier suggestion that ROQ recognizes its target RNAs through shape- rather than sequence-specific interactions and locks them in a canonical arrangement that is unique for CDEs (Supplementary Figure S5D). Most remarkably, all three CDEs shown in the overlay resemble an identical orientation of loop nucleotides, which is not reflected by the respective free forms (Supplementary Figure S5E and F). Previous work on the *Tnf* CDE has shown the role of CDE-intrinsic sequence features for the recognition by the ROQ domain (38,43,44). Comparison of the free and ROQ-bound *Tnf* CDE also suggested that major rear-

rangements of the loop nucleotides are necessary for complex formation (Supplementary Figure S5G). In line with that, comparing all individual unbound RNAs of our study with their respective conformations when bound to ROQ suggests, that the ROQ domain either provokes an induced-fit mechanism during engagement with the three CDEs, or that all of them are able to sample the final fold in solution (conformational selection).

Previous data showed a mutual stabilization of both the ROQ domain and CDE RNA upon complex formation (38), which was suggested to be a driving force for the high-affinity interaction between them and underlines the central role of CDE elements for Roquin-mediated post-transcriptional regulation (36,38,40–43,46). Using EMSA experiments, we previously estimated 100 nM affinity between the ROQ domain and the wild type *UCP3* tandem CDE-AREs (44) (Figure 1B). Mutational analysis suggested similar contributions from either *UCP3* CDE, with affinities in the low micromolar range. This affinity is in line with a CDE found in the *Ox40* 3'-UTR (39), but considerably higher than measured for the *Tnf* CDE (38,43). However, the measurements of the single CDEs were performed in the context of the *UCP3* tandem RNA, with possible interference from the surrounding sequence context, due to transient interactions and thus weakening of the CDE fold. In order to quantify the individual binding affinity of the *UCP3* CDEs, we used ITC and determined K_D values for ROQ binding to CDE1GG and CDE2GG. The resulting K_D s show tight, low-nanomolar binding, comparable to that of the *Tnf* CDE (Figure 4C). This reveals two remarkable features: First, these affinities are in full agreement with the ROQ-*UCP3* ARE complex arrangement in a CDE-like manner as proven by the NMR data and crystal structures. Importantly, they do not support binding of the CDE-AREs as linear sequence, as ssRNA recognition by the ROQ domain occurs with 2–3 orders of magnitude lower affinity (45). Previous studies on CDEs and ADEs suggested high on-rates during complex formation as one cause for the high affinity (39). From the affinities observed, we here assume a similar kinetic basis, based on preformed SL RNAs. Second, the low K_D values suggest a functionally relevant and active recognition of structured ARE elements by a single high-affine RBD, the ROQ domain. The latter is remarkable as the high-affine recognition of AREs had so far only been described for single-stranded ARE species together with the concerted action of multiple RBDs (17,67). In the case of AUF1 for example, both RRM are necessary for and contribute equally to high affinity binding (68).

Biophysical investigation of *UCP3* CDE-ARE-ROQ complexes

Next, we analyzed a potential CDE-stabilizing effect of the Roquin ROQ domain. Such stabilization had been suggested for the *Tnf* CDE before (38), while no clear experimental proof was provided. Thus, for the *UCP3* CDEs, we first assessed differences in stability between the free *UCP3* CDEs and their complexed form with ROQ in CD-melting experiments (Figure 5A). All SLs showed comparable and characteristic CD spectra with a maximum ellipticity in the range of wavelengths between 262 and 265 nm and a mini-

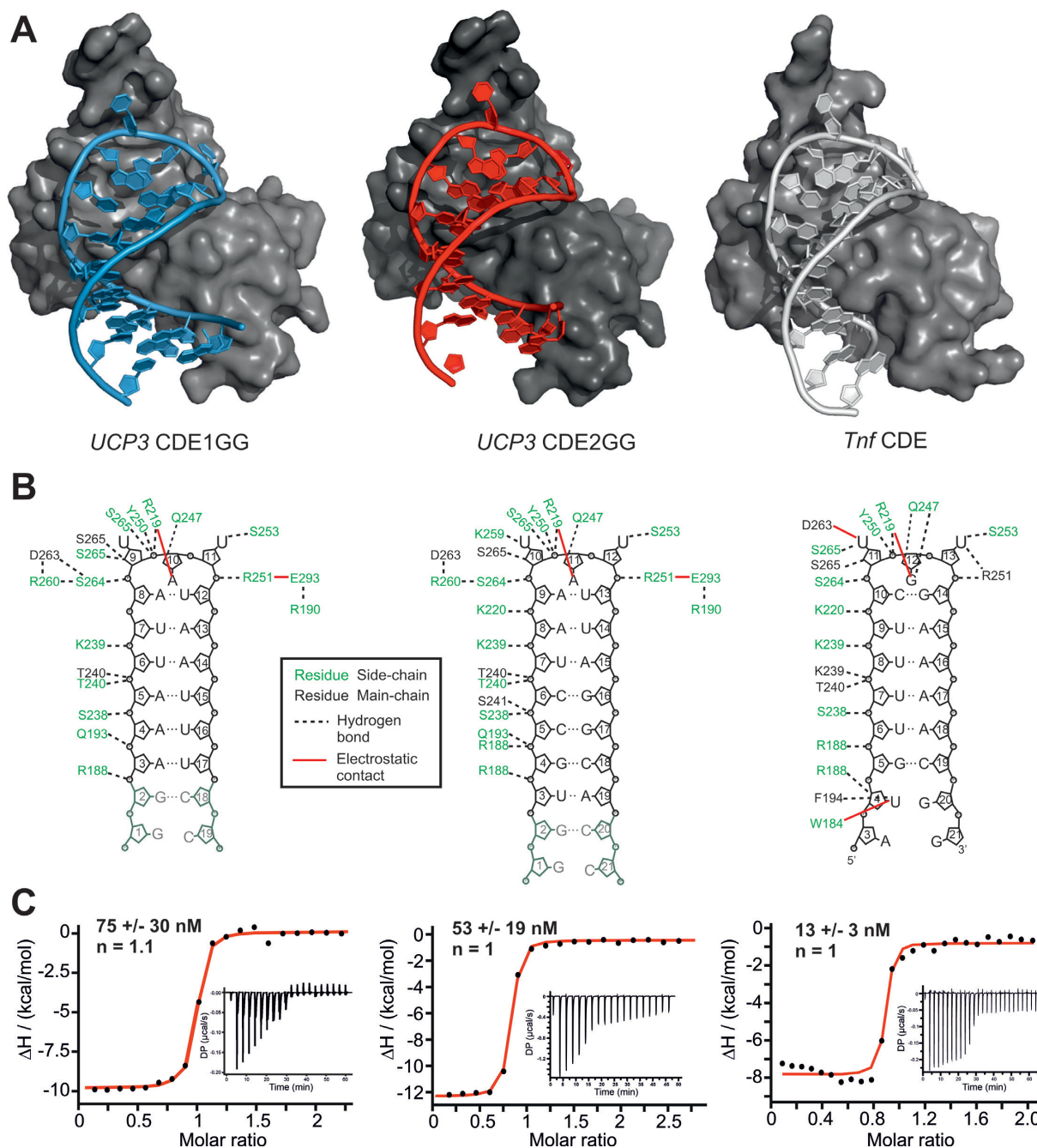


Figure 4. Crystal structures of CDE1GG and CDE2GG in complex with the ROQ domain. (A) Overall structures of ROQ (grey surface) in complex with CDE1GG (blue cartoon) and CDE2GG (red cartoon). For comparison, the same view of ROQ in complex with the *Tnf* CDE is included (grey cartoon) (38). (B) Internal contact map of ROQ domain residues with *UCP3* CDE1GG and CDE2GG in panel (A). For comparison, the same is shown for ROQ in complex with the *Tnf* CDE (38). (C) Representative ITC curves of ROQ titrated with the RNAs as shown in (B) reveal binding affinities comparable to the *Tnf* CDE. K_D values are mean +/- standard deviation from three independent experiments. See also Supplementary Figure S5.

num between 235 and 245 nm (Supplementary Figure S6). As expected, melting points (T_m) largely differed between stabilized and non-stabilized CDEs, owing to the two stabilizing GC base pairs and rationalizing them for structure determination of free RNAs. In line with that, we also found a remarkable difference in melting temperatures of CDE1 (27°C) versus CDE2 (69°C) and CDE1GG (52°C) versus CDE2GG (82°C) (Figure 5A), where in the latter the effect

of stabilization is less pronounced due to the intrinsic two GCs base pairs in CDE2 variants (compare RNA sequences in Figure 1B). This finding underlines the drastic lability of pure ARE sequences, but suggests at the same time an inherent potency to be also recognized in their linear form.

The T_m of ROQ alone is 42°C (Supplementary Figure S6) and thus below the melting points of CDE2 and the stabilized CDE variants. We thus used CDE1 to examine the

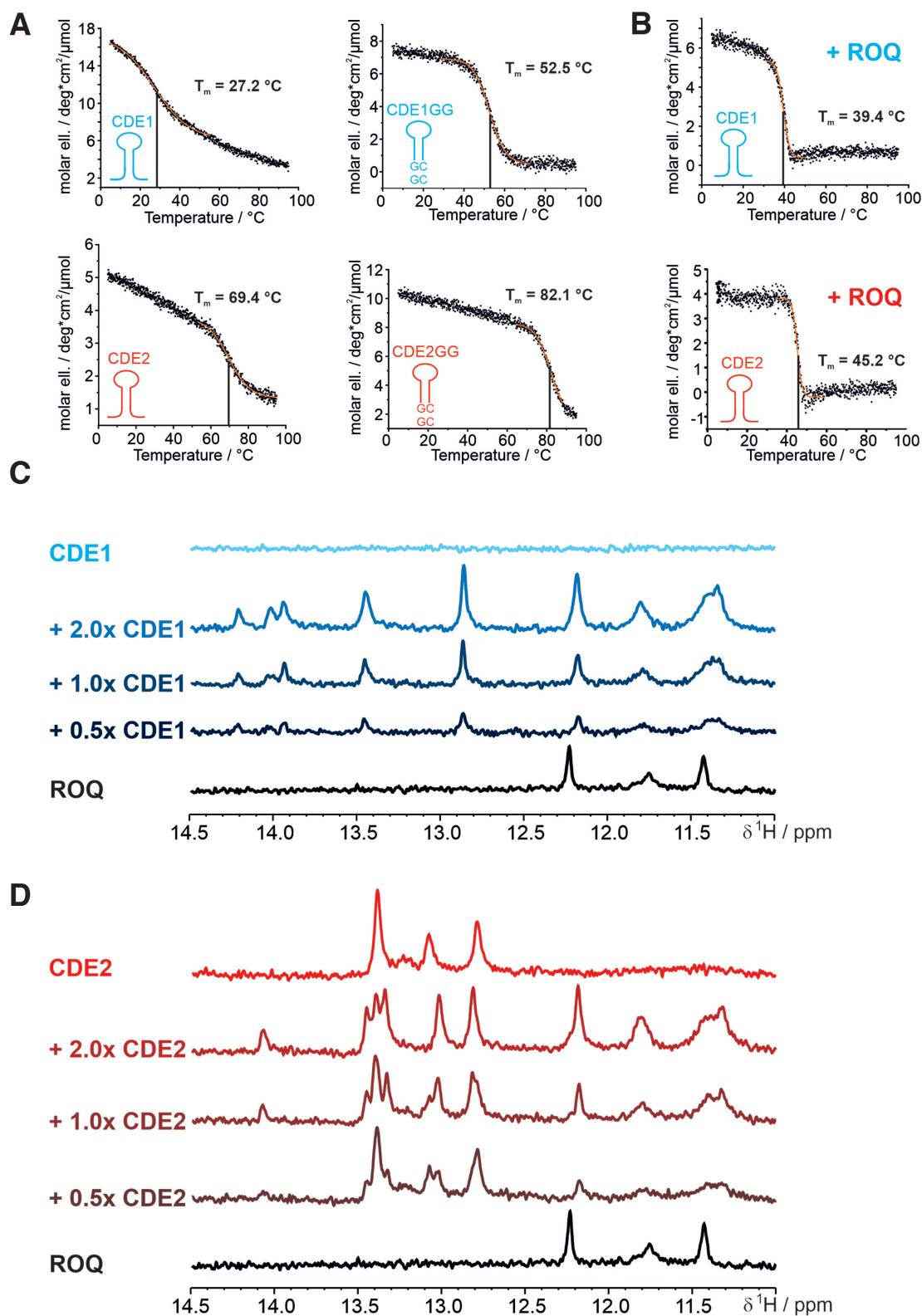


Figure 5. The ROQ domain locks the *UCP3* CDE-AREs in the CDE-fold, independent of their thermodynamic instability. (A and B) CD melting curves of CDE1 and CDE2, their respective stabilized constructs (GG) (A) and after addition of equimolar amounts of ROQ domain (B). Respective sigmoidal fitting regions are indicated in orange. Melting points and their errors are indicated as obtained from the fitting procedure. (C and D) Imino proton spectra of the non-stabilized CDE1 (C) and CDE2 (D) at room temperature alone and in complex with different ratios of ROQ domain. As a control, the same region is shown for the ROQ domain alone in order to indicate protein-derived signals. See also Supplementary Figure S6.

role of the ROQ domain in CDE stabilization. We observed a strong stabilizing effect of ROQ on CDE1 by CD spectroscopy (Figure 5B) evidenced by an increase of T_m from 27°C to 39°C. We thus reason, Roquin captures and stabilizes the transient SL of the pure-AU CDE1 at physiological conditions, while the *cis*-element alone will be mainly unfolded.

To further verify the stabilizing role of ROQ towards the *UCP3* CDEs at atomic level, we tested the ability of the ROQ domain to shift the equilibrium of unfolded CDE1 and CDE2 to the folded species in dependence of protein-RNA ratios (Figure 5C and D). To this end, we used the detectability of imino protons in SL RNAs, which indicate H-bonding of base-paired nucleotides, i.e. structured RNA (Supplementary Figure S1). Indeed, both wild type CDE1 and CDE2 revealed a significant increase in imino proton resonances in response to the presence of the ROQ domain. This suggests formation of stable ROQ-CDE1 and -CDE2 complexes with RNA-structure lifetimes observable during the NMR experiment (here a few hundred milliseconds) that clearly exceed the lifetime of imino protons in the free RNAs. For CDE2, we also observed small changes in imino proton chemical shifts between the apo RNA and the 1:1 complex with the ROQ domain (Figure 5D). We thus conclude that the ROQ domain, upon binding, induces the formation and promotes the stabilization of the CDE-fold in the wild type *UCP3* 3'-UTR. Surprisingly, this feature appears to be independent from the obvious differences in intrinsic stabilities between the two *UCP3* SLs as determined by CD spectroscopy.

This observation suggests that tight binding to the two CDEs is additionally controlled by a low off-rate. Such a binding mechanism mainly requires the apical AUAUU pentamer, while the actual stem sequence is not of apparent relevance, presuming a general stem formation. This is in line with the completely sequence independent consensus for active Roquin binding sites suggested by mutational analyses of the *UCP3* CDE-AREs (44).

Independent recognition of the two *UCP3* CDEs by Roquin

The accumulated data suggest simultaneous and highly similar binding of ROQ to the two *UCP3* CDEs in their endogenous context. To verify that, we made use of a tandem *UCP3* RNA construct (CDE1-2, Supplementary Table S1) that harbors both CDEs in close sequential proximity (Figure 1B). To prevent dimerization observed at high RNA concentrations with the wild type sequence (Supplementary Figure S7A), a point mutation had to be introduced in the linker region between the two CDEs. Mutations in the linker region were shown not to affect mRNA repression *in vivo* (44). In order to ensure that the tandem construct used for NMR studies retains full repressive capacity, it was tested in a luciferase reporter assay. As expected the point mutation had no significant effect on the regulation *in vivo* (Supplementary Figure S7B). Thus, the construct is suitable for further analyses.

Earlier data had given good evidence for a 1:2 complex between the *UCP3* RNA and ROQ domain (44). Consequently, to directly monitor the formation of comparable complexes (compare also Figure 2) at atomic resolution we

titrated tandem *UCP3* to the ROQ domain and recorded HSQC spectra (Supplementary Figure S2). Excess RNA ($>0.5\times$ with respect to the number of available CDEs) allows the interpretation of detectable complex resonances. The overall occupancy of CDEs showed a 1:1 distribution, which indicates that ROQ does not discriminate between the two CDEs for the formation of tight complexes. Consequently, in the tandem complex, both CDE-complexes appear with a comparable lifetime, indicating similar complex stability. However, we find a number of resonances that suggest a slight preference for CDE2, which is in line with its relatively higher intrinsic stability measured by CD. Strikingly, the two SLs behave fully independent as evidenced by the unaffected chemical shifts of individual imino protons (Supplementary Figure S1).

Competition of RBPs for AU-rich elements

The NMR and CD data clearly demonstrate that ROQ binding can rigidify the tandem-CDE regulatory hub within the 3'-UTR of the *UCP3* mRNA. However, in the absence of Roquin binding, this region will be largely unstructured and thus accessible for RBPs that specifically recognize linear AREs. Thus, Roquin will be able to compete with these RBPs by removing their ssRNA target sites in the 3'-UTR, while conversely these RBPs will be able to compete with Roquin by masking the AREs and preventing them from adopting a CDE-fold.

To investigate competition for identical ARE-foldamers by two RBPs we included the well-studied protein AUF1 (hnRNP D) into the Roquin-*UCP3* interactome (Figures 1A and 6A). The high affinity of AUF1 to AU-rich sequences has already been described (69), and such sequences are present in the studied *UCP3* tandem CDE (Figure 6A). Additionally, previous work has shown that structure formation in AREs abolished binding and mRNA regulation by AUF1 (30,31). AUF1 thus represents a promising study case for the regulatory capacity that centers on dynamic CDE-AREs. The AUF1 protein comprises two canonical RRM domains, arranged in tandem that account for RNA-binding (69). With NMR, we found that the two AUF1 RRM domains are independent from each other, as shown by a ^{15}N -HSQC overlay comparing tandem RRM1-2 and RRM2. Here, the single RRM2 domain provides resonances that are in full overlay with the respective resonances in the tandem RRM1-2 indicating that the two RRMs are structurally independent from each other with respect to their two-site RNA-binding capacity (Supplementary Figure S8A). As a prerequisite for the following experiments, we also ruled out an interaction of AUF1 RRMs with the Roquin ROQ domain (Supplementary Figure S8B).

We next tested whether AUF1 is capable of binding to the tandem *UCP3* RNA using RNA immunoprecipitation. Notably, AUF1 isoforms can readily be isolated from cell lysates using the wild type *UCP3* tandem CDE (Figure 6B). Further, overlays of ^{15}N -HSQC spectra confirm the interaction of the two AUF1 RRMs with both individual *UCP3* CDEs and the tandem construct. The spectra show large chemical shift perturbations accompanied by line broadening which is in line with the nanomolar affinities that had

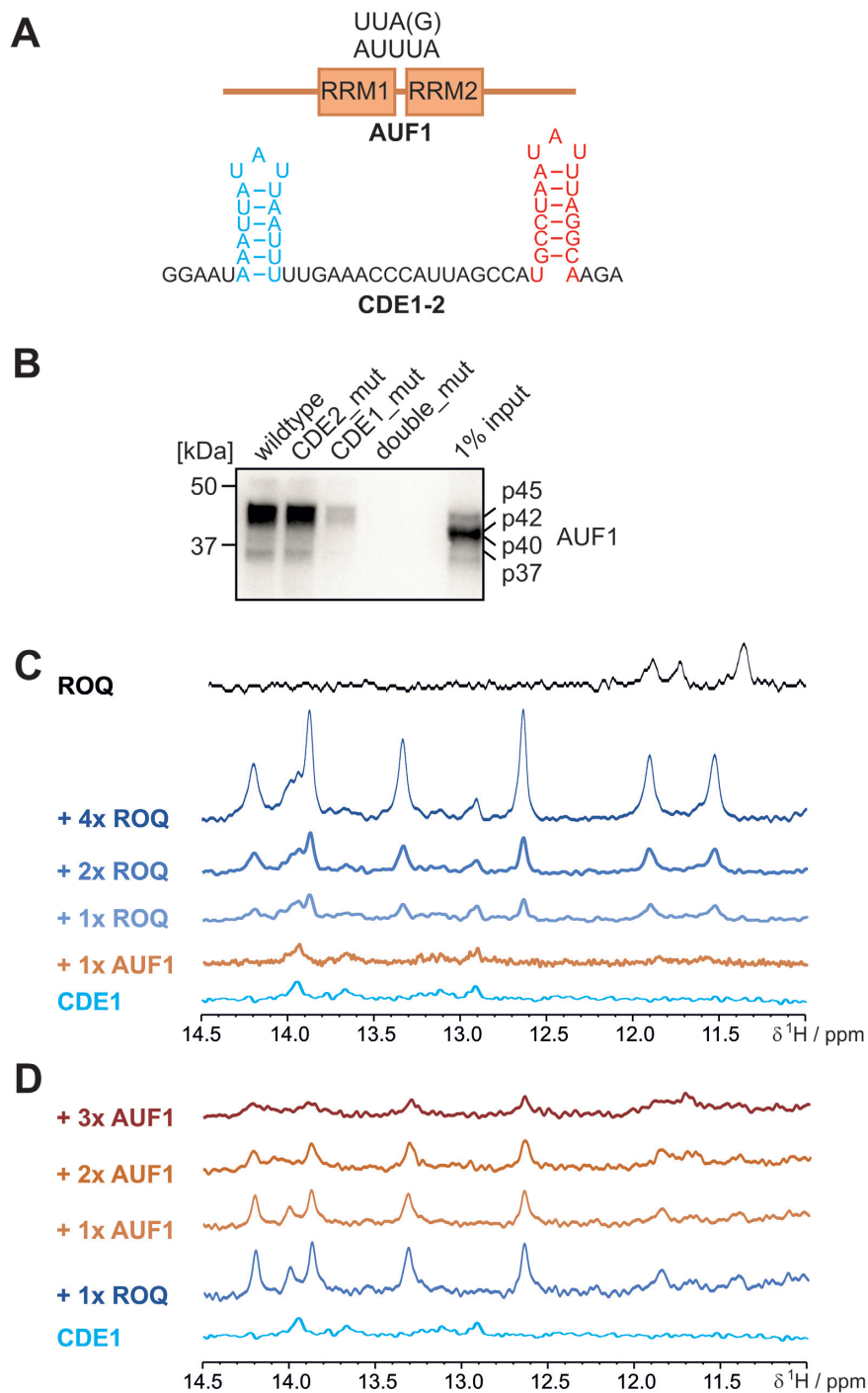


Figure 6. Competition of Roquin with AUF1 for the *UCP3* ARE hub. (A) Top, schematic presentation of AUF1 domain arrangement with approximate position of the tandem RRM1-2 domains. Below that, the wild type *UCP3* ARE hub is shown as used in this study and for the pulldown experiment in panel (B). (B) Analysis of Roquin binding to the *UCP3* ARE hub. For RNA affinity purification, HEK293 whole cell lysates were incubated with the indicated *UCP3* RNAs. AUF1 was visualized by Western blot using anti-AUF1 antibody. (C and D) NMR-based *in vitro* competition experiments between the Roquin ROQ domain and the AUF1 tandem RRM1-2. (C) Imino proton spectra of WT *UCP3* CDE1 alone, in a 1:1 complex with AUF1 and after addition of increasing amounts of ROQ. (D) The same as in (C), but starting from a 1:1 complex of CDE1 with ROQ and subsequent titration of increasing amounts of AUF1 RRM1-2. For panels (C) and (D), the same spectral region is shown for ROQ alone for clarity towards protein-derived peaks. See also Supplementary Table S1 and Supplementary Figure S8.

been reported for AUF1-ARE interactions earlier (70). As expected for canonical RRM domains, binding is primarily mediated by the intrinsic RNP motifs (Supplementary Figure S8C–E).

Interestingly, mutants of the RNA within CDE1 or 2 of the tandem *UCP3* RNA revealed a significantly more pronounced loss in binding for CDE1, which is solely composed of AU residues (Figure 6A and B). This is in line with AUF1's primary preference for extended AU-rich regions (71,72) and suggests a slight preference for the CDE1 region within the tandem RNA construct.

We thus decided to monitor the presence of imino protons in CDE1 as a measure of the RBP-competition for different RNA-foldamers. The wild type CDE1 did not show a significant proportion of structure on the NMR time scale (Figure 5C). Addition of AUF1 to the wild type CDE1 yielded no visible increase in RNA structure (Figure 6C), in line with the previously described target specificity of AUF1 for linear AREs (70). We next added increasing stoichiometric amounts of Roquin ROQ domain to the mixture. As expected, we observed a stepwise increase in imino signals indicating the stabilization of the CDE-fold in a ROQ-RNA complex (compare to Figure 5). Taking into account the high affinity interaction of AUF1 RRMs with CDE1 (Supplementary Figure S8), it suggests that AUF1 has to be occluded from the RNA-interface in favor of the ROQ domain. ROQ-binding sequesters the former linear ARE and thereby excludes AUF1 from its potential regulatory function on dynamic AREs. To exclude the potential scenario of both RBPs being bound to CDE1, we next started from a pre-formed ROQ-CDE1 complex and added increasing stoichiometric amounts of AUF1 RRM1-2 (Figure 6D). As expected, the increase in AUF1 led to a stepwise loss of imino resonances in CDE1 indicating that AUF1 is able to replace Roquin, and thus unfolds the CDE1.

To estimate the occurrence and folding of further CDE-AREs *in vivo*, we analyzed whether AU-pure CDEs overlap with binding sites of AUF1 derived from PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) data in HEK293 cells (22). A total of 146 AU-pure CDEs with a UAU tri-loop, i.e. analogous to CDE1, were predicted in the 3'-UTRs of 128 human genes (44). Of these, 48 genes (38%), containing 55 AU-pure CDEs (38%) in their 3'-UTRs, are recognized by AUF1 (Supplementary Table S3). Although these 48 genes together harbor 119 binding sites, only 4 binding sites (3%) overlap with CDE-AREs. Importantly, most of the genes targeted by AUF1 are also bound by Roquin-1 in HEK293 cells (37 genes, 77%), as shown by PAR-CLIP (42). Together, the data suggest that CDE-AREs are widespread and most often excluded from AUF1 binding, likely due to their structured nature and/or subsequent occupation by competing *trans*-acting factors. Interestingly, the four CDE-AREs bound by AUF1 were also recognized by the mRNA stabilizing factor HuR (22), thus further extending the landscape of possible protein interactors.

In summary, the data suggest competitive binding between Roquin ROQ and the AUF1 RRMs to the same *cis*-regulatory RNA sequence, while targeting different folding states and actively interfering with the equilibrium between the two forms.

DISCUSSION

The ROQ domain recognizes *UCP3 cis*-regulatory AREs in a pre-formed CDE-like manner

While AREs and CDEs had long been conceived as separate entities, functioning as parallel, redundant or eventually cooperative *cis*-elements (29,34,46), we have recently shown that AREs can be functioning as CDEs (44) and consequently to be regulated by the protein Roquin. Their low intrinsic half-lives as SLs suggested that they could fulfil additional roles as linear *cis*-elements. However, an unambiguous, i.e. structural proof, for the existence of CDE-shaped AREs and their recognition by Roquin in a CDE-like manner had remained open.

We show here that the two adjacent Roquin-regulatory *cis*-elements in the 3'-UTR of the *UCP3* mRNA indeed form CDE-like folds. This holds true for the CDEs alone (Figure 3) and in complex with the protein (Figure 4). We also show that in solution, Roquin engages with these CDEs in a manner identical to the binding of the canonical *Tnf* CDE (38) and with similar, low nanomolar affinities. Thus, our data prove the previously described functional role for the *UCP3* SLs as CDEs. Moreover, they generally suggest CDE-AREs as a subclass of CDEs that allow for unique regulatory features. Such features are particularly obvious for the pure-AU CDE1 of *UCP3*, which did not reveal base pairing in its free form, suggesting it to be unfolded in average on the NMR-observed time scale. Notably, we found CDE1 to be stabilized in its folded form by the Roquin protein, with a melting temperature of 27°C in its free form, while ROQ domain binding increases CDE1 stability to 39°C. We speculate that this narrow, physiologically relevant temperature window for stabilization by an RBP expresses the functional dynamics of interactions between Roquin and transiently folded AREs. Under physiological conditions, those CDE-AREs are in constant exchange with the unfolded form, and Roquin is replaced by competing RBPs through the adjustment of relative protein levels.

For Roquin, it has been discussed that RNA binding requires a certain RNA flexibility, while the stable ROQ-CDE complex is of a highly conserved geometry. Interestingly, this geometry is also present in ROQ-ADE complexes (39). It clearly highlights the shape-specificity of Roquin for RNA, rather than a particular sequence requirement. We found that both *UCP3* CDEs in complexes with the ROQ domain reflect an identical RNP arrangement, which is furthermore identical to all other ROQ-CDE structures (38,40,73–75). Interestingly, other than for the *Tnf* CDE, the two stabilized *UCP3* CDEs adopt in their free forms a fold, which is reminiscent of the ROQ-bound form. It remains open, whether this pre-arranged element supports their interaction with ROQ and if it is a particular feature of the *UCP3* CDEs.

Taking into account the labile nature of free *UCP3* CDEs (Supplementary Figure S1), we assume that complex stability is mainly driven by a low off-rate, where the ROQ domain locks CDEs in a compact structure. Ever since the discovery of CDEs, their intrinsic instability has been suggested to account for their regulatory potency. Indeed, former work has shown that GCs are by default underpre-

sented in CDEs (36) and a higher GC content does not *per se* increase affinity for ROQ, as e.g. seen for the CDE-like SL in the *Ox40* 3'-UTR (38,39). This is also supported by the U-rich SLs found in a PAR-CLIP study that are recognized by Roquin and show an almost exclusive content of AUs in their stems (42). In fact, the role of intrinsic lability for enabling an induced-fit with the ROQ domain was correlated with a preference for polyA-stretches along the 3'-side of the stem, i.e. this unique geometry was suggested to be a driving force of high-affine binding to Roquin as shown for the *Tnf* CDE (38). However, we recently showed that these stretches are no requirement for active CDEs (44).

As contradictive findings from mutational data on CDEs have been reported (36,38,43,44), our data essentially suggest binding to be facilitated by the pure possibility that a potential CDE can be formed *in vivo*. In principle, a plethora of short (~20 nt) sequences will be capable of adopting CDE-like folds. Accordingly, hundreds of low-populated, potential CDEs have been predicted in a genome-wide bioinformatic screen (44). Despite their potential to form a stable complex with Roquin, they were thought to be mainly unfolded at physiological conditions or to be hidden in a larger RNA secondary structure context. Nevertheless, such sequences might also be functional Roquin targets, when they transiently adopt a CDE-like fold.

In sum, we present the first pure-AU CDE that resembles a *bona-fide* CDE-like fold and thus extend the possible *cis*-element target group of Roquin for mRNA regulation on a structural basis. Indeed, while all previous studies have missed to provide clear structural evidence for a CDE-like fold of AREs, we here unambiguously prove the existence of such CDE-AREs in the free and ROQ-bound form with high resolution. However, while we assume a gain in complex stability for both RNA and ROQ, the question remains whether the rate-limiting step *in vivo* will be the prior formation of CDE-folded SLs; assuming a constant pressure against RNA unfolding (76) and the competition with miRNAs and proteins that favor linear AREs (77).

Opposing specificity for *UCP3* ARE conformers by the Roquin ROQ domain vs. AUF1 RRM

Our experiments show such a competition between Roquin and AUF1 for the same ARE *in vitro*, targeting either the folded or the linear form, respectively. As such, we suggest a mutually exclusive, i.e. functionally competitive or redundant action of RBPs on AREs with transient structure like the herein described CDE-AREs. Notably, the AUF1 isoform p45 is capable of dissolving viral SL structures, thus promoting genome replication of flaviviruses (78,79), and we found this isoform to be most effectively enriched after pulldown with the *UCP3* CDEs (Figure 6). At the same time, an increase in RNA structure at its target *cis*-element environment clearly interferes with regulation of mRNA fate by the AUF1 isoform p37 (30). Strikingly, ARE structure formation was shown to differentially affect *trans*-factor binding (31), indicating a complex interplay of *cis-trans* interactions for mRNA regulation depending on mRNA structure and the exact preferences of regulatory RBPs.

Two scenarios are thus to be considered for post-transcriptional regulation: First, the relative proportions of RNA forms will decide about the accessibility for the respective RBPs. Herein, the interconversion between folded and unfolded RNA is a decisive parameter for the possible RNPs to be formed. Second, relative affinities and amounts of available protein will decide about the occupancy of AREs and their tight stabilization in the linear or folded forms, respectively. The individual AUF1 RRMs have an affinity of 300–800 nM for AREs, depending on the exact RNA sequence and protein isoform (69). Those numbers are in the identical order of affinity with Roquin binding to CDEs and ADEs (Figure 4C and (38,39)). Consequently, mRNA fate will depend on the relative RBP levels and their local availability. As shown in Figure 6, replacement of ROQ requires an excess of AUF1 on the investigated time scale. This is consistent with the finding that potentially folded CDE-AREs are underrepresented as binding sites of AUF1. In the case of counter-acting RBPs that compete for (CDE-)AREs, folding probability, affinity and availability of the *trans*-acting factors represent critical contributions and regulatory levels for individual fates of target mRNAs.

In immune-regulatory proteins like Regnase and Roquin, high-affine and specific binding of structured decay elements is accompanied by the recognition of adjacent AREs through the co-existing zinc finger domains. Despite the discussable gain in specificity and affinity (42,45,80), the role of ARE recognition by the Roquin ZnF has appeared to be clearly subordinated to the function of the ROQ domain after the discovery of CDE and ADE elements (36,39,46), with the major role for the ROQ domain in mRNA regulation (38,40). Our findings now provide an additional option, in that the Roquin ROQ domain would also account for mRNA regulation through pure AU-rich sites, which can fold into CDE-shaped SLs. This would increase the likelihood of Roquin to engage with its mRNA target regions when either using its ZnF or ROQ domains for identical AU-rich *cis*-elements.

Interestingly, genome-wide bioinformatic analyses predicted >100 CDEs consisting exclusively of AUs (see Supplementary Table S3). Moreover, several of these show evolutionary conservation, highly suggestive of further instances of CDE-AREs (44). This might explain possible target overlap with proteins that do not apparently recognize CDEs. In this context, recent work has also presented the novel RBP Arid5a that appears to counteract Roquin and Regnase binding, possibly by destabilization or refolding of SL elements (81,82). Further, Roquin recognition of an SL element in the *Pten* 3'-UTR competes with miRNA binding (83). Since miRNAs recognize ssRNA sequences by direct base pairing, at least partial melting of the *Pten* SL would be necessary.

Our findings pose the question whether also other *cis*-elements recognized by sequence-specific proteins in their linear form, might be structured more often and thus could be bound by RBPs with shape-specific domains. Strikingly, considering *cis*-element accessibility significantly increases the prediction of *in vivo* binding sites of RBPs, compared to predictions based only on sequence preferences (84). This suggests that linear motifs are often masked within RNA

folds, limiting recognition by *trans*-acting factors that require single strand regions. There are >1000 proteins encoded in the human genome that are capable of interacting with mRNA, while binding preferences for the vast majority of them are unknown (85,86). Likely, the herein presented findings of protein competition for differently folded forms of the same ARE can be extended to other *cis*-elements that have so far been regarded as purely linear acting motifs. In essence, our data imply a role for transiently folded *cis*-elements in mRNA regulation. Our work highlights how *cis*-elements are used as signal-integrative hubs, and mRNA regulation is a consequence of *cis*-element availability/accessibility and the relative abundance of competitive RBPs.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures were deposited in the Protein Data Bank under the accession numbers 6TQB and 6TQA for the ROQ domain with UCP3 CDE1GG and CDE2GG, respectively. UCP3 CDE1GG and CDE2GG NMR structural ensembles were deposited under the accession numbers 6XXW and 6XJW in the Protein Data Bank and under IDs 34483 and 34482 in the BMRB, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the Frankfurt BMRZ NMR facility for excellent support. We are grateful to Michael Sattler (Technical University and Helmholtz Centre Munich) and the Bavarian NMR centre for the possibility to perform some of the NMR and ITC experiments.

FUNDING

Deutsche Forschungsgemeinschaft [SFB902/B14, WE5819/3-1 to J.E.W., SFB902/B16, SCHL2062/2-1 to A.S., SFB902/A1 to H.S.]; Johanna Quandt Young Academy at Goethe [2019/AS01 to A.S.]. Funding for open access charge: Deutsche Forschungsgemeinschaft. *Conflict of interest statement.* None declared.

REFERENCES

- Baumjohann,D. and Heissmeyer,V. (2018) Posttranscriptional gene regulation of T follicular helper cells by RNA-Binding proteins and microRNAs. *Front. Immunol.*, **9**, 1794.
- Yoshinaga,M. and Takeuchi,O. (2019) Post-transcriptional control of immune responses and its potential application. *Clin Transl Immunol.*, **8**, e1063.
- Beisang,D. and Bohjanen,P.R. (2012) Perspectives on the ARE as it turns 25 years old. *Wiley Interdiscip. Rev RNA*, **3**, 719–731.
- Caput,D., Beutler,B., Hartog,K., Thayer,R., Brown-Shimer,S. and Cerami,A. (1986) Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 1670–1674.
- Chen,C.Y. and Shyu,A.B. (1994) Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol. Cell. Biol.*, **14**, 8471–8482.
- Kruys,V.I., Wathelet,M.G. and Huez,G.A. (1988) Identification of a translation inhibitory element (TIE) in the 3' untranslated region of the human interferon-beta mRNA. *Gene*, **72**, 191–200.
- Jones,T.R. and Cole,M.D. (1987) Rapid cytoplasmic turnover of c-myc mRNA: requirement of the 3' untranslated sequences. *Mol. Cell. Biol.*, **7**, 4513–4521.
- Shaw,G. and Kamen,R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659–667.
- Wilson,T. and Treisman,R. (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature*, **336**, 396–399.
- Fallmann,J., Sedlyarov,V., Tanzer,A., Kovarik,P. and Hofacker,I.L. (2016) AREsite2: an enhanced database for the comprehensive investigation of AU/GU/U-rich elements. *Nucleic Acids Res.*, **44**, D90–D95.
- Khabar,K.S. (2017) Hallmarks of cancer and AU-rich elements. *Wiley Interdiscip. Rev. RNA*, **8**, e1368.
- Bolognani,F. and Perrone-Bizzozero,N.I. (2008) RNA-protein interactions and control of mRNA stability in neurons. *J. Neurosci. Res.*, **86**, 481–489.
- Jing,Q., Huang,S., Guth,S., Zarubin,T., Motoyama,A., Chen,J., Di Padova,F., Lin,S.C., Gram,H. and Han,J. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell*, **120**, 623–634.
- Ma,F., Liu,X., Li,D., Wang,P., Li,N., Lu,L. and Cao,X. (2010) MicroRNA-4661 upregulates IL-10 expression in TLR-triggered macrophages by antagonizing RNA-binding protein tristetraprolin-mediated IL-10 mRNA degradation. *J. Immunol.*, **184**, 6053–6059.
- von Roretz,C., Di Marco,S., Mazroui,R. and Gallouzi,I.E. (2011) Turnover of AU-rich-containing mRNAs during stress: a matter of survival. *Wiley Interdiscip. Rev. RNA*, **2**, 336–347.
- Tran,H., Maurer,F. and Nagamine,Y. (2003) Stabilization of urokinase and urokinase receptor mRNAs by HuR is linked to its cytoplasmic accumulation induced by activated mitogen-activated protein kinase-activated protein kinase 2. *Mol. Cell. Biol.*, **23**, 7177–7188.
- Wang,X. and Tanaka Hall,T.M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.*, **8**, 141–145.
- Ma,W.J., Chung,S. and Furneaux,H. (1997) The Elav-like proteins bind to AU-rich elements and to the poly(A) tail of mRNA. *Nucleic Acids Res.*, **25**, 3564–3569.
- Brewer,G. (1991) An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro. *Mol. Cell. Biol.*, **11**, 2460–2466.
- White,E.J., Matsangos,A.E. and Wilson,G.M. (2017) AUF1 regulation of coding and noncoding RNA. *Wiley Interdiscip. Rev. RNA*, **8**, doi:10.1002/wrna.1393.
- Wilson,G.M., Sun,Y., Lu,H. and Brewer,G. (1999) Assembly of AUF1 oligomers on U-rich RNA targets by sequential dimer association. *J. Biol. Chem.*, **274**, 33374–33381.
- Yoon,J.H., De,S., Srikantan,S., Abdelmohsen,K., Grammatikakis,I., Kim,J., Kim,K.M., Noh,J.H., White,E.J., Martindale,J.L. et al. (2014) PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat. Commun.*, **5**, 5248.
- Barreau,C., Paillard,L. and Osborne,H.B. (2005) AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.*, **33**, 7138–7150.
- Gronlund,G.R. and Ramos,A. (2017) The devil is in the domain: understanding protein recognition of multiple RNA targets. *Biochem. Soc. Trans.*, **45**, 1305–1311.
- Schneider,T., Hung,L.H., Aziz,M., Wilmen,A., Thaum,S., Wagner,J., Janowski,R., Müller,S., Hüttelmaier,S., Niessing,D. et al. (2019) Combinatorial recognition of clustered RNA elements by a multidomain RNA-binding protein, IMP3. *Nat. Commun.*, **10**, 2266.
- Schlundt,A., Tants,J.N. and Sattler,M. (2017) Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods*, **118–119**, 119–136.
- Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 479–490.

28. Bohjanen, P.R., Petryniak, B., June, C.H., Thompson, C.B. and Lindsten, T. (1992) AU RNA-binding factors differ in their binding specificities and affinities. *J. Biol. Chem.*, **267**, 6302–6309.
29. Paschoud, S., Dogar, A.M., Kuntz, C., Grisoni-Neupert, B., Richman, L. and Kuhn, L.C. (2006) Destabilization of interleukin-6 mRNA requires a putative RNA stem-loop structure, an AU-rich element, and the RNA-binding protein AUF1. *Mol. Cell. Biol.*, **26**, 8228–8241.
30. Wilson, G.M., Sutphen, K., Chuang, K. and Brewer, G. (2001) Folding of A+U-rich RNA elements modulates AUF1 binding. Potential roles in regulation of mRNA turnover. *J. Biol. Chem.*, **276**, 8695–8704.
31. Fialcowitz, E.J., Brewer, B.Y., Keenan, B.P. and Wilson, G.M. (2005) A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. *J. Biol. Chem.*, **280**, 22406–22417.
32. Vinuesa, C.G., Cook, M.C., Angelucci, C., Athanasopoulos, V., Rui, L., Hill, K.M., Yu, D., Domasch, H., Whittle, B., Lambe, T. *et al.* (2005) A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature*, **435**, 452–458.
33. Matsushita, K., Takeuchi, O., Standley, D.M., Kumagai, Y., Kawagoe, T., Miyake, T., Satoh, T., Kato, H., Tsujimura, T., Nakamura, H. *et al.* (2009) Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *Nature*, **458**, 1185–1190.
34. Fu, M. and Blakeshear, P.J. (2017) RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nat. Rev. Immunol.*, **17**, 130–143.
35. Mino, T., Murakawa, Y., Fukao, A., Vandenberg, A., Wessels, H.H., Ori, D., Uehata, T., Tartey, S., Akira, S., Suzuki, Y. *et al.* (2015) Regnase-1 and Roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms. *Cell*, **161**, 1058–1073.
36. Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M.C. and Stoecklin, G. (2013) Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell*, **153**, 869–881.
37. Wilamowski, M., Gorecki, A., Dziedzicka-Wasylewska, M. and Jura, J. (2018) Substrate specificity of human MCP1P1 endoribonuclease. *Sci. Rep.*, **8**, 7381.
38. Schlundt, A., Heinz, G.A., Janowski, R., Geerloff, A., Stehle, R., Heissmeyer, V., Niessing, D. and Sattler, M. (2014) Structural basis for RNA recognition in Roquin-mediated post-transcriptional gene regulation. *Nat. Struct. Mol. Biol.*, **21**, 671–678.
39. Janowski, R., Heinz, G.A., Schlundt, A., Wommelsdorf, N., Brenner, S., Gruber, A.R., Blank, M., Buch, T., Buhmann, R., Zavolan, M. *et al.* (2016) Roquin recognizes a non-canonical hexaloop structure in the 3'-UTR of Ox40. *Nat. Commun.*, **7**, 11032.
40. Tan, D., Zhou, M., Kiledjian, M. and Tong, L. (2014) The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nat. Struct. Mol. Biol.*, **21**, 679–685.
41. Schuetz, A., Murakawa, Y., Rosenbaum, E., Landthaler, M. and Heinemann, U. (2014) Roquin binding to target mRNAs involves a winged helix-turn-helix motif. *Nat. Commun.*, **5**, 5701.
42. Murakawa, Y., Hinz, M., Mothes, J., Schuetz, A., Uhl, M., Wyler, E., Yasuda, T., Mastrobuoni, G., Friedel, C.C., Dolken, L. *et al.* (2015) RC3H1 post-transcriptionally regulates A20 mRNA and modulates the activity of the IKK/NF-kappaB pathway. *Nat. Commun.*, **6**, 7367.
43. Codutti, L., Leppek, K., Zalesak, J., Windeisen, V., Masiewicz, P., Stoecklin, G. and Carlomagno, T. (2015) A distinct, Sequence-Induced conformation is required for recognition of the constitutive decay element RNA by Roquin. *Structure*, **23**, 1437–1447.
44. Braun, J., Fischer, S., Xu, Z.Z., Sun, H., Ghoneim, D.H., Gimbel, A.T., Plessmann, U., Urlaub, H., Mathews, D.H. and Weigand, J.E. (2018) Identification of new high affinity targets for Roquin based on structural conservation. *Nucleic Acids Res.*, **46**, 12109–12125.
45. Essig, K., Kronbeck, N., Guimaraes, J.C., Lohs, C., Schlundt, A., Hoffmann, A., Behrens, G., Brenner, S., Kowalska, J., Lopez-Rodriguez, C. *et al.* (2018) Roquin targets mRNAs in a 3'-UTR-specific manner by different modes of regulation. *Nat. Commun.*, **9**, 3810.
46. Stoecklin, G., Lu, M., Rattenbacher, B. and Moroni, C. (2003) A constitutive decay element promotes tumor necrosis factor alpha mRNA degradation via an AU-rich element-independent pathway. *Mol. Cell. Biol.*, **23**, 3506–3515.
47. Kemmerer, K., Fischer, S. and Weigand, J.E. (2018) Auto- and cross-regulation of the hnRNPs D and DL. *RNA*, **24**, 324–331.
48. Kemmerer, K. and Weigand, J.E. (2014) Hypoxia reduces MAX expression in endothelial cells by unproductive splicing. *FEBS Lett.*, **588**, 4784–4790.
49. Richter, C., Reif, B., Wörner, K., Quant, S., Marino, J.P., Engels, J.W., Griesinger, C. and Schwalbe, H. (1998) A new experiment for the measurement of nJ(C,P) coupling constants including 3J(C4'i,Pi) and 3J(C4'i,Pi+1) in oligonucleotides. *J. Biomol. NMR*, **12**, 223–230.
50. Schwalbe, H., Samstag, W., Engels, J.W., Bermeil, W. and Griesinger, C. (1993) Determination of 3J(C,P) and 3J(H,P) coupling constants in nucleotide oligomers with FIDS-HSQC. *J. Biomol. NMR*, **3**, 479–486.
51. Rinnenthal, J., Richter, C., Ferner, J., Duchardt, E. and Schwalbe, H. (2007) Quantitative gamma-HCNH: determination of the glycosidic torsion angle chi in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation by solution NMR spectroscopy. *J. Biomol. NMR*, **39**, 17–29.
52. Linge, J.P., Habeck, M., Rieping, W. and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, **19**, 315–316.
53. Nozinovic, S., Furtig, B., Jonker, H.R., Richter, C. and Schwalbe, H. (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.*, **38**, 683–694.
54. Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.*, **147**, 138–146.
55. Schwieters, C.D., Kuszewski, J.J. and Clore, G.M. (2006) Using Xplor-NIH for NMR molecular structure determination. *Progr. NMR Spectrosc.*, **48**, 47–62.
56. Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C.A. and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, **84**, 5066–5073.
57. Kabsch, W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.
58. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 72–82.
59. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.*, **50**, 760–763.
60. Read, R.J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D. Biol. Crystallogr.*, **57**, 1373–1382.
61. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.
62. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.*, **53**, 240–255.
63. Winn, M.D., Murshudov, G.N. and Papiz, M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.*, **374**, 300–321.
64. Laskowski, R., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.
65. Chen, V.B., Arendall, W.B. 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 12–21.
66. Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D. *et al.* (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 465–481.
67. Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. and Wright, P.E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat. Struct. Mol. Biol.*, **11**, 257–264.
68. Meyer, A., Golbik, R.P., Sanger, L., Schmidt, T., Behrens, S.E. and Friedrich, S. (2019) The RGG/RG motif of AUF1 isoform p45 is a key modulator of the protein's RNA chaperone and RNA annealing activities. *RNA Biol.*, **16**, 960–971.

69. Kajita, Y., Nakayama, J., Aizawa, M. and Ishikawa, F. (1995) The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. *J. Biol. Chem.*, **270**, 22167–22175.
70. Zucconi, B.E., Ballin, J.D., Brewer, B.Y., Ross, C.R., Huang, J., Toth, E.A. and Wilson, G.M. (2010) Alternatively expressed domains of AU-rich element RNA-binding protein 1 (AUF1) regulate RNA-binding affinity, RNA-induced protein oligomerization, and the local conformation of bound RNA ligands. *J. Biol. Chem.*, **285**, 39127–39139.
71. Katahira, M., Miyanoi, Y., Enokizono, Y., Matsuda, G., Nagata, T., Ishikawa, F. and Uesugi, S. (2001) Structure of the C-terminal RNA-binding domain of hnRNP D0 (AUF1), its interactions with RNA and DNA, and change in backbone dynamics upon complex formation with DNA. *J. Mol. Biol.*, **311**, 973–988.
72. Nagata, T., Kurihara, Y., Matsuda, G., Saeki, J., Kohno, T., Yanagida, Y., Ishikawa, F., Uesugi, S. and Katahira, M. (1999) Structure and interactions with RNA of the N-terminal UUAG-specific RNA-binding domain of hnRNP D0. *J. Mol. Biol.*, **287**, 221–237.
73. Schlundt, A., Niessing, D., Heissmeyer, V. and Sattler, M. (2016) RNA recognition by Roquin in posttranscriptional gene regulation. *Wiley Interdiscip. Rev. RNA*, **7**, 455–469.
74. Sakurai, S., Ohto, U. and Shimizu, T. (2015) Structure of human Roquin-2 and its complex with constitutive-decay element RNA. *Acta Crystallogr. F Struct. Biol. Commun.*, **71**, 1048–1054.
75. Zhang, Q., Fan, L., Hou, F., Dong, A., Wang, Y.X. and Tong, Y. (2015) New insights into the RNA-binding and E3 Ubiquitin ligase activities of Roquins. *Sci. Rep.*, **5**, 15660.
76. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. and Weissman, J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
77. Plass, M., Rasmussen, S.H. and Krogh, A. (2017) Highly accessible AU-rich regions in 3' untranslated regions are hotspots for binding of regulatory factors. *PLoS Comput. Biol.*, **13**, e1005460.
78. Friedrich, S., Schmidt, T., Geissler, R., Lilie, H., Chabierski, S., Ulbert, S., Liebert, U.G., Golbik, R.P. and Behrens, S.E. (2014) AUF1 p45 promotes West Nile virus replication by an RNA chaperone activity that supports cyclization of the viral genome. *J. Virol.*, **88**, 11586–11599.
79. Friedrich, S., Engelmann, S., Schmidt, T., Szczepankiewicz, G., Bergs, S., Liebert, U.G., Kummerer, B.M., Golbik, R.P. and Behrens, S.E. (2018) The host factor AUF1 p45 supports flavivirus propagation by triggering the RNA switch required for viral genome cyclization. *J. Virol.*, **92**, e01647-17.
80. Glasmacher, E., Hoefig, K.P., Vogel, K.U., Rath, N., Du, L., Wolf, C., Kremmer, E., Wang, X. and Heissmeyer, V. (2010) Roquin binds inducible costimulator mRNA and effectors of mRNA decay to induce microRNA-independent post-transcriptional repression. *Nat. Immunol.*, **11**, 725–733.
81. Hanieh, H., Masuda, K., Metwally, H., Chalise, J.P., Mohamed, M., Nyati, K.K., Standley, D.M., Li, S., Higa, M., Zaman, M.M. *et al.* (2018) Arid5a stabilizes OX40 mRNA in murine CD4(+) T cells by recognizing a stem-loop structure in its 3' UTR. *Eur. J. Immunol.*, **48**, 593–604.
82. Masuda, K., Ripley, B., Nishimura, R., Mino, T., Takeuchi, O., Shioi, G., Kiyonari, H. and Kishimoto, T. (2013) Arid5a controls IL-6 mRNA stability, which contributes to elevation of IL-6 level in vivo. *Proc. Natl Acad. Sci. U.S.A.*, **110**, 9409–9414.
83. Essig, K., Hu, D., Guimaraes, J.C., Alterauge, D., Edelmann, S., Raj, T., Kranich, J., Behrens, G., Heiseke, A., Floess, S. *et al.* (2017) Roquin suppresses the PI3K-mTOR signaling pathway to inhibit T helper cell differentiation and conversion of treg to Tfr cells. *Immunity*, **47**, 1067–1082.
84. Li, X., Quon, G., Lipshitz, H.D. and Morris, Q. (2010) Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA*, **16**, 1096–1107.
85. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Stein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
86. Gerstberger, S., Hafner, M. and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 829–845.

Structural basis for the recognition of transiently structured AU-rich elements by Roquin

Oliver Binas^{1†}, Jan-Niklas Tants^{2†}, Stephen A. Peter³, Robert Janowski⁴, Elena Davydova⁴, Johannes Braun^{3#}, Dierk Niessing^{4,5}, Harald Schwalbe¹, Julia E. Weigand^{3,*} and Andreas Schlundt^{2,*}

¹ Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany

² Institute for Molecular Biosciences, Goethe University Frankfurt and Center for Biomolecular Magnetic Resonance (BMRZ), 60438 Frankfurt, Germany

³ Department of Biology, Technical University of Darmstadt, Darmstadt 64287, Germany

⁴ Institute of Structural Biology, Helmholtz-Zentrum München, 85764, Neuherberg, Germany

⁵ Institute of Pharmaceutical Biotechnology, Ulm University, 89081, Ulm, Germany

* To whom correspondence should be addressed. Tel: +49 69 798 29699; Fax: +49 69 798 29225; Email: schlundt@bio.uni-frankfurt.de

Correspondence may also be addressed to Tel: +49 6151 16 22005; Fax: +49 6151 16 22003;

Email: julia.weigand@tu-darmstadt.de

† Joint Authors

#Present Address: Center for Thrombosis and Homeostasis, Johannes Gutenberg University Medical Center, Mainz 55131, Germany

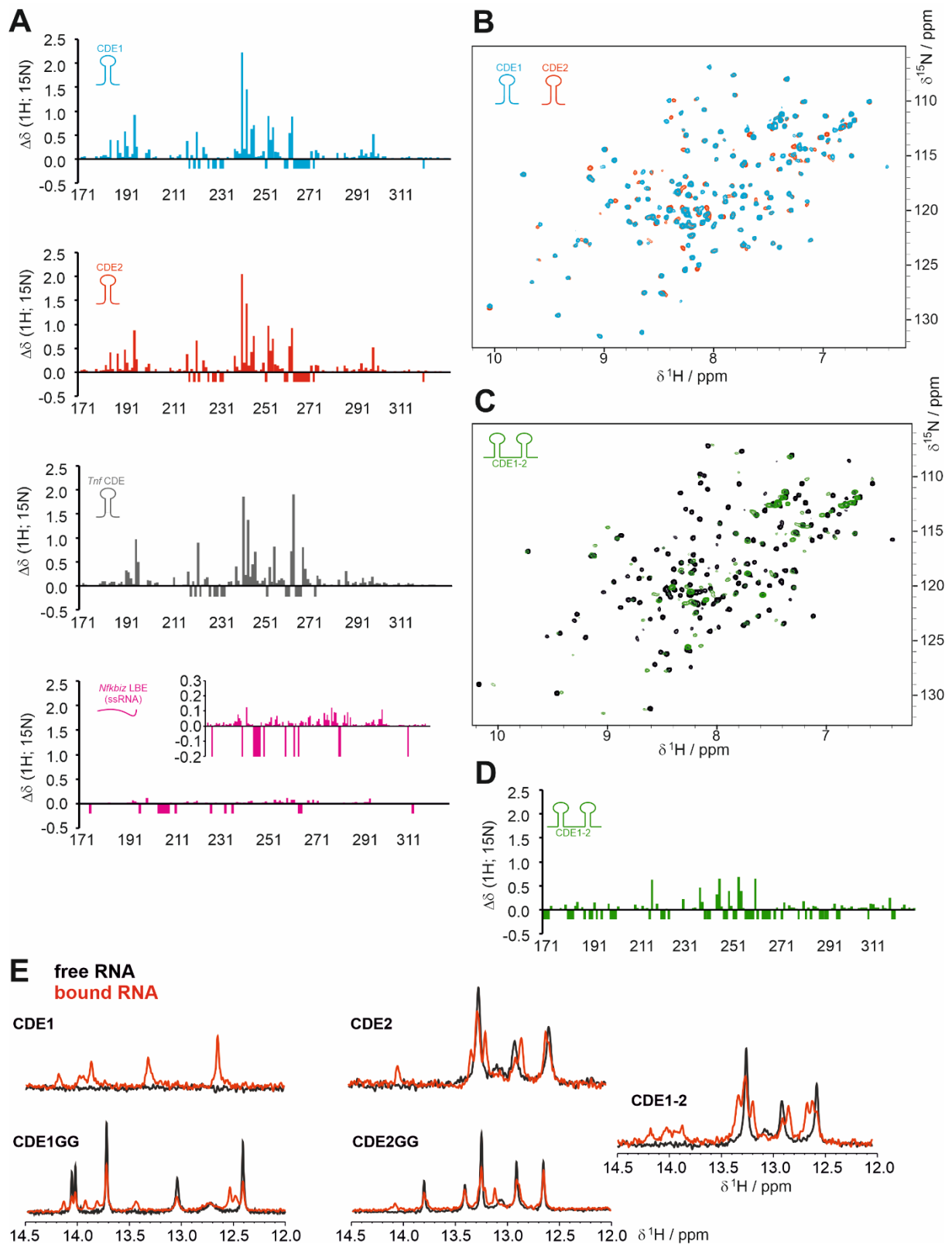
This Supplementary information comprises three tables, eight figures and references. Due to its size and complexity, Supplementary Table 3 will be provided and uploaded as a separate file.

Supplementary Table 1: CDE sequences. Mutations with respect to the wild type are shown in red.

Construct	Sequence 5' → 3'
CDE1	GGAAUAAAUAUAUUAAUUUUUGAAAC
CDE2	GAUUAGCCAUGCCUAAUAUUUAGGCAAGA
CDE1GG	GGAAAUUAUAUUAAUUUCC
CDE2GG	GGUGCCUAAUAUUUAGGCACC
tandem CDE1-2	GGAAUAAAUAUAUUAAUUUUUGAAACCCAUUAGCCAUGGCUAAUAUUUAGGCAAGA
Tnf CDE	ACAUGUUUUCUGUGAAAACGGAG
Luciferase reporter gene assay	
UCP3_wt	CAAGAUGGAAAAUAAAUAUAUUAAUUUUUGAAACCCAUUAGGCAUGCCUAAUAUUUA GGCAAG
UCP3_NMR	CAAGAUGGAAAAUAAAUAUAUUAAUUUUUGAAACCCAUUAG C CAUGCCUAAUAUUUA GGCAAG
RNA affinity purification	
UCP3_wt	GGCAAGAUGGAAAAUAAAUAUAUUAAUUUUUGAAACCCAUUAGGCAUGCCUAAUAUU UAGGCAAG
UCP3_CDE1_mut	GGCAAGAUGGAAAAUA CCGGCGC UUAUUUUUGAAACCCAUUAGGCAUGCCUAAUAUU UAGGCAAG
UCP3_CDE2_mut	GGCAAGAUGGAAAAUAAAUAUAUUAAUUUUUGAAACCCAUUAGGCAUGCCUAAUA GG G AGGCAAG
UCP3_double_mut	GGCAAGAUGGAAAAUA CCGGCGC UUAUUUUUGAAACCCAUUAGGCAUGCCUAAUA GG G AGGCAAG

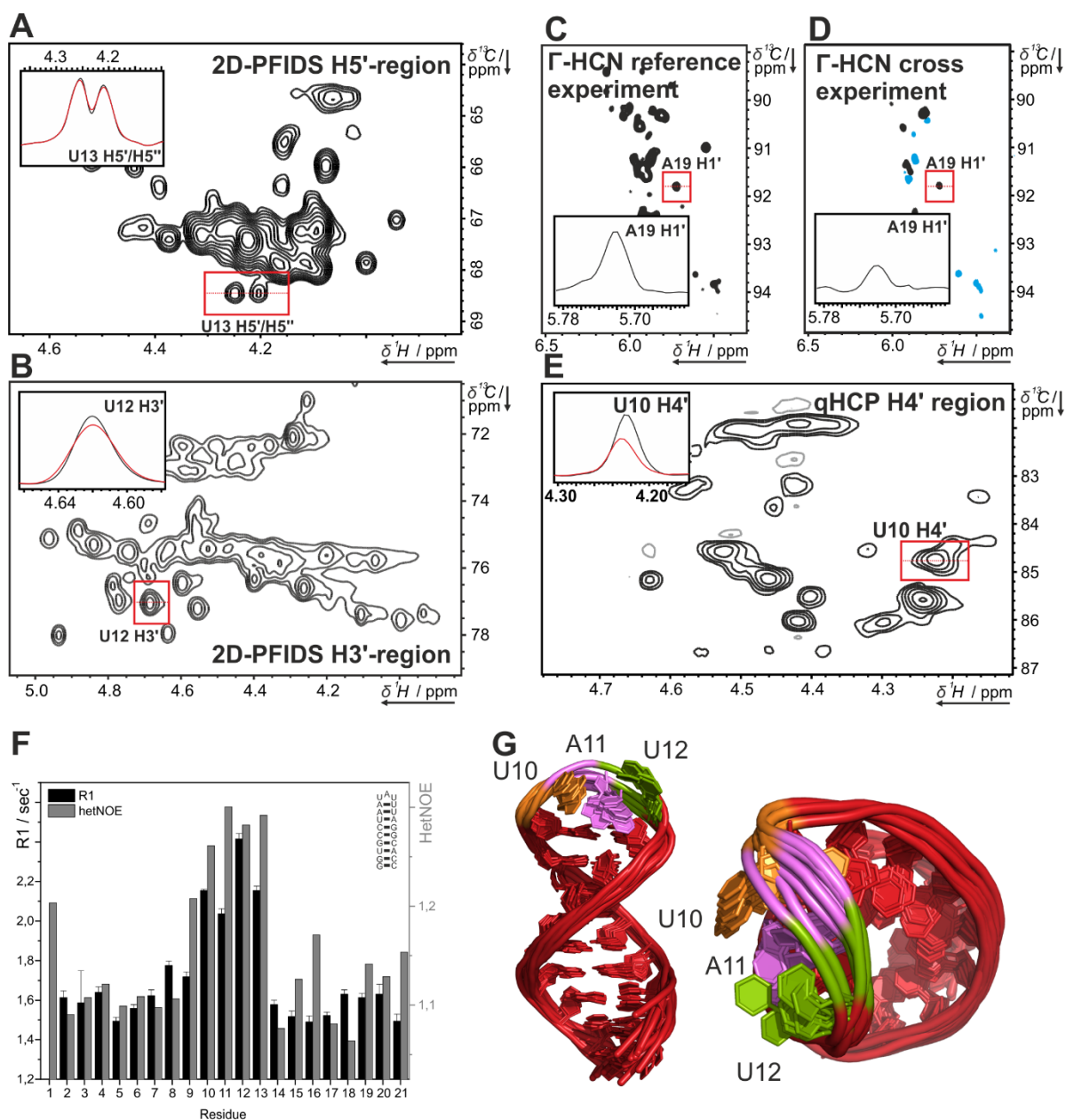
Supplementary Table 2: Roquin ROQ domain and AUF1 RRM domains as used in this study, i.e. after TEV cleavage from His₆-Thioredoxin-fusion proteins.

Construct	Sequence N → C
Roquin_ROQ	GAMALQHQNPPQLSSNLWAAVRARGCQFLGPAMQEEALKLVLLALEDGSAL SRKVLVLFVVRLEPRFPQASKTSIGHVVQLLYRASCVKVTKRDEDSSLMQLK EEFRTYEALRREHDSQIVQIAMEAGLRIAPDQWSSLLYGDQSHKSHMQSIIDKL QT
AUF1_RRM2	GKTKEPVKKIFVGGGLSPDTPEEKIREYFGGFGEVESIELPMDNKTNRKRG FCFITFKEEEPVKKIMEKKYHNVGLSKCEIKVAMSKE
AUF1_RRM1-2	GASKNEEDEGKMFIGGLSWDTTKKDLKDYFSKFGEVVDCTLKLDPITGRSRGF GFVLFKESESVDKVMQKEHKLNGKVIDPKRAKAMKTKEPVKKIFVGGGLSPDT PEEKIREYFGGFGEVESIELPMDNKTNRKRGFCFITFKEEEPVKKIMEKKYHNV GLSKCEIKVAMSKE

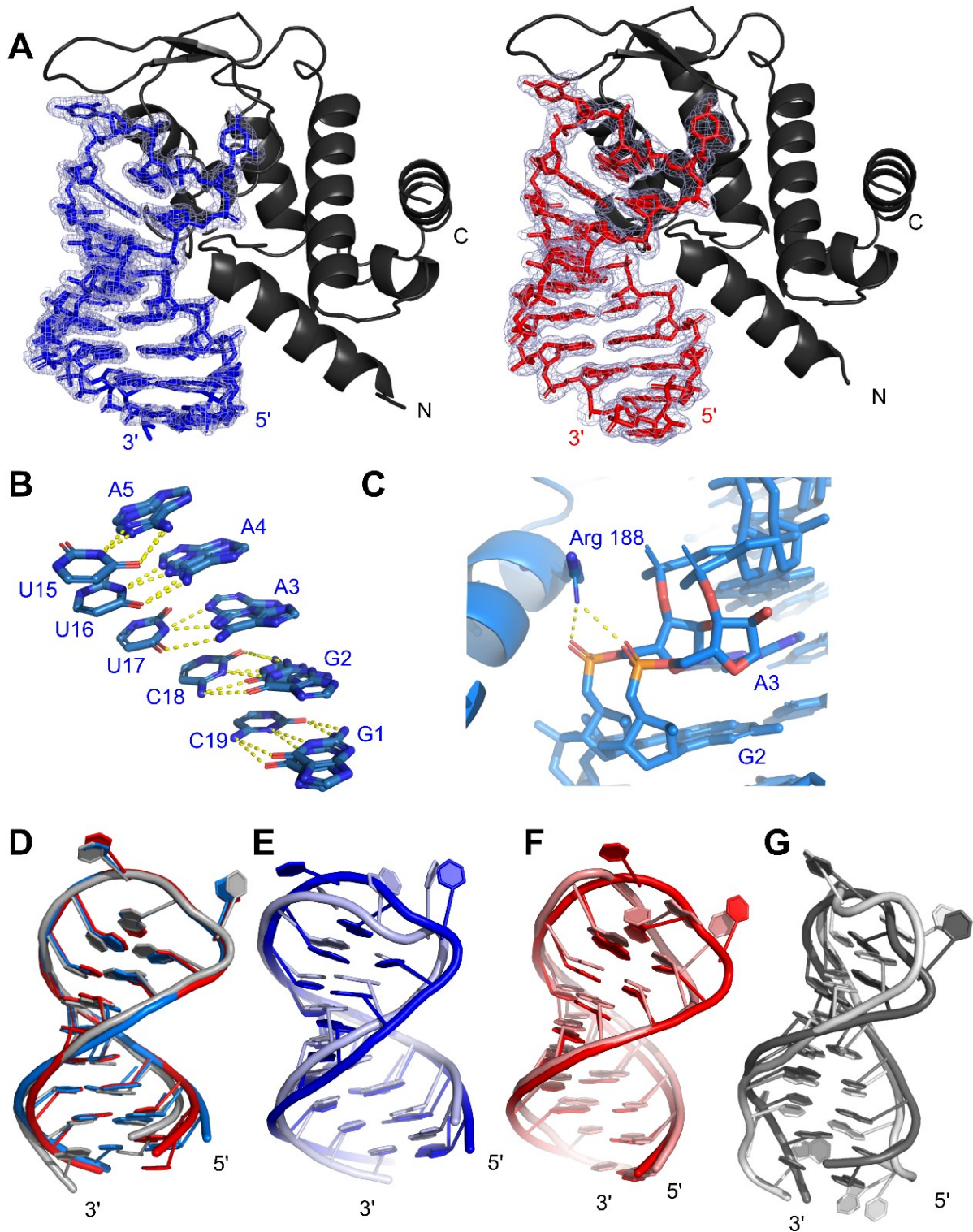


Supplementary Figure S2. The *UCP3* mRNA stem-loops are bound by the Roquin ROQ domain in a CDE-like manner. **(A)** Comparison of bar plots showing combined chemical shift perturbations (CSPs) against the ROQ domain sequence calculated from free vs. CDE-bound protein. The respective RNAs are shown as insets. The CSPs of the *UCP3* wild type CDE1 and CDE2 are derived from spectra shown in **Figures 2C and D**. The *Tnf* CDE CSP plot is adjusted from (1). For comparison, the panel also shows a CSP plot of the ROQ domain when titrated with the 15-mer ssRNA (LBE) found in the 3'-UTR of *Nfk1b*

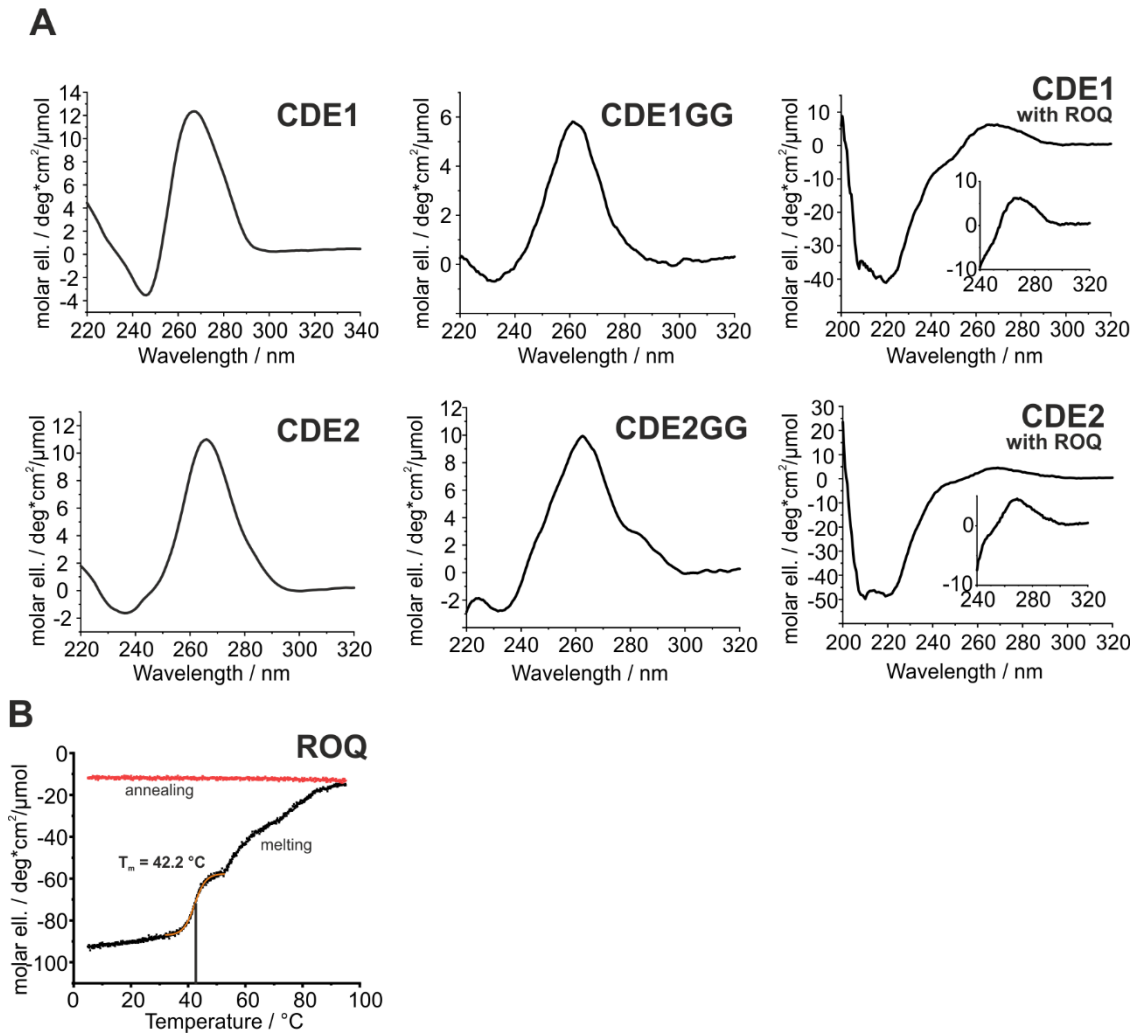
and is described with spectra in (2). Negative bars indicate residues with missing assignments in at least one of the two spectra or prolines. CSP patterns are highly similar. **(B)** Overlay of ^1H - ^{15}N -HSQC NMR spectra showing the Roquin ROQ domain in complex with either wild type *UCP3* CDE1 or CDE2 (in twofold excess, as shown in **Figures 2C and D**). Spectra of the final complexes show highly comparable chemical shifts indicating similar complexes. **(C)** Overlay of ^1H - ^{15}N -HSQC NMR spectra showing the Roquin ROQ domain alone (black) or in complex with the tandem *UCP3* CDE1-2 (green). **(D)** CSP plot of spectral overlay shown in **(C)**. **(E)** Overlay of ^1H -1D-NMR spectra zoomed into the imino proton region and showing the RNAs as denoted either alone (black) or in complex with equimolar amounts of ROQ domain protein (red).



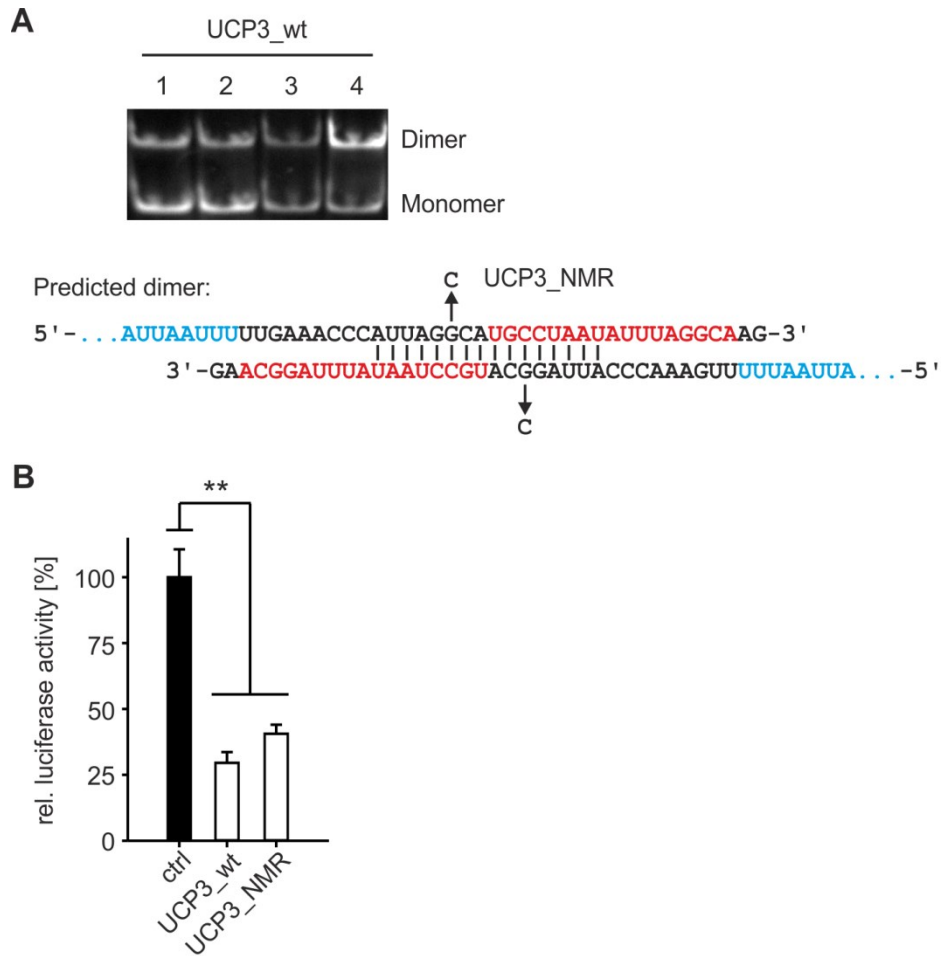
Supplementary Figure S4. NMR experiments for structure determination (dihedral restraints) and dynamics of CDE2GG. **(A and B)** H5' (**A**) and H3' (**B**) regions 2D-PFIDS experiment for determination of H-P coupling constants. A 1D-slice demonstrating the difference between decoupled and coupled spectra is shown in the insert. **(C and D)** Γ -HCN experiment for determination of the glycosidic bond angles via C-H dipolar cross-correlated relaxation rates. Reference (**C**) and cross experiment (**D**) are shown with the insert highlighting the intensity difference in 1D-slices (see reference (3)). **(E)** H4' region of the 2D quantitative HCP experiment for the determination of C-P coupling constants. The insert shows the difference in intensity between the coupled and uncoupled experiment. **(F)** Diagram of spin-lattice relaxation rate and hetNOE of CDE2GG against residue number. Data was measured on sugar C1' nuclei (R1) and aromatic protons (hetNOE) of each residue. **(G)** Unrefined NMR structure of CDE2GG (PDB: 6XXB) Tri-loop residues are labeled in sequential order (orange, magenta and green respectively) to highlight differences in loop geometry.



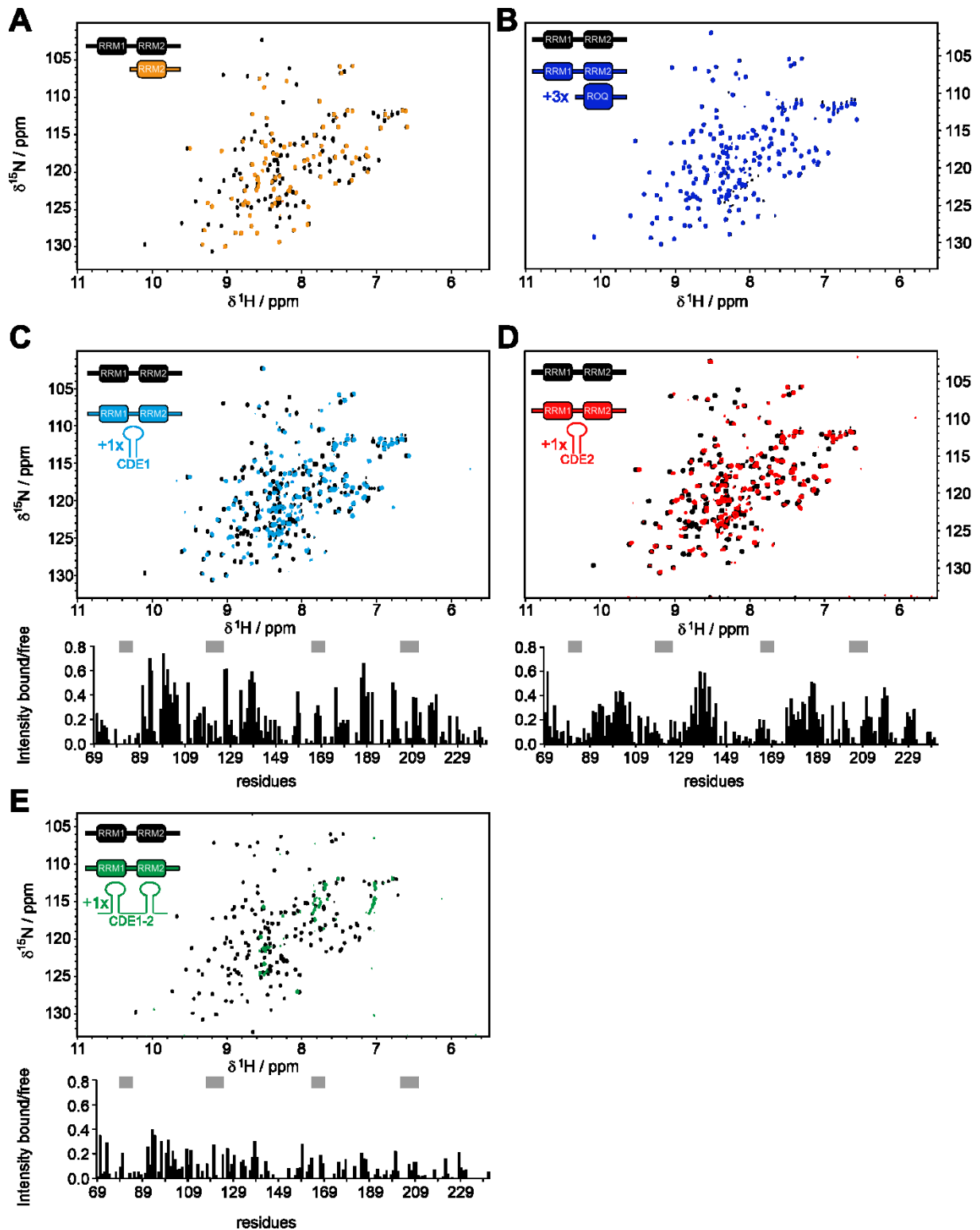
Supplementary Figure S5. Structures of *UCP3* CDE-AREs alone and in complex with ROQ. **(A)** Crystal structures of the ROQ domain in complex with CDE1GG (left) or CDE2GG (right). The RNAs are shown with their 2Fo-Fc electron density maps contoured at +1 sigma. **(B)** H-bond pattern between complementary bases as shown within the two bound conformations of CDE1GG. **(C)** Polar interaction between the side chain of Roquin Arg 188 and the CDE1GG RNA backbone at A3, the latter of which is shown in its two conformers. **(D)** Superimposition of the *UCP3* CDE1GG, CDE2GG and the CDE of *Tnf α* in their ROQ-bound form, respectively. **(E - G)** Superimposition of the free (lighter colour) and ROQ-bound forms (darker colour) of *UCP3* CDE1GG (**E**), CDE2GG (**F**) and *Tnf* CDE (**G**).



Supplementary Figure S6. Stabilization of *UCP3* CDEs by the ROQ domain. **(A)** Raw CD spectra of CDE1 and CDE2 with and without addition of ROQ, as well as CDE1GG and CDE2GG alone. Shown is the molar ellipticity as a function of wavelength. deg=degree **(B)** CD melting curve of ROQ showing the melting and annealing curve. Shown is the molar ellipticity as a function of temperature. deg=degree. No annealing is observable due to precipitation of the protein. T_m was obtained from sigmoidal fitting in the region marked orange.



Supplementary Figure S7. Verification of the NMR tandem CDE construct. **(A)** The *UCP3* wild type (*UCP3_wt*) RNA dimerizes at higher concentrations. Native PAGE of 200 ng *UCP3_wt* folded in structure buffer (10 mM Tris-HCl pH 7.0, 10 mM MgCl₂, 100 mM KCl). 1-4) Different folding conditions: 1) 5 min at 95°C, snap-cooling, 2) 5 min at 95°C, slow cooling, 3) 5 min at 65°C, snap-cooling, 4) 5 min at 65°C, slow cooling. The predicted interaction between two *UCP3_wt* RNAs is shown. A G-to-C conversion in the NMR-adjusted tandem CDE (*UCP3_NMR*) inhibits dimerization. **(B)** Repression mediated by the *UCP3* wild type and NMR-adjusted tandem CDE. Wild type tandem CDE (*UCP3_wt*) and the mutant used in NMR experiments (*UCP3_NMR*) were fused to firefly luciferase. Firefly luciferase activity was normalized to *Renilla* luciferase as internal transfection control. Values are normalized to an empty vector control, without *UCP3* 3'-UTR sequences. $n = 3$. (**) P -value < 0.01 (Student's t -test, two-tailed, paired).



Supplementary Figure S8. AUF1 RRM1-2 NMR experiments. **(A)** The AUF1 RRMs are independent in solution: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -labelled AUF1 RRM1-2 tandem alone and with ^{15}N -RRM2 alone. **(B)** The AUF1 RRMs do not interact with the Roquin ROQ domain: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -AUF1 RRM1-2 tandem alone and with threefold stoichiometric excess of ROQ domain (unlabelled). **(C-E)** *UCP3* CDE-binding of AUF1: Overlay of ^1H - ^{15}N -HSQC spectra of ^{15}N -labelled AUF1 RRM1-2 tandem alone and with wild type *UCP3* CDE1 **(C)**, CDE2 **(D)** or the tandem CDE1-2 **(E)** at equimolar amounts. The RNAs are unlabelled. For quantification, the intensity ratios of bound over free tandem RRM amide peaks are plotted versus primary sequence below the respective

spectrum. The grey bars indicate the RNP motifs that are by default responsible for RNA-recognition in a canonical RRM and well cover the dips in the respective plots.

Supplementary references

1. Schlundt, A., Heinz, G.A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D. and Sattler, M. (2014) Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nat Struct Mol Biol*, **21**, 671-678.
2. Essig, K., Kronbeck, N., Guimaraes, J.C., Lohs, C., Schlundt, A., Hoffmann, A., Behrens, G., Brenner, S., Kowalska, J., Lopez-Rodriguez, C. *et al.* (2018) Roquin targets mRNAs in a 3'-UTR-specific manner by different modes of regulation. *Nat Commun*, **9**, 3810.
3. Rinnenthal, J., Richter, C., Ferner, J., Duchardt, E. and Schwalbe, H. (2007) Quantitative gamma-HCNCH: determination of the glycosidic torsion angle chi in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation by solution NMR spectroscopy. *J Biomol NMR*, **39**, 17-29.

11. Research article III: Structure validation for G-rich RNAs in non-coding regions of the human genome.

Oliver Binas, Irene Bessi and Harald Schwalbe

ChemBioChem, **2020**, 21, 1656 – 1663

The cellular roles RNA G-quadruplexes (G4s) are an intensively discussed topic in current research (see **Chapter 4.4**). In many biological or bioinformatical studies, G-rich and possibly G4 forming sequences are identified which are claimed to have regulatory roles. However, these RNAs are rarely sufficiently characterized by biophysical methods *in vitro* in order to satisfy the claim of G4 formation *in vivo*. In this article, a simple protocol for biophysical characterization of G-rich RNAs is applied to sequences obtained from a striking study by Wendel et al., in which G4s are found to be associated with action of the cancer therapeutic silvestrol on the human initiation factor eIF4A.^[266] The protocol is divided into three increasingly complex steps, starting with CD characterization, followed by 1D-NMR spectroscopy and finally 2D-NMR spectroscopy. It was found that, of the six RNAs investigated, only one clearly formed a G4, while two were in an equilibrium with a hairpin structure. The three remaining sequences formed unspecific aggregates.

All experiments, data evaluation were performed by the author of the thesis. CD and 1D-NMR studies were performed under supervision of Dr. I. Bessi. The manuscript was written by the author, I. Bessi and H. Schwalbe.

Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome

Oliver Binas,^[a] Irene Bessi,^[b] and Harald Schwalbe*^[a]

We present the rapid biophysical characterization of six previously reported putative G-quadruplex-forming RNAs from the 5'-untranslated region (5'-UTR) of silvestrol-sensitive transcripts for investigation of their secondary structures. By NMR and CD spectroscopic analysis, we found that only a single sequence—[AGG]₂[CGG]₂C—folds into a single well-defined G-quadruplex

structure. Sequences with longer poly-G strands form unspecific aggregates, whereas CGG-repeat-containing sequences exhibit a temperature-dependent equilibrium between a hairpin and a G-quadruplex structure. The applied experimental strategy is fast and provides robust readout for G-quadruplex-forming capacities of RNA oligomers.

Introduction


G-quadruplexes are noncanonical inter- or intramolecular structural motifs formed by G-rich DNA or RNA sequences. Their fundamental building blocks are planar tetrads composed of four guanine nucleobases (G-tetrad) that engage in Hoogsteen-type base pairing (Scheme 1A).^[1,2] Typically, two to four^[3] of these tetrads are stacked on top of each other to form the structural motif, though larger assemblies have been reported as well (Scheme 1B).^[4,5] G-quadruplex folding requires monovalent cations, such as K⁺ or Na⁺, because the positive charge stabilizes the partially negatively charged O6 atoms of the guanine bases.^[6] Although DNA G-quadruplexes show strong structural polymorphism due to loop variety, leading to various possible folding motifs,^[7] the presence of the 2'-OH groups in RNA G-quadruplexes favors a C3'-endo sugar pucker and an all-parallel folding topology (Scheme 1B).^[8] The strong stacking interactions between the tetrads lead to a remarkable thermodynamic stability,^[9] with the stability of RNA G-quadruplexes exceeding that of DNA G-quadruplexes.^[10] DNA G-quadruplexes can be detected in vivo^[11] and their fundamental role in telomere maintenance as well as in gene regulation is generally


acknowledged.^[12] Bioinformatics and in vitro studies^[11,13] suggested the presence of G-rich putative quadruplex-forming sites in untranslated regions (UTRs)^[14] of mRNAs, as well as in the transcripts of human telomers known as telomeric repeat-containing RNA. RNA G-quadruplexes might be involved in processes such as polyadenylation,^[15] modulation of translational efficiency,^[16,17] and splicing.^[18] However, the existence of RNA G-quadruplexes in vivo is a matter of current scientific debate.^[19] In 2016, in-cell mapping experiments suggested that G-rich cellular transcripts capable of forming G-quadruplex structures after refolding in vitro are largely unfolded in eukaryotic cells.^[20] Later studies that combined RNA G-quadruplex-specific precipitation with sequencing were able to detect transient formation of RNA G-quadruplex structures in human cells.^[21] Collectively, these findings suggest the existence of RNA G-quadruplexes in vivo as dynamic structures, their folding/unfolding dynamics governed by the cellular machinery (possibly helicases).


By using high-throughput probing methods, deep sequencing, and bioinformatics, several studies have identified G-rich motifs that, it was speculated, might form G-quadruplexes.^[22–25] Moreover, investigation of the interaction between RNA G-quadruplexes and protein counterparts often involves the use of short G-rich oligomers^[26] that are assumed to form G-quadruplexes in vitro, but without appropriate experimental evidence in support of this assumption. High abundances of guanosine residues in RNA transcripts can indicate G-quadruplex formation, but structure, stability, and other biophysical properties remain unclear without detailed biophysical characterization of these G-rich motifs. Although the RNA G-quadruplex topologies are limited in terms of strand orientation by the strong preference of RNA to form all-anti all-parallel G-quadruplexes, dimerization or multimerization can lead to a number of different general topologies that can alter biological function immensely (Scheme 1). While focusing primarily on the biological context and putting great emphasis on the role of G-quadruplexes in regulatory systems, we wish to remark here

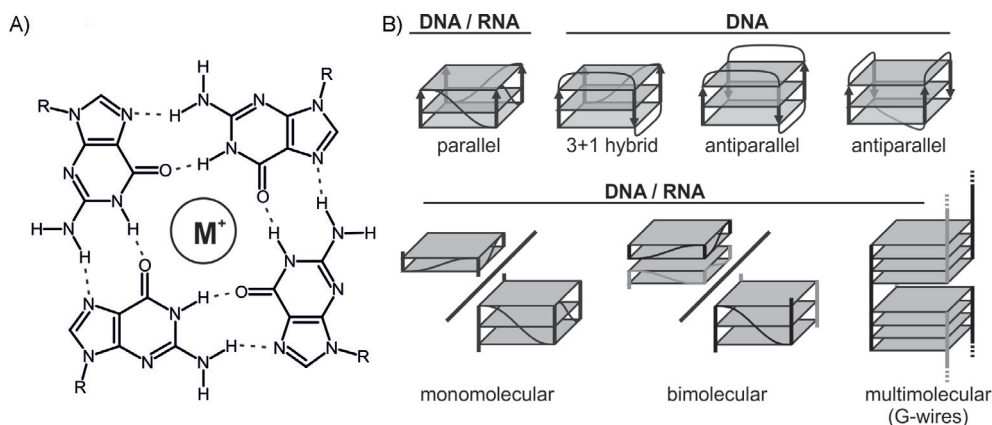
[a] O. Binas, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Goethe University Frankfurt
Max-von-Laue Strasse 7, 60438 Frankfurt (Germany)
E-mail: schwalbe@em.uni-frankfurt.de

[b] I. Bessi
Institute for Organic and Biomolecular Chemistry
Julius Maximilians University Würzburg
Am Hubland 97074 Würzburg (Germany)

 Supporting information and the ORCID identification numbers for the authors of this article can be found under <https://doi.org/10.1002/cbic.201900696>.

 © 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. 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.

 This article is part of a Special Collection on the occasion of Horst Kessler's 80th birthday. To view the complete collection, visit our homepage



Scheme 1. A) General chemical structure of a G-tetrad featuring four guanine residues and a single monovalent cation (M⁺). B) Examples of the possible topologies featured in DNA and RNA G-quadruplexes.

that many cell biology studies do not involve any (or involve only sparse) biophysical characterization of G-quadruplex structures.^[27–31]

Herein, we show that such advanced biophysical structural characterization can be time- and cost-effective and should be mandatory for any high-profile study. It should be performed in order to establish whether putative G-quadruplex-forming sequences are indeed actually forming G-quadruplex structures under the tested in vitro experimental conditions.

After an illustration of the general protocol for a preliminary biophysical screening, the analysis of six putative G-quadruplex-forming RNAs from the 5'-UTR of silvestrol-sensitive transcripts, previously reported in Nature, is presented in detail.^[32]

A simple biophysical protocol for validation of the formation of G-quadruplex structures

We propose a stepwise screening method involving CD and NMR spectroscopic studies and incorporating increasingly sophisticated spectroscopic methods in each step, ultimately providing a strong dataset for the folding potential of RNA sequences predicted to form G-quadruplex structures.

In a first step, CD spectroscopy is used to determine the thermodynamic stabilities of the putative G-quadruplex structures. Additionally, general information about the folding topology of the G-quadruplex can be inferred from the CD signature (Figure 1).^[33] A strong response to addition of a monovalent cation (typically K⁺) is an indicator for G-quadruplex formation; however, the formation of a different structure cannot be excluded unambiguously.

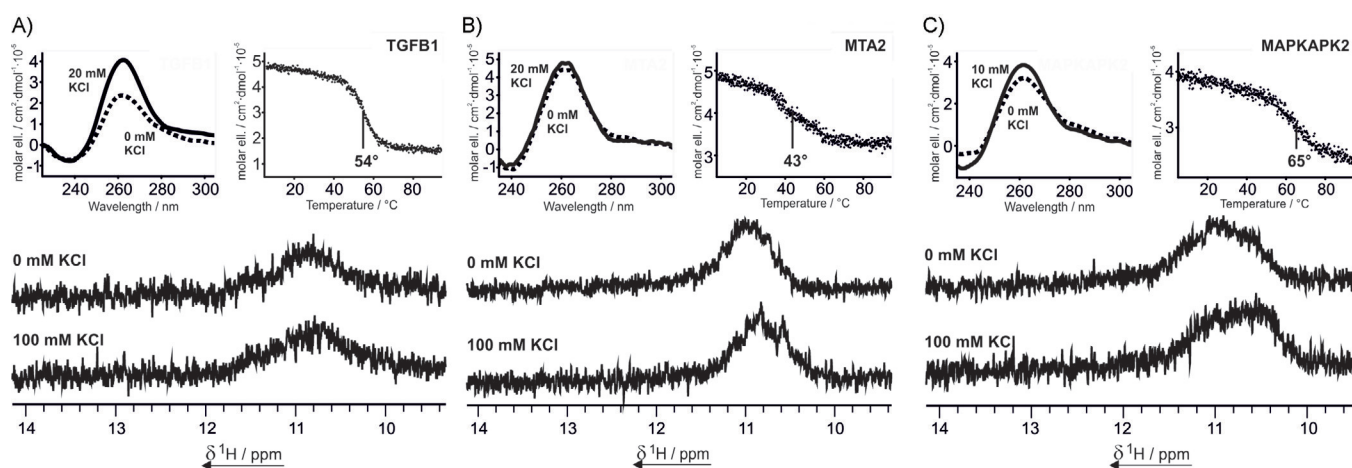


Figure 1. Top left in each panel: CD spectra of A) TGFB1, B) MTA2, and C) MAPKAPK2 in the presence of 0 mM KCl (—) and A), B) 20, or C) 10 mM KCl (-----). The solid line spectra mark the end point of a KCl titration (Figure S1). Samples contained 10 μM RNA and 10 mM BisTris-HCl buffer (pH 6.8). Top right in each panel: CD melting curves of A) TGFB1, B) MTA2, and C) MAPKAPK2 at 262 nm in the presence of 20, 20, and 10 mM KCl, respectively. Melting points were derived from sigmoidal fitting. Samples contained 10 μM RNA in 10 mM potassium phosphate buffer (pH 6.8). Bottom in each panel: imino region of the 1D ¹H NMR spectrum of A) TGFB1, B) MTA2, and C) MAPKAPK2 in the presence of 0 mM and 100 mM KCl. Samples contained 100 μM RNA, 25 mM BisTris-HCl buffer (pH 6.8), and 10% D₂O in H₂O. Severe peak broadening is observed due to the formation of large molecular aggregates.

Although CD spectra are helpful for determining G-quadruplex strand topology, the question of whether or not a G-quadruplex is formed cannot be answered conclusively because the spectra of other structures such as G-wires^[34] can look comparable or identical to those of G-quadruplexes. 1D ¹H NMR spectroscopy can help in identifying a G-quadruplex unambiguously, because reporter signals of imino atoms involved in G-tetrads resonate at a characteristic frequency of around 11 ppm. Other than determining the number of guanine residues involved in the G-quadruplex, 1D NMR data lack further structural information, but 2D NMR spectroscopy and, in particular, 2D ¹H,¹H NOESY can serve as tools for more advanced structural characterization in a third step.^[35] Essentially, 2D ¹H,¹H NOESY spectra contain all distances below a threshold of ≈ 6 Å between the protons of the oligonucleotide. Although these data might even be sufficient for the calculation of a basic 3D structure, the assignment of all protons can be tedious and time-consuming and requires data from other 2D NMR methods. Even without an in-depth analysis, however, the general structural features of a G-quadruplex can be determined by analysis of 2D ¹H,¹H NOESY spectra. Further, NMR-spectroscopic methods such as DOSY^[36] or heterocorrelated NMR can yield additional structural information.

By following these increasingly complex steps, it is possible to achieve the biophysical and structural characterization of any small G-quadruplex-forming oligonucleotide. It should be noted that CD and NMR measurements require 0.05–0.5 mM samples in a volume of approx. 0.5 mL, a quantity that can easily be ordered and delivered within only a few working days at comparatively low cost.

A case study—silvestrol-sensitive G-rich transcripts from human 5'-UTR

Recent studies have suggested the involvement of G-quadruplexes in the function of the anticancer therapeutic silvestrol,^[37] which inhibits the initiation factor eIF4A in human T-ALL-infected cell lines.^[38] Wendel et al. found an accumulation of G-rich sequences in the 5'-UTR of human mRNA that experienced downregulation of the translational efficiency under the influence of silvestrol.^[32] Several G-rich sequential motifs that showed a CD profile indicative of G-quadruplex formation were identified.

In this study, we have characterized the tendency of these six short putatively G-quadruplex-forming RNAs from the 5'-UTR of silvestrol-sensitive mRNA transcripts to undergo G-

quadruplex formation. By using the protocol described above, including NMR spectroscopy, we were able to monitor directly the secondary structures actually formed and to assess the influence of varying conditions, such as K⁺, concentration, and temperature. Five of the G-rich RNA sequences, screened in this work (Table 1), are among the most silvestrol-sensitive RNA transcripts, as determined in studies by Wendel et al.^[32] In addition, two flanking U residues were added to EP300 to assess their effect on the overall structure and to impede G-quadruplex stacking,^[39] resulting in the sequence UEP300U.

CD spectroscopic examination of TGFB1, MTA2, and MAPKAPK2 showed a profile indicative of an all-parallel G-quadruplex structure^[40,41] with a maximum at 262 nm and a minimum at 240 nm at 20 mM, 20 mM, and 10 mM KCl, respectively (Figure 1, top left in each panel). No spectral changes were visible after addition of more K⁺ ions to the system (Figure S1 in the Supporting Information). Even without K⁺ ions, a maximum in ellipticity at 262 nm was already observable. G-tetrad formation requires monovalent cations, so this is atypical for a G-quadruplex and hints at the formation of a different secondary structure. The melting points could be determined by CD melting curve analysis (Figure 1, top right in each panel) and were determined as 54, 43, and 65 °C for TGFB1, MTA2, and MAPKAPK2, respectively.

The type of secondary structure was further investigated by using NMR spectroscopy. The 1D ¹H spectra of all three sequences showed only a very broad signal in the imino region between 10 and 12 ppm (Figure 1, bottom panels). This region is typical for Hoogsteen-paired residues as observed in G-quadruplexes.^[42] Although G-quadruplexes, as compact structures, show distinct peaks in the imino region in 1D ¹H NMR spectra, we assume that these G-rich RNA sequences form a higher-order polymorphic structure. The NMR data suggest the formation of high-order unspecific aggregates, because large structures lead to broadening of peaks in NMR spectra, while the peak positions differ throughout the numerous different possible lengths of such structures.^[43] These aggregates might interact through GG N1-carbonyl symmetric base pairs, which can be formed even in the absence of monovalent cations, as recently shown by Plavec et al.^[44] Addition of KCl did not lead to any observable change in the NMR spectra, whereas the CD spectra, at least in the case of TGFB1, showed that a rearrangement takes place. The strong positive signal at 262 nm is also in agreement with the formation of G-wires,^[5] as observed by Protozanova and Macgregor.^[34] For larger aggregates of those G-wires, the detection of a broad signal in 1D ¹H NMR spectra

Table 1. RNA sequences screened throughout the study and the corresponding effects of KCl addition as observed by CD and NMR spectroscopy.

Name	Sequence	Length [nt]	Effect of KCl addition (CD signal at 260 nm)	Effect of KCl addition (NMR signal 10–12 ppm)	Type of NMR signals observed
TGFB1	5'-GGGAGGAGGGGA-3'	13	moderate increase	none observed	broad bulge
MTA2	5'-GGGGCGGGGUA-3'	13	none observed	slight upfield shift	broad bulge
MAPKAPK2	5'-GGGGCGGCGGG-3'	13	minor increase	slight upfield shift	broad bulge
ADAM10	5'-AGGAGCGCGGC-3'	13	strong increase	signals appear	defined imino signals
EP300	5'-CGGCGGCGGG-3'	12	minor increase	temperature-dependent chemical shift changes	defined imino signals
UEP300U	5'-UCGGCGGCGGGU-3'	14	minor increase	temperature-dependent chemical shift changes	defined imino signals

is expected.^[45] In native PAGE experiments the bands of TGFB1 and MAPKAPK2 are strongly broadened, thus supporting the proposed folding scheme of highly polymorphic structures (Figure S2). MTA2 shows, apart from the same broad bands, a small defined band with an intensity that diminishes upon re-folding (Figure S2).

ADAM10 showed no significant circular dichroism signal without the addition of KCl (Figure 2, top left). This contrasts with the CD data for TGFB1, MTA2, and MAPKAPK2 (Figure 1), because those sequences showed strong signals at around 260 nm even without addition of KCl. When, however, KCl was added to a sample of ADAM10, a positive ellipticity at 260 nm could be observed. The end point of the titration was reached at 30 mM KCl, and no further increase in ellipticity could be measured. CD melting analysis was thus carried out at a KCl concentration of 30 mM. The melting curve (Figure 2, bottom left) shows a clear sigmoidal profile, and the melting point could be determined as 44 °C.

G-quadruplex formation could be confirmed by 1D ¹H NMR data (Figure 2, right). In the absence of KCl, no signal apart from a small broadened bulge was observed in the imino region. This hints at the formation of unspecific aggregates, analogously to the cases of TGFB1, MTA2, and MAPKAPK2. After KCl had been added, however, three sharp separated resonances and several overlapped signals at 11.2–11.3 ppm could be detected. Eight signals hint at a stack of two G-tetrads, as can be expected from the sequence [AGG]₂[CGG]₂C.

The formation of a highly symmetric dimer, though, cannot be ruled out by 1D ¹H NMR spectroscopy alone.^[46] We acquired 1D ¹H and 2D ¹H,¹H NOESY spectra at a higher concentration (700 μM) and discovered two additional imino proton signals, resonating at 10.3 and 9.3 ppm (Figure 3A), that were only faintly visible at 50 μM. A possible hypothesis that could ex-

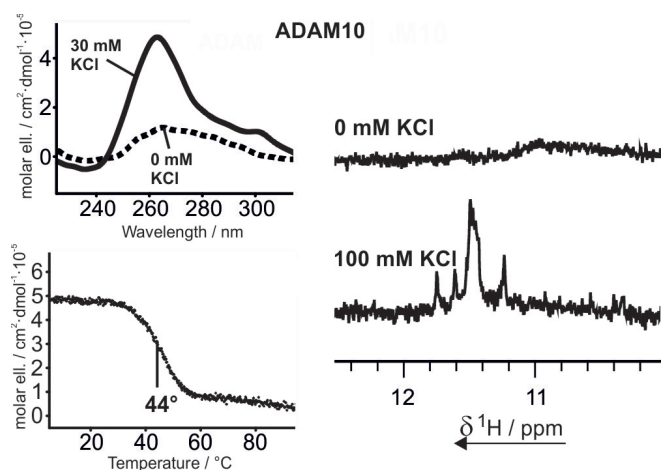


Figure 2. Top left: CD spectrum of ADAM10 in the presence of 0 (----) and 30 mM (—) KCl; 30 mM KCl marked the endpoint of the titration (Figure S1). The sample contained 10 μM RNA and 10 mM BisTris-HCl buffer (pH 6.8). Bottom left: CD melting curve of ADAM10 at 262 nm in the presence of 30 mM KCl. The melting point was derived from sigmoidal fitting. The sample contained 10 μM RNA in 10 mM potassium phosphate buffer (pH 6.8). Right: Imino region of the 1D ¹H NMR spectrum of ADAM10 in the presence of 0 and 100 mM KCl. Samples contained 100 μM RNA, 25 mM Bis-Tris-HCl buffer (pH 6.8), and 10% D₂O in H₂O.

Figure 3. Further investigation of ADAM10 by 2D NMR spectroscopy. A) Imino region of the 1D ¹H spectrum. B) Imino–aromatic region, highlighting strand interactions. C) Four-tetrad G-quadruplex layout as suggested from 2D data. D) CD melting curve at an RNA concentration of 300 μM. E) Full imino region of the 2D ¹H,¹H NOESY spectrum, highlighting two peaks showing three NOESY crosspeaks. F) DOSY of ADAM10 showing zooms of the 1,4-dioxane reference peak and the aromatic RNA peaks.

plain this strong chemical shift perturbation might be ring current effects due to the stacking of terminal or loop residues on the G-tetrad. We investigated the general topology of the G-quadruplex by analysis of the ¹H,¹H NOESY spectrum (Fig-

ure 3B, E). On inspection of the imino proton crosspeaks, we found that crosspeaks from imino proton signals to three other signals are visible (Figure 3E, solid and dashed lines); this would not be expected from a two-tetrad G-quadruplex but could be explained by these residues lying between two tetrads, as in three- or four-tetrad G-quadruplexes. In this case two imino proton crosspeaks to the adjacent tetrads and a third crosspeak to the neighboring base in the same tetrad would explain the observation. Additionally, the imino to aromatic crosspeak region shows that it is possible to follow cross-signals over more than two H1–H8 layers (Figure 3B).

From these observations, the topology of the ADAM10 G-quadruplex would be expected to include four G-tetrads, so a bimolecular G-quadruplex should be formed. We were able to confirm the formation of such a higher-order structure by DOSY spectroscopy (Figure 3F), in which we determined a hydrodynamic radius of 13.7 Å. A comparable value of 13.9 Å was observed in the four-tetrad G-quadruplex of the modified DNA GG-Az1-GG.^[46] The DOSY peaks of ADAM10 and of the reference substance 1,4-dioxane are shown in Figure 3F. Additional peaks with a higher diffusion coefficient are observed, corresponding to an even higher-order structure with a hydrodynamic radius of 18.8 Å that is partially populated.

To provide additional verification of the formation of dimerization we conducted CD melting experiments at a higher concentration (300 μM, Figure 3D). We observed a rise in the melting temperature from 44 to 71 °C, in comparison with the lower concentration (10 μM) investigated before (Figure 2). From the number of imino peaks observed, symmetry can be proposed. Combined with the requirement of an all-anti all-parallel G-quadruplex, this leads to two possible topologies (Figure 3C). These differ in the stacking tetrads, which feature the 5'- or 3'-terminal residues.^[47] The exact determination of the topology including the G-residue polarity would require a full assignment of the NMR signals. Only in rare cases can this be achieved from NOESY data alone; most commonly it involves isotopic labeling of single nucleotides in various NMR samples^[48] or uniform labeling of one NMR sample. These techniques require considerable preparative effort, so their application is beyond the scope of a topological screening, as carried out here.

In circular dichroism spectra of EP300 (Figure 4A, top left) only moderate changes in ellipticity were visible upon addition of KCl. Whereas the peak at 260 nm only showed slight variation in intensity, a small positive peak at 240 nm and a negative peak at 290 nm appeared at high KCl concentrations. Subsequently, circular dichroism at 260 nm was measured over a temperature range between 5 and 95 °C, revealing biphasic behavior with two different transition points: 25 and 44 °C. Subsequent analysis of the NMR imino proton region on titration with KCl, as well as of its temperature dependence, led to the same finding. At 0 °C signals with equal intensity could be observed in the regions characteristic of Hoogsteen base pairs (10–12 ppm) and of Watson–Crick base pairs (12–14 ppm). Upon heating, the low-field signals diminished, vanishing completely above 30 °C. Simultaneously, the intensity of the high-field imino signals increased. This observation indicates a ther-

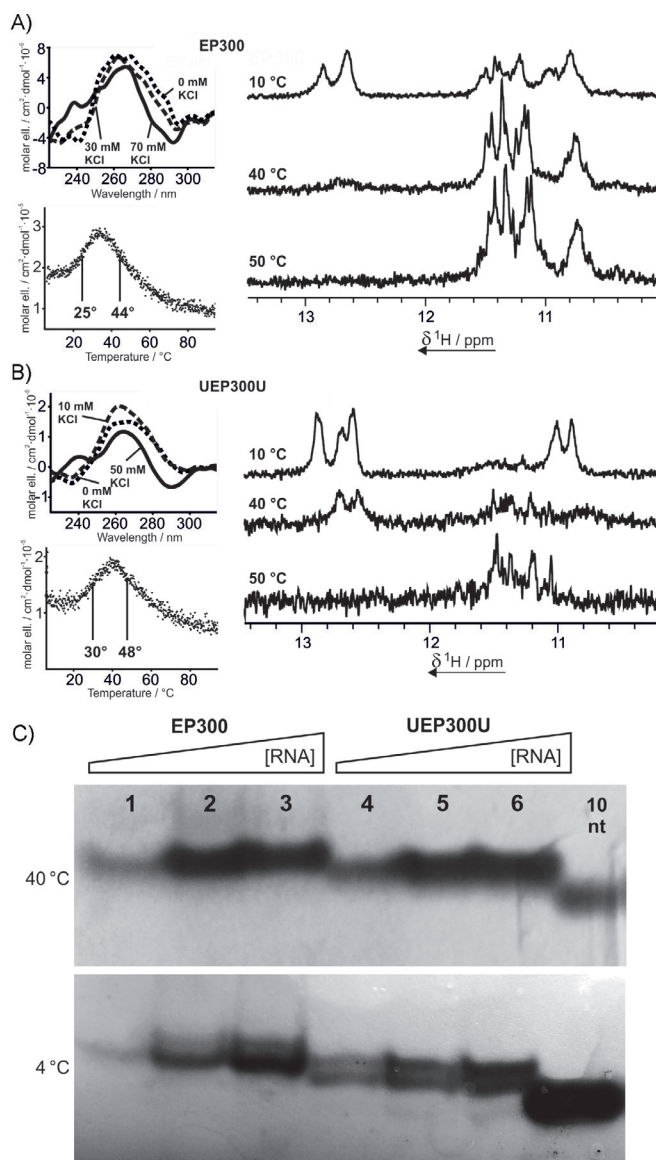


Figure 4. Top left: CD spectra of A) EP300, and B) UEP300U in the presence of 0 mM (dotted lines), A) 30 mM, or B) 10 mM (dashed lines), and A) 50 mM, or B) 70 mM KCl (solid lines). Concentrations of 50 and 70 mM KCl marked the end points of the titration. The sample contained 10 μM RNA and 10 mM BisTris-HCl buffer (pH 6.8). A), B) Bottom left: CD melting curves of EP300 and UEP300U at 262 nm in the presence of 50 or 70 mM KCl, respectively. Melting points were derived from sigmoidal fitting. The sample contained 10 μM RNA in 10 mM potassium phosphate buffer (pH 6.8). Right: Imino regions of the 1D ¹H NMR spectra of EP300 and UEP300U at 10, 40, and 50 °C. Samples contained 100 μM RNA, 25 mM BisTris-HCl buffer (pH 6.8), and 10% D₂O in H₂O. C) Native PAGE (15%) of EP300 and UEP300U at 4 and 40 °C at increasing RNA concentrations. The gel was run for 4 h at 0.5 W with 1 × TBE buffer (50 mM KCl) as running buffer. Samples contained 12, 24, and 36 μM (EP300) or 20, 40, and 60 μM (UEP300U) RNA, 30% glycerol, 50 mM KCl, and 1 × TBE buffer. 10 nt ssRNA was used as a reference.

mal equilibrium between a duplex or hairpin and a G-quadruplex, with the G-quadruplex showing a higher thermal stability. The two states showed different electrophoretic mobilities in native PAGE, resulting in two slightly separated bands at 4 °C but in no such observation being made at 40 °C (Figure 4, lanes 1–3). The electrophoretic mobility of the duplex or hair-

pin structure is comparable with that observed for the G-quadruplex structure.

The same set of experiments was conducted on UEP300U, with comparable results. Interestingly, addition of two flanking U residues to the sequences enhances the thermal stability of the low-temperature structure (possibly duplex or hairpin) relative to that of the G-quadruplex. Thus, imino signals of Watson–Crick bound residues were observed even at 40 °C, whereas in the case of EP300 these signals had completely vanished at this temperature. Analysis of the CD melting curve supports this observation, with the melting point of the low-temperature structure being shifted to 30 °C for UEP300U. The CD profiles of UEP300U showed a more pronounced potassium dependency than those of EP300. The spectral changes upon addition of small amounts of KCl were subtle and led to a spectrum characteristic of an all-anti all-parallel quadruplex, whereas at higher concentrations a positive signal at 240 nm and a negative signal at 290 nm were observed, accompanied by a decrease in the positive signal at 260 nm. Such spectra have previously also been observed in the cases of Z-DNA^[49] and Z-RNA.^[50] These helices can form from purine-pyrimidine repeat sequences.^[51] Hairpin formation with two [CG] base pairs and a GG mismatch has been observed previously,^[52] so formation of a duplex mimicking the CD characteristics of Z-RNA at high salt concentrations could explain the unusual CD spectral characteristics of EP300 and UEP300U. In contrast, Rypniewski et al. observed an A-type helical structure of CGG repeats in crystallographic studies.^[53] As in the case of EP300, the two structures of UEP300U at low temperature can also be observed in native PAGE (Figure 4, lanes 4–6).

Discussion

Herein, we report a three-step protocol for the characterization of putative G-quadruplex-forming RNA oligomers. The protocol involves CD and NMR screening, including CD melting curve analysis and 1D and 2D NMR.

We have demonstrated the application of the protocol by using NMR and circular dichroism studies to probe the structural preferences of six different G-rich RNAs—from the 5'-UTR of human mRNA—that are involved in the modulation of eIF4A suppression by silvestrol. It was found that three of the sequences, despite being G-rich, did not form distinct G-quadruplex structures, but produced aggregates in an unspecific manner. The [XGG]₄ repeat sequences ADAM10, EP300, and UEP300U, however, did fold into G-quadruplexes, with EP300 and UEP300U existing in thermal equilibrium with hairpin structures.

Circular dichroism studies of four of the putative G-quadruplex-forming sequences showed high spectral similarities, especially under the influence of KCl. CD spectra of TGFB1, MTA2, MAPKAPK2, and ADAM10 each showed a strong positive band at 260 nm after addition of KCl, with differences being merely visible before the addition, when the RNA was assumed to be unfolded. Before KCl addition, MTA2, MAPKAPK2, and—to some extent—TGFB1 already showed a positive peak. By using NMR spectroscopy and PAGE we were only able to con-

firm G-quadruplex folding in the case of ADAM10, which forms a four-layer quadruplex. This quadruplex shows a low thermal stability with a melting point of only 44 °C at a concentration of 10 μM. At 300 μM this rises to 71 °C, likely due to dimer formation, which leads to a four-tetrad structure. These have already been observed to be thermodynamically more stable than two-tetrad G-quadruplexes.^[54] The bimolecular nature of this quadruplex is only evident through 2D NMR investigation but could be confirmed by DOSY NMR spectroscopy. NMR spectra revealed that TGFB1, MTA2, and MAPKAPK2 form unspecific aggregates of high molecularity with or without addition of KCl. The absence of a monovalent cation hinders G-tetrad formation, so interaction through GG N1-carbonyl symmetric base pairs is assumed in this case. Under the influence of potassium, a change in the CD spectrum of TGFB1 suggests the formation of G-wires, whereas the behavior of MTA2 and MAPKAPK2 remains unclear. Because NMR spectroscopy fails to resolve structures of such a size, those structural preferences cannot be confirmed beyond doubt, but G-quadruplex formation can be ruled out. The structural polymorphism is supported by native PAGE, with the gel bands being broadened almost beyond detection. The small defined band of MTA2, which is observed alongside the broad main band, diminishes upon refolding, thus indicating a small thermodynamic barrier to aggregate formation.

The structure of CGG-repeat-containing RNA is a subject of current scientific debate.^[55,56] Contributing to this, we collected CD and NMR spectroscopic data for the sequences [CGG]₄ (EP300) and U[CGG]₄U (UEP300U). CGG repeats are known to form helical structures with a non-Watson–Crick GG base pair.^[52] We were able to identify Watson–Crick bases in the NMR spectra of EP300 and UEP300U accordingly. Additionally a high-temperature structure exists in a temperature-dependent structural equilibrium. It is assumed to be a G-quadruplex because the involved imino protons resonate in the region typical of Hoogsteen-bound G residues. Although the temperature-dependent transition between two structural states was determined by CD melting studies, the structural characteristics of each state could be deconvoluted only by NMR. We were able to show that high temperature favors the formation of the G-quadruplex over the duplex. The transition temperature from hairpin to G-quadruplex was higher in the case of the uridine-flanked sequence UEP300U. This might be because of the capping effect of dangling ends, which stabilizes RNA duplexes.^[57] In addition, flanking sequences can negatively affect the stability of G-quadruplex structures;^[58] this adds to the higher relative stability of the hairpin. Even though G-quadruplexes are stabilized by monovalent cations, the duplex seems to be the preferred conformation at high salt concentrations, if room temperature data as observed from CD experiments are considered. Despite CGG repeats forming an A-helix in crystals and Z-RNA in general forming at very high salt concentrations,^[59] CD spectra of EP300 and UEP300U hint at duplexes that exhibit characteristics of Z-RNA. The reason for this behavior remains unclear. However, such structural equilibria are of great biological importance and can be tuned by cellular key factors such as cation levels^[60] or tRNA concentration.^[61]

The electrophoretic mobility of the duplex structure is slightly lower than the mobility of the G-quadruplex in both RNAs. Because the electrophoretic mobility in native PAGE depends not only on size, but also on compactness of the structure, the molecular weight cannot directly be deduced from this data point. NMR spectra of the duplex show at least five imino signals, thus hinting at a bimolecular duplex.

Conclusions

We were able to elucidate the conformational space of six small G-rich mRNA fragments, from human 5'-UTR of mRNA transcripts that are sensitive to silvestrol, by applying a three-step screening protocol involving CD and NMR spectroscopy. Three of the six oligonucleotides do not fold into G-quadruplexes as expected, but instead aggregate unspecifically. ADAM10 ([AGG]₂[CGG]₂C) forms a four-tetrad all-anti parallel G-quadruplex. EP300 and UEP300U ([CGG]₄ and U[CGG]₄U) each fold either into a duplex or into a G-quadruplex depending on the conditions, in particular temperature and salt concentration. G-quadruplex structures are believed to be involved in the therapeutic mechanisms of the anticancer drug silvestrol, so understanding of the structural characteristics of the investigated G-rich sequences is of great importance for discriminating between potential G-quadruplexes and other G-rich sequences. By applying the three steps of our screening protocol in order, we were able to show that CD spectroscopy, although offering a rapid and cost-effective method for obtaining preliminary information on the nature of a secondary structure, cannot be used alone to assess the actual conformation of a G-rich oligonucleotide. NMR spectroscopy is necessary to unravel the information obtained by CD spectroscopy and to shed light on the actual conformation(s) present in solution. The protocol could be useful for obtaining all data points necessary for characterization of the structural characteristics of such oligonucleotides.

Experimental Section

RNA sample preparation: The oligoribonucleotide sequences [GGGAGGAGGGGGA] (TGFB1), [GGGGCGGGGUA] (MTA2), [GGGGCGGGG] (MAPKAPK2), [AGG]₂[CGG]₂C (ADAM10), [CGG]₄ (EP300), and U[CGG]₄U (UEP300U) were bought from Dharmacon (GE Healthcare). RNA samples were purified (HPLC), desalted, precipitated with LiClO₄ (2%, w/v in acetone, 5 volumes) and stored in aqueous stock solutions.

Circular dichroism: All CD experiments were carried out with a Jasco J-810 CD spectrometer (Jasco, GmbH) and use of quartz optical cuvettes with 0.1 cm path length (0.01 cm path length was used for ADAM10 at 300 μM concentration). For CD titration experiments with KCl, RNA samples with a concentration of 10 μM in potassium-free BisTris-HCl buffer (pH 6.8, 10 mM) were prepared. The titration range was 0 to 70 mM KCl. Resulting CD spectra were baseline-corrected and corrected for sample dilution. The data were smoothed by application of a Savitzky–Golay filter^[62] (10 points).

For CD melting curves the samples contained the RNA of interest (10 μM, 300 μM in one additional experiment for ADAM10), potassi-

um phosphate buffer (pH 6.8, 10 mM), and potassium chloride according to the previously determined end point of the titration. The peak ellipticity was monitored over a temperature range from 5 to 95 °C with a scan rate of 1 °C min⁻¹.

Native PAGE: Aqueous stock solutions of RNA were diluted to concentrations of 10–30 μM in TBE buffer (1×) with addition of glycerol (30%) and KCl (50 mM). Polyacrylamide gels (15%) containing TBE (1×) and KCl (50 mM) were cast and run with the samples at 0.5 to 0.8 W for 4 h at 4 or 40 °C. The running buffer contained TBE (1×) and KCl (50 mM). Gel visualization was achieved by GelRed staining and subsequent imaging under UV light.

Nuclear magnetic resonance: 1D ¹H NMR spectra were acquired with 800 MHz Bruker AVIII (MTA2, MAPKAPK2), 600 MHz Bruker AVIII HD (TGFB1, ADAM10), and 600 MHz Bruker AVII (EP300, UEP300U) spectrometers (Bruker Biospin) equipped with Cryo TCI ¹H[¹³C,¹⁵N], Prodigy TCI ¹H[¹³C,¹⁵N], and Cryo TCI ¹H[¹³C,¹⁵N] probes, respectively. Samples contained RNA (100 μM), BisTris-HCl (pH 6.8, 25 mM) for KCl titrations or potassium chloride buffer (pH 6.8, 25 mM) for temperature series, and DSS (25 μM) in H₂O/D₂O (9:1, v/v). The final sample volumes of 280 μL were transferred into 5 mm Shigemi tubes (Shigemi, Inc.). 1D ¹H NMR spectra were recorded with 256 scans, 4096 points, and 1.5 s relaxation delay. Water suppression was applied by using a jump-and-return echo sequence.^[63]

Acknowledgements

The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) in SFB902 and the state of Hesse (BMRZ). We also acknowledge Dr. Christian Richter for his technical support and Elke Stirnal for the HPLC purification of RNA.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: biophysical investigation · circular dichroism · G-quadruplexes · NMR spectroscopy · RNA

- [1] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- [2] J. T. Davis, *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698; *Angew. Chem.* **2004**, *116*, 684–716.
- [3] Y. Chen, D. Yang, *Curr. Protoc. Nucleic Acid Chem.* **2012**, *50*, 17.5.1–17.5.17.
- [4] N. Borbone, J. Amato, G. Oliviero, V. D'Atri, V. Gabelica, E. De Pauw, G. Piccialli, L. Mayol, *Nucleic Acids Res.* **2011**, *39*, 7848–7857.
- [5] T. C. Marsh, E. Henderson, *Biochemistry* **1994**, *33*, 10718–10724.
- [6] T. van Mourik, A. J. Dingley, *Chem. Eur. J.* **2005**, *11*, 6064–6079.
- [7] M. Webba Da Silva, *Chem. Eur. J.* **2007**, *13*, 9738–9745.
- [8] M. M. Fay, S. M. Lyons, P. Ivanov, *J. Mol. Biol.* **2017**, *429*, 2127–2147.
- [9] C. J. Lech, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 2034–2046.
- [10] A. Arora, S. Maiti, *J. Phys. Chem. B* **2009**, *113*, 10515–10520.
- [11] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 75–80.
- [12] J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.* **2019**, <https://doi.org/10.1016/j.trechm.2019.07.002>.
- [13] J. L. Huppert, A. Bugaut, S. Kumari, S. Balasubramanian, *Nucleic Acids Res.* **2008**, *36*, 6260–6268.
- [14] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.

- [15] J.-D. Beaudoin, J.-P. Perreault, *Nucleic Acids Res.* **2013**, *41*, 5898–5911.
- [16] G. D. Balkwill, K. Derecka, T. P. Garner, C. Hodgman, A. P. F. Flint, M. S. Searle, *Biochemistry* **2009**, *48*, 11487–11495.
- [17] M. J. Morris, S. Basu, *Biochemistry* **2009**, *48*, 5313–5319.
- [18] M. M. Ribeiro, G. S. Teixeira, L. Martins, M. R. Marques, A. P. de Souza, S. R. P. Line, *Hum. Genet.* **2015**, *134*, 37–44.
- [19] C. K. Kwok, G. Marsico, S. Balasubramanian, *Cold Spring Harbor Perspect. Biol.* **2018**, *10*, <https://doi.org/10.1101/cshperspect.a032284>.
- [20] J. U. Guo, D. P. Bartel, *Science* **2016**, *353*, aaf5371.
- [21] S. Y. Yang, P. Lejault, S. Chevrier, R. Boidot, A. G. Robertson, J. M. Y. Wong, D. Monchaud, *Nat. Commun.* **2018**, *9*, 4730.
- [22] C. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers, S. Balasubramanian, *Nat. Methods* **2016**, *13*, 841–844.
- [23] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.
- [24] D. Benhalevy, S. K. Gupta, C. H. Danan, S. Ghosal, H. W. Sun, H. G. Kazemier, K. Paeschke, M. Hafner, S. A. Juranek, *Cell Rep.* **2017**, *18*, 2979–2990.
- [25] J.-M. Garant, J.-P. Perreault, M. S. Scott, *Bioinformatics* **2017**, *33*, 3532–3537.
- [26] X. Wang, R. D. Paucek, A. R. Gooding, Z. Z. Brown, E. J. Ge, T. W. Muir, T. R. Cech, *Nat. Struct. Mol. Biol.* **2017**, *24*, 1028–1038.
- [27] T. Endoh, N. Sugimoto, *Sci. Rep.* **2016**, *6*, 22719.
- [28] K. Matsumura, Y. Kawasaki, M. Miyamoto, Y. Kamoshida, J. Nakamura, L. Negishi, S. Suda, T. Akiyama, *Oncogene* **2017**, *36*, 1191–1199.
- [29] G. Mirihana Arachchilage, A. C. Dassanayake, S. Basu, *Chem. Biol.* **2015**, *22*, 262–272.
- [30] M. Sauer, S. A. Juranek, J. Marks, A. De Magis, H. G. Kazemier, D. Hilbig, D. Benhalevy, X. Wang, M. Hafner, K. Paeschke, *Nat. Commun.* **2019**, *10*, 2421.
- [31] M. Beltran, M. Tavares, N. Justin, G. Khandelwal, J. Ambrose, B. M. Foster, K. B. Worlock, A. Tvardovskiy, S. Kunzelmann, J. Herrero, T. Bartke, S. J. Gamblin, J. R. Wilson, R. G. Jenner, *Nat. Struct. Mol. Biol.* **2019**, *26*, 899–909.
- [32] A. L. Wolfe, K. Singh, Y. Zhong, P. Drewe, V. K. Rajasekhar, V. R. Sanghvi, K. J. Mavrikis, M. Jiang, J. E. Roderick, J. Van der Meulen, et al., *Nature* **2014**, *513*, 65–70.
- [33] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648; *Angew. Chem.* **2011**, *123*, 10833–10836.
- [34] E. Protozanova, R. B. Macgregor, Jr., *Biophys. J.* **1998**, *75*, 982–9.
- [35] A. T. Phan, *FEBS J.* **2010**, *277*, 1107–1117.
- [36] A. Ambrus, D. Yang, *Anal. Biochem.* **2007**, *367*, 56–67.
- [37] S. Kim, Y. H. Bang, B. N. Su, H. Chai, Q. Mi, A. D. Kinghorn, R. Wild, S. M. Swanson, *Anticancer Res.* **2007**, *27*, 2175–2183.
- [38] M. E. Bordeleau, F. Robert, B. Gerard, L. Lindqvist, S. M. H. Chen, H. G. Wendel, B. Brem, H. Greger, S. W. Lowe, J. A. Porco, Jr., J. Pelletier, *J. Clin. Invest.* **2008**, *118*, 2651–2660.
- [39] N. Q. Do, K. W. Lim, M. H. Teo, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2011**, *39*, 9448–9457.
- [40] S. Pandey, P. Agarwala, S. Maiti, *J. Phys. Chem. B* **2013**, *117*, 6896–6905.
- [41] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648; *Angew. Chem.* **2011**, *123*, 10833–10836.
- [42] M. Webba da Silva, *Methods* **2007**, *43*, 264–277.
- [43] A. Sengar, B. Heddi, A. T. Phan, *Biochemistry* **2014**, *53*, 7718–7723.
- [44] S. Ceručer, P. Sket, I. Prislán, J. Lah, J. Plavec, *Angew. Chem. Int. Ed.* **2014**, *53*, 4881–4884; *Angew. Chem.* **2014**, *126*, 4981–4984.
- [45] T. Ilc, P. Šket, J. Plavec, M. Webba da Silva, I. Drevenšek-Olenik, L. Spindler, *J. Phys. Chem. C* **2013**, *117*, 23208–23215.
- [46] J. Thevarpadam, I. Bessi, O. Binas, D. P. N. Gonçalves, C. Slavov, H. R. A. Jonker, C. Richter, J. Wachtveitl, H. Schwalbe, A. Heckel, *Angew. Chem. Int. Ed.* **2016**, *55*, 2738–2742; *Angew. Chem.* **2016**, *128*, 2788–2792.
- [47] T. Mashima, A. Matsugami, F. Nishikawa, S. Nishikawa, M. Katahira, *Nucleic Acids Res.* **2009**, *37*, 6249–6258.
- [48] A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
- [49] J. Kypr, I. Kejnovská, D. Renčíuk, M. Vorlíčková, *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
- [50] J. H. Riazance, W. A. Baase, W. C. Johnson, K. Hall, P. Cruz, I. Tinoco, I. Tinoco, Jr., *Nucleic Acids Res.* **1985**, *13*, 4983–4989.
- [51] K. Hall, P. Cruz, I. Tinoco, T. M. Jovin, J. H. van de Sande, *Nature* **1984**, *311*, 584–586.
- [52] K. Sobczak, M. de Mezer, G. Michlewski, J. Krol, W. J. Krzyzosiak, *Nucleic Acids Res.* **2003**, *31*, 5469–5482.
- [53] A. Kiliszek, R. Kierzek, W. J. Krzyzosiak, W. Rypniewski, *Nucleic Acids Res.* **2011**, *39*, 7308–7315.
- [54] A. Joachimi, A. Benz, J. S. Hartig, *Bioorg. Med. Chem.* **2009**, *17*, 6811–6815.
- [55] M. Malgowska, D. Gudanis, R. Kierzek, E. Wyszko, V. Gabelica, Z. Gdaniec, *Nucleic Acids Res.* **2014**, *42*, 10196–10207.
- [56] M. Zumwalt, A. Ludwig, P. J. Hagerman, T. Dieckmann, *RNA Biol.* **2007**, *4*, 93–100.
- [57] S. M. Freier, B. J. Burger, D. Alkema, T. Neilson, D. H. Turner, *Biochemistry* **1983**, *22*, 6198–6206.
- [58] A. Arora, D. R. Nair, S. Maiti, *FEBS J.* **2009**, *276*, 3628–3640.
- [59] M. Popenda, J. Milecki, R. W. Adamiak, *Nucleic Acids Res.* **2004**, *32*, 4044–4054.
- [60] A. Bugaut, P. Murat, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19953–19956.
- [61] A. B. Rode, T. Endoh, N. Sugimoto, *Angew. Chem. Int. Ed.* **2016**, *55*, 14315–14319; *Angew. Chem.* **2016**, *128*, 14527–14531.
- [62] A. Savitzky, M. J. E. Golay, *Anal. Chem.* **1964**, *36*, 1627–1639.
- [63] G. Marius Clore, B. J. Kimber, A. M. Gronenborn, *J. Magn. Reson.* **1983**, *54*, 170–173.

Manuscript received: November 20, 2019

Accepted manuscript online: January 14, 2020

Version of record online: February 26, 2020

CHEMBIOCHEM

Supporting Information

Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome

Oliver Binas,^[a] Irene Bessi,^[b] and Harald Schwalbe^{*[a]}

cbic_201900696_sm_miscellaneous_information.pdf

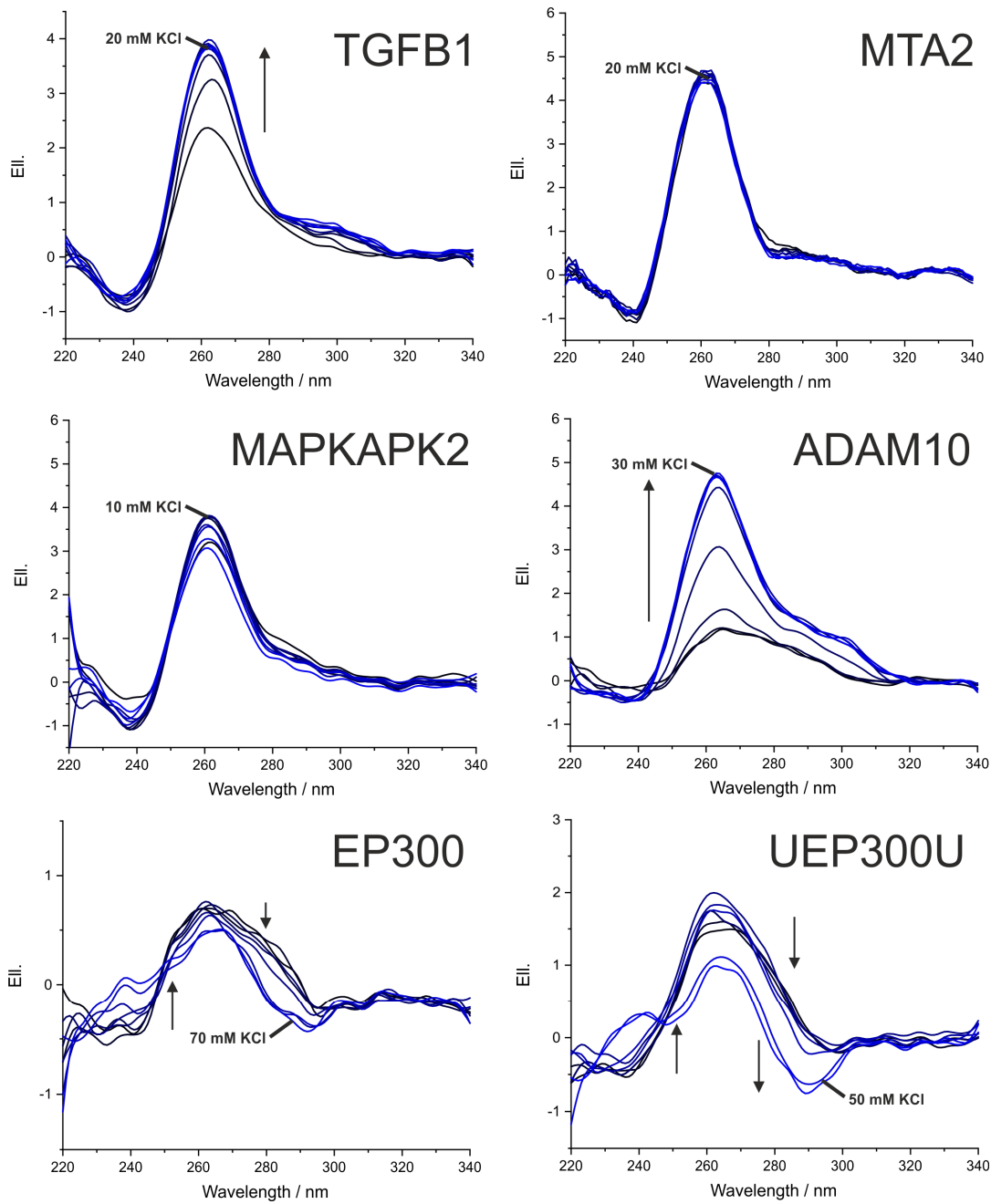


Figure S1. Raw CD KCl titration data of investigated RNAs. CD curves are colored from black (0 mM KCl) to blue (70 mM KCl). Arrows indicate changes in the CD signatures observed at higher KCl concentration. Endpoints of the titration mentioned in the manuscript are marked.

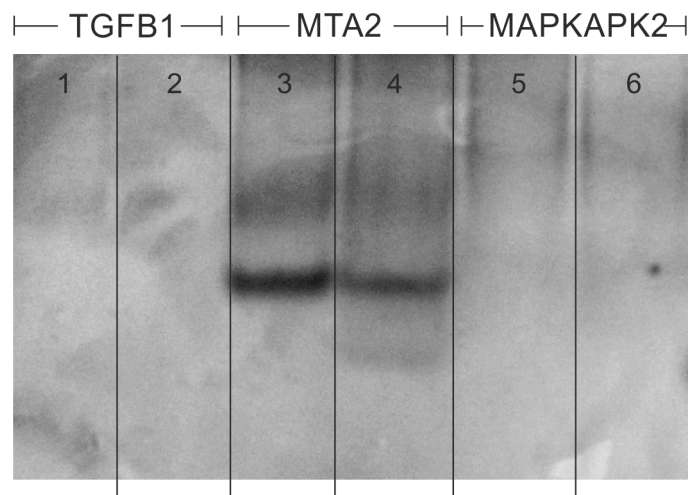


Figure S2. 15%-Native PAGE of TGFB1, MTA2 and MAPKAPK2. Without refolding (lanes 1, 3, 5) and after refolding by heating to 95 °C and rapid cooling (lanes 2, 4, 6).

12. Research article IV: Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif.

Julie Thevarpadam*, Irene Bessi*, Oliver Binas*, Diana P. N. Gonçalves, Chavdar Slavov, Hendrik R. A. Jonker, Christian Richter, Josef Wachtveitl, Harald Schwalbe and Alexander Heckel

*authors contributed equally to the work

Angewandte Chemie International Edition, **2016**, *55*, 2738 – 2742

G-quadruplexes (G4s) can serve as highly interesting and versatile building block in the currently emerging field of nano structure and nano machine development (see **Chapter 5.3**). While many of these approaches rely on the addition of a “fuel” molecule to exert molecular action or sense certain cell metabolites, light can also serve as an external trigger, if photoresponsive moieties are involved. In this article, the development and structural characterization of a highly symmetrical photosensitive DNA G4 is shown, which consists of an azobenzene moiety linked to two guanosine residues on each side. The G4 forms upon potassium addition and can be disassembled by UV-illumination as the azobenzene moiety conformationally switches. Since azobenzene is a reversible photoswitch, subsequent illumination with visible light leads to reformation of the G4 structure, an effect that can be reliably cycled. Photoswitching as well as formation and decomposition of the structure was monitored using CD- and NMR- and FTIR difference spectroscopy. The three-dimensional structure of the assembled G4 was subsequently determined by NMR-spectroscopy, involving angular restraints derived from E.COSY and high-resolution J-sensitive HSQC experiments. Interestingly, it was found that the G4, which was assumed to be bimolecular formed a stacked dimer, leading to a tetramolecular structure, as proven by careful investigation of NOESY-signals, diffusion ordered NMR spectroscopy^[425] and comparative structure calculation.

The author of the thesis was provided with the DNA by Dr. J. Thevarpadam of the group of Prof. Heckel. All NMR experiments involved in the structural characterization as well as NMR experiments monitoring photoswitching and subsequent data evaluation were carried out by the author of the thesis under supervision of Dr. I. Bessi. The structure calculation was performed by Dr. H.R.A. Jonker. The author contributed to the corresponding paragraphs of the manuscript.

G-Quadruplexes

International Edition: DOI: 10.1002/anie.201510269
German Edition: DOI: 10.1002/ange.201510269

Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif

Julie Thevarpadam⁺, Irene Bessi⁺, Oliver Binas⁺, Diana P. N. Gonçalves, Chavdar Slavov, Hendrik R. A. Jonker, Christian Richter, Josef Wachtveitl, Harald Schwalbe,* and Alexander Heckel*

In memory of Gerhard Quinkert

Abstract: The ability of three different bifunctional azobenzene linkers to enable the photoreversible formation of a defined intermolecular two-tetrad G-quadruplex upon UV/Vis irradiation was investigated. Circular dichroism and NMR spectroscopic data showed the formation of G-quadruplexes with K⁺ ions at room temperature in all three cases with the corresponding azobenzene linker in an E conformation. However, only the para-para-substituted azobenzene derivative enables photoswitching between a nonpolymorphic, stacked, tetramolecular G-quadruplex and an unstructured state after E–Z isomerization.

G-Quadruplexes are important DNA secondary structures. The structures form in an intramolecular manner or by association of multiple strands and can exist in many polymorphic forms.^[1–4] G-Quadruplexes are important regulatory elements in the genome^[5,6] but they are also selected in SELEX procedures as versatile aptamers (SELEX = systematic enrichment of ligands by exponential amplification).^[7] For example a simple G-rich DNA 15mer can inhibit blood clotting.^[8] In DNA nanoarchitectures, G-quadruplexes can act as interesting structural scaffolds.^[9–14]

Light is an ideal external trigger signal that can be highly selective and superior to changes in temperature and pH value. Localized irradiation, for example in laser scanning

microscopes, allows for very precise spatiotemporal and dose control,^[15–18] far beyond the precision of the injection of trigger compounds, as recently shown for blood clotting and miRNA.^[19,20]

The ability to control a process by light can be introduced either by using photolabile groups or photoswitches such as azobenzene.^[24,25] The use of photolabile groups has been applied to G-quadruplexes which were either formed or destroyed irreversibly upon light irradiation.^[21–23]

In the G-quadruplex field, Ogasawara and Maeda have used stilbene-type substituents to control the formation of G-quadruplexes by light-induced E–Z isomerization.^[26] Spada et al. controlled the self-assembly of guanosine monomers in a similar fashion.^[27] The formation of G-quadruplexes can also be induced by small molecules that act as molecular chaperons in the absence of cations.^[28–30] Zhou et al. developed azobenzene-containing small-molecule chaperones to regulate G-quadruplex formation.^[31] Tan et al. synthesized an azobenzene-modified antisense DNA linked to a thrombin-binding aptamer to regulate G-quadruplex formation.^[32]

Herein, we present a minimal light-switchable DNA module enabling the formation of an intermolecular and conformationally well-defined G-quadruplex structure with a photoswitchable azobenzene residue as part of the backbone structure (Figure 1).

Azobenzene derivatives **Az1**, **Az2**, and **Az3** (Figure 2a) were employed as photoswitchable linkers between two sets of two consecutive guanosine moieties and were introduced using DNA solid-phase synthesis. The size and substitution patterns on the azo linkers were chosen to offer a suitable balance between the rigidity and flexibility of the overall structure, such that the photoswitch in the E conformation should permit the formation of a G-quadruplex, whereas in the Z conformation no such G-quadruplex formation should be possible. Simple predictions suggested that **Az1** can bridge a distance of 13.2–13.6 Å in the E conformation and 7.7–

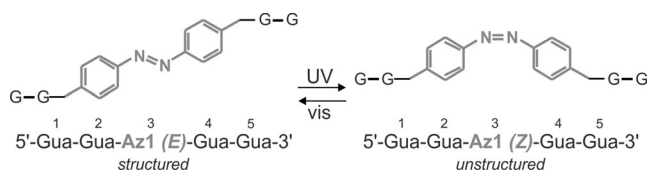


Figure 1. Sequence and numbering of the azobenzene derivative used and characterized in this study, with the azo unit (**Az1**) shown in gray.

[*] J. Thevarpadam,^[+] Dr. D. P. N. Gonçalves, Prof. Dr. A. Heckel
Goethe University Frankfurt
Institute for Organic Chemistry and Chemical Biology
Buchmann Institute for Molecular Life Sciences
Max-von-Laue-Strasse 9, 60438 Frankfurt (Germany)
E-mail: heckel@uni-frankfurt.de
Homepage: <http://photochem.uni-frankfurt.de>

I. Bessi,^[+] O. Binas,^[+] Dr. H. R. A. Jonker, Dr. C. Richter,
Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Center for Biomolecular Magnetic Resonance (BMRZ)
Max-von-Laue-Strasse 9, 60438 Frankfurt (Germany)
E-mail: schwalbe@nmr.uni-frankfurt.de
Homepage: <http://schwalbe.org.chemie.uni-frankfurt.de>

Dr. C. Slavov, Prof. Dr. J. Wachtveitl
Institute for Physical and Theoretical Chemistry
Max-von-Laue-Strasse 7, 60438 Frankfurt (Germany)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201510269>.

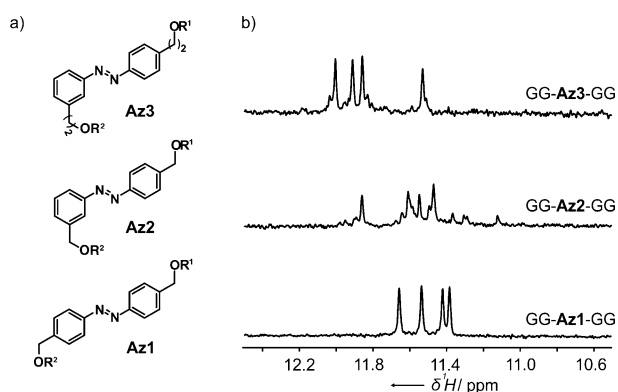


Figure 2. a) The structures of the azo units employed in this study. b) Imino regions of the 1D ^1H NMR spectra of GG-**Az1**-GG, GG-**Az2**-GG, and GG-**Az3**-GG in the presence of KCl (100 mM). Experimental conditions: DNA (50 μM), Tris-HCl buffer (50 mM; pH 7.4), 298 K, 600 MHz.

11.5 Å in the *Z* conformation between the two oxygen atoms adjacent to the azobenzene core. Going from a *para-para* to a *para-meta* substitution pattern in **Az2**, these values change to 11.1–13.6 Å and 9.0–11.9 Å, respectively. The O–O distance at equivalent positions in the narrow grooves in G-quadruplex structures can reach 11–12 Å (see for example PDB 2GKU). We also included the double homologue **Az3** (8.8–14 Å in the *E* conformation and 7.3–12.4 Å in the *Z* conformation). Clearly, minor changes in the structure of the azo linker result in significant changes of the “hinge qualities” of the photoswitch linkers. All of these considerations should be considered with reservation given the highly polymorphic nature of G-quadruplex structures and their respective structural flexibility. We specifically refrained from using longer homologues so as not to dissipate the perturbation induced by the *E-Z* transition into too many internal degrees of freedom.

Initial ^1H NMR (100 mM K^+) characterization of the short azobenzene-linked sequences chosen for this investigation showed signals in the imino region of the spectrum, typical for Hoogsteen-type hydrogen bonds (Figure 2b), suggesting the formation of G-quadruplex structures. Circular dichroism (CD) studies were performed to assess the conformational properties of the modified G-rich sequences. After the addition of 25 mM of K^+ ions, a positive signal around $\lambda = 295$ nm and a negative signal near $\lambda = 260$ nm were detected for all three systems, indicating formation of antiparallel G-quadruplex structures (see Figure 3a for GG-**Az1**-GG and the Supporting Information for GG-**Az2**-GG and GG-**Az3**-GG).^[33,34] Interestingly, ^1H NMR analysis of the different derivatives revealed structural polymorphism for GG-**Az2**-GG and GG-**Az3**-GG which was not detectable by CD. In all three cases, no formation of higher order aggregates was detected (see Figure S58 in the Supporting Information). For details on the thermal stability of the structures formed by the three sequences, please see Figure S5.

Only four ^1H NMR signals for imino groups, indicating a highly symmetric G-quadruplex structure, were detected after addition of K^+ to GG-**Az1**-GG. The ^1H NMR spectra of GG-**Az1**-GG also showed that at a K^+ ion concentration of 25 mM, the intensity of the set of imino signals was already

65% of its maximum value, whereas at a 5 mM K^+ ion concentration, the signal intensity was at 10% of its maximum (Figure 3b). In absence of K^+ ions there is no detectable interaction between the nucleobases, indicating that neither G-quadruplexes nor any other aggregates of the oligonucleotide strands are formed. Additionally, at Na^+ concentrations of up to 500 mM, we did not detect the formation of any secondary structure (Figures 3c and 3d). In this case, G-quadruplex formation was induced only after adding an additional 100 mM of K^+ ions (Figure 3d; Figure S39).

For another recent example of ion selectivity in G-quadruplexes and possible application in nanotechnology, see also Ref. [36]. Cation selectivity has also been reported for the thrombin-binding aptamer G-quadruplex,^[37] whose G-tetrad core adopts the same folding topology as we have determined for GG-**Az1**-GG (see Figure 4 and the corresponding discussion). We propose that the smaller size of the Na^+ ion is not optimal to coordinate all eight O6 atoms in the two-tetrad cavity, which is necessary to keep the GG-**Az1**-GG in a stable quadruplex structure. Furthermore, the azobenzene linkers may introduce additional strain to the ionic channel, leading to the observed ion selectivity.

We investigated structural changes upon photoswitching of the azobenzene units. Two distinguishable sets of CD spectra could be obtained upon irradiation with either UV or visible light (Figure 3e). After an initial irradiation with UV light for 5 min, the CD signal at $\lambda = 295$ nm disappeared completely. Over the course of 30 min, the system reached thermal equilibrium (red data points in Figures 3e–g). Whereas the thermal *Z-E* isomerization occurs over the course of several days (Figure S42), irradiation with visible light for a short period of time (2 min) led to the almost complete recovery of the initial CD signal after 15 min of equilibration (Figure 3f) which could also be paralleled with corresponding NMR experiments (Figure 3h; Figure S60a). ^1H NMR spectroscopy showed the complete disappearance of signals for the imino protons upon irradiation with UV light and their almost complete recovery after irradiation with visible light (Figure 3i, 100 mM of K^+). We found that the degree of recovery after irradiation with visible light is a function of the DNA concentration: although at 50 μM the recovery of signals attributable to a G-quadruplex structure after irradiation with visible light is basically complete, a sample containing 125 μM DNA shows 75% recovery of resonance signals for the G-quadruplex 15 min after irradiation. However, the recovery of the G-quadruplex structure is complete after thermal equilibration (Figure S60a).

We speculate that at DNA concentrations greater than 50 μM , after UV illumination unspecific aggregates are formed (indicated by a broad signal in the aromatic region of the ^1H NMR spectrum) that slowly convert into the quadruplex folded state and/or to the completely unfolded state.

No effects of UV degradation (such as the photo-oxidation of guanine) were detected in the aromatic region of the 1D ^1H NMR spectra after repetitive UV/visible-light irradiation cycles (Figure S60b). Additionally, UV/Vis difference spectra (Figure S6) showed the differential absorption signature typical for G-quadruplexes.^[38]

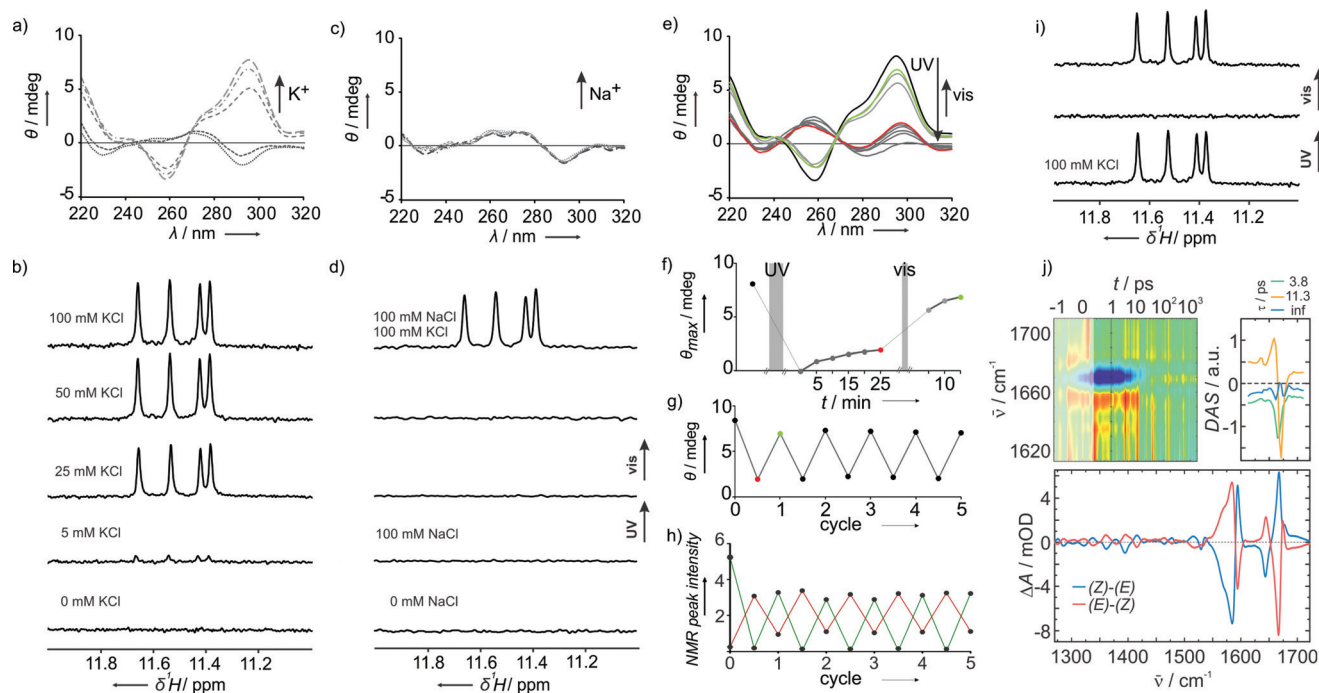


Figure 3. Spectroscopic characterization of the GG-Az1-GG sequence. a) CD spectra recorded with increasing K^+ concentrations (0, 5, 25, 50, 100 mM). b) 1D 1H NMR spectra showing the imino region recorded in the presence of different KCl concentrations. c) CD spectra with increasing concentrations (0, 50, 100, 200, 300, 400, 500 mM) of Na^+ . d) Imino region of the 1D 1H NMR spectra in the presence of increasing amounts of NaCl and KCl. e) CD spectra of GG-Az1-GG with photoirradiation (5 min UV, 2 min visible light) and f) corresponding time course of the signal at $\lambda = 295$ nm under the indicated treatment. Cycling of the photoresponsive structural conversion of GG-Az1-GG by alternate irradiation with UV (15 min) and visible (4 min) light as monitored g) by CD at $\lambda = 295$ nm and h) by 1H NMR ([DNA] = 125 μ M, [KCl] = 100 mM) at $\delta = 11.55$ ppm (green line; monitoring the intensity of the signal for the imino proton of residue G5 in the *E* conformation) and at 6.64 ppm (red line; monitoring the intensity of the signal for the aromatic proton from the azobenzene moiety in the *Z* conformation). Absolute NMR peak intensities are referenced to an internal standard. i) Imino region of the 1D 1H NMR spectra ([DNA] = 50 μ M, [KCl] = 100 mM) showing spectral changes with photoirradiation (30 min UV, 2 min visible light). j) IR difference spectra (bottom); IR transient absorption data recorded from GG-Az1-GG after excitation of the azobenzene moiety at $\lambda = 335$ nm (top left); decay-associated spectra from the global lifetime analysis^[35] of the transient absorption data (top right).

Photoswitching of GG-Az1-GG was also evident in the FTIR difference spectra (Figure 3j, bottom; Figure S53). The azobenzene isomerization and the subsequent disruption of the G-quadruplex structure led to a bathochromic shift ($1675\text{ cm}^{-1} \rightarrow 1667\text{ cm}^{-1}$) of the C=O stretching band^[39–41] and to batho- and hypsochromic shifts of the bands associated with the C=C and C=N purine ring vibrations ($1550\text{--}1590\text{ cm}^{-1}$ range),^[39,41] features reported in the literature to be characteristic of G-quadruplex melting.^[39,41] Ultrafast UV-pump/mid-IR-probe experiments^[42] in the carbonyl stretching vibration range (Figure 3j, top left; Figure S54–55) were performed to investigate the dynamics of the G-quadruplex after *E* azobenzene excitation ($\lambda = 335$ nm). In the first picoseconds after laser excitation, the transient absorption data are dominated by the cooling dynamics of the C=O stretching band ($1645\text{--}1680\text{ cm}^{-1}$) with a lifetime of circa 11 ps. The absorption and bleach bands for the product state become visible after circa 10 ps (see the infinity spectrum in Figure 3j, top right). The last spectrum in the transient absorption data is essentially a *Z–E* IR spectrum of GG-Az1-GG at about 1800 ns. Evidently, this spectrum does not fully match the corresponding FTIR difference spectrum (Figure 3j, bottom), which indicates that despite the nearly instantaneous disruption of the FTIR features of a G-quad-

ruplex, residual conformational dynamics occur on longer time scales.

For studies of GG-Az1-GG derivatives elongated or shortened at the 3'- and 5'-end and for the results obtained with the sequence GGG-Az1-GGG, please see Figures S40/S41 and S57, respectively.

To elucidate the molecular structure of the homogeneously folded GG-Az1-GG, 2D NMR experiments (2D $^1H\text{--}^1H$ NOESY, $^1H\text{--}^{13}C$ HSQC, and $^1H\text{--}^{13}C$ HMBC) were conducted to obtain the complete proton chemical shift assignment. Distance restraints were derived from NOE data and additional angular restraints were obtained from high-resolution $^1H\text{--}^{13}C$ HSQC, 2D $^1H\text{--}^1H$ P.E.COSY, and $^1H\text{--}^{31}P$ TOCSY experiments.^[43] Assignment and *J*-coupling analysis are reported in the Supporting Information.

The folding topology of GG-Az1-GG was determined on the basis of NOESY data. As indicated by intra-tetrad H1–H8 connectivities (Figures S44), GG-Az1-GG adopts a symmetric, antiparallel G-quadruplex structure with tetrads composed as shown in Figure 4 a, b. The edgewise loops containing the azobenzene moieties are located above the G2–G4 tetrad. The anomeric-aromatic region of the NOESY spectrum (Figure S45) indicated a *syn* conformation for the glycosidic bond of residues G1 and G4 and an *anti* conformation for residues G2 and G5.

Intriguingly, the structure of GG-Az1-GG resembles very much the ones of the G-quadruplex formed by the thrombin-binding aptamer (PDB: 148D, NMR; and PDB: 4DII, X-ray)^[37,44] and in the promoter region of the B-raf gene (PDB: 4H29; see Figure S59 for an overlay).^[45]

Hydrogen–deuterium exchange experiments (Figure S46) revealed that imino protons belonging to residues G1 and G5 are protected from solvent exchange. This result suggests a tetrameric G-quadruplex consisting of two dimeric units with the previously defined topology, in which the G1–G5 tetrads of both dimers face each other (Figure 4c). The stacking of two dimeric G-quadruplex units is supported by DOSY data (Figure S47). The structure of the tetramer (Figure 4d) was calculated using ARIA (details in the Supporting Information). The tetrameric arrangement was further supported by the fact that structure calculations as either monomer or dimer led to NOE violations. A significant number of NOEs are unambiguously assigned for the dimer and tetramer (Table 1). The prevalent conformation of the sugar moieties of G1, G2, and G4 is C2'-endo (confirmed by typical strong NOE cross-peaks and ³J coupling constants). Only the 3'-terminal G5 sugar moiety is less well-defined, in part caused by resonance overlap, but manual inspection of the NOE cross-peaks (which are weaker than for the others) also suggest that this sugar possibly interconverts between the C3'-endo and C2'-endo conformations. Therefore no additional torsion angle restraints were included for the G5 nucleoside. The final bundle of structures was refined in explicit water, including three coordinated potassium ions within the tetrads. The structure is well-defined with an average root-mean-square deviation (RMSD) to mean for all atoms of 0.60 Å.

In summary, we have developed a photoswitchable G-quadruplex module and have characterized its photochemical behavior and its 3D conformation. Out of three investigated

Table 1: Statistics of the structure calculation.^[a]

NOE distance restraints	193
Unambiguous NOEs:	146
intra-residue	87
sequential	42
long-range	1
dimer	4
tetramer	12
Ambiguous NOEs:	47
intra-residue	5
intra-monomer	22
monomer or dimer	7
monomer or tetramer	3
dimer or tetramer	8
mono-, di-, or tetramer	2
Distance restraints	14
intra-monomer hydrogen bonds	4
inter-monomer hydrogen bonds	4
potassium site coordination	6
Base planarity	4
intra-monomer	2
inter-monomer	2
Torsion angles	35
backbone	13
β (from ³ J(H5' _{1,2} ,P))	3
glycosidic (χ)	4
sugar pucker	15
Violations	0
distances (>0.3 Å)	0
dihedral angles (>5°)	0
RMSD (average to mean)	
monomer (all atom)	0.55 Å
tetramer (all atom)	0.60 Å

[a] Statistics per monomer.

photoswitchable linkers in a number of sequence contexts, only GG-Az1-GG showed a defined and robust structural behavior. The system shows excellent photocontrol by UV/

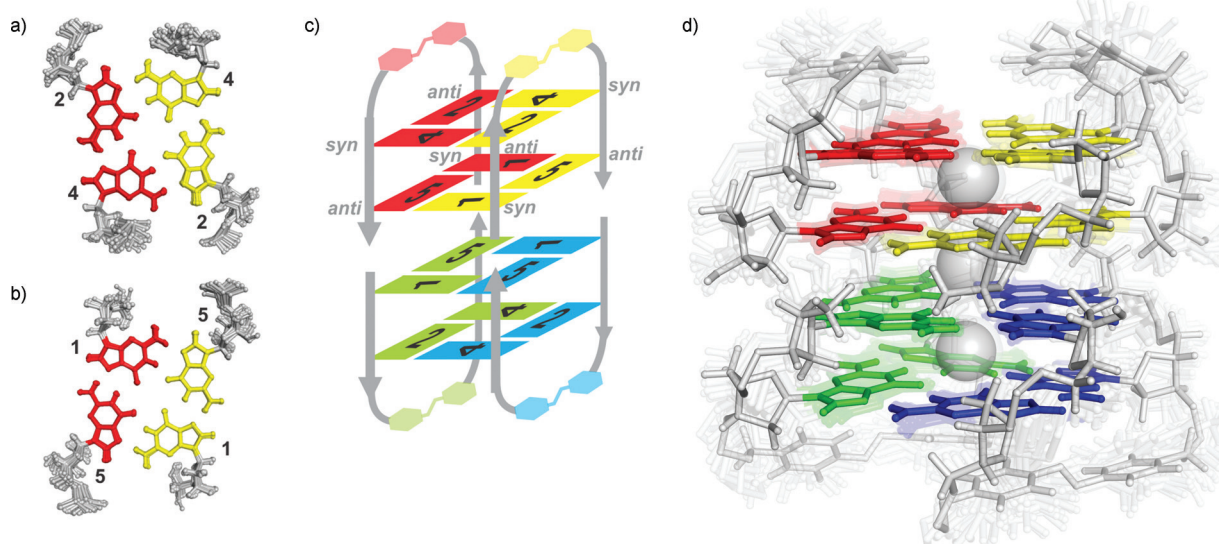


Figure 4. a) The G2–G4 tetrad and b) the G1–G5 tetrad with color coding corresponding both to the structure shown in (c) and to residue numbering given in Figure 1. c) Representation of the structure of the tetramer, with each monomer displayed in a different color. Conformations of the bases (*syn* or *anti*) are indicated. d) NMR solution structure of the G-quadruplex, showing the best representative of the bundle (indicated faintly in the background). K⁺ ions are shown as gray spheres. PDB-code: 2N9Q; BMRB-code: 25915.

Vis irradiation, qualifying it as the smallest photocontrollable DNA switch reported to date. Irradiation with UV light of wavelength $\lambda = 365$ nm and with visible light is known not to be harmful to DNA or to cells and tissues.^[17]

Numerous applications for such a photoswitchable G-quadruplex can be envisioned. The structure may, for example, find application as the glue for photocontrolled (dis)assembly of new fine-tunable nanoarchitectures or as an optomechanical molecular motor.^[46]

Acknowledgements

The authors gratefully acknowledge help by Dr. Anna Lena Lieblein and funding by the Deutsche Forschungsgemeinschaft (DFG) in SFB902 and CLiC. J.W., H.S., and A.H. are members of the DFG-funded cluster of excellence: Macromolecular Complexes (EXC 115). Work in the group of H.S. has been funded by LOEWE project SynChemBio and by the state of Hesse (BMRZ). C.S. and J.W. acknowledge funding by the DFG (WA 1850/4-1).

Keywords: azobenzene · DNA structures · G-quadruplexes · IR spectroscopy · NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 2738–2742
Angew. Chem. **2016**, *128*, 2788–2792

-
- [1] T. Simonsson, *Biol. Chem.* **2001**, *382*, 621–628.
- [2] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402–5415.
- [3] A. I. Karsisiotis, C. O’Kane, M. Webba da Silva, *Methods* **2013**, *64*, 28–35.
- [4] I. Bessi, H. R. A. Jonker, C. Richter, H. Schwalbe, *Angew. Chem. Int. Ed.* **2015**, *54*, 8444–8448; *Angew. Chem.* **2015**, *127*, 8564–8568.
- [5] D. Rhodes, H. J. Lipps, *Nucleic Acids Res.* **2015**, *43*, 8627–8637.
- [6] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
- [7] W. O. Tucker, K. T. Shum, J. A. Tanner, *Curr. Pharm. Des.* **2012**, *18*, 2014–2026.
- [8] K. Y. Wang, S. McCurdy, R. G. Shea, S. Swaminathan, P. H. Bolton, *Biochemistry* **1993**, *32*, 1899–1904.
- [9] E. A. Venczel, D. Sen, *J. Mol. Biol.* **1996**, *257*, 219–224.
- [10] Z. Li, C. A. Mirkin, *J. Am. Chem. Soc.* **2005**, *127*, 11568–11569.
- [11] M. Biyani, K. Nishigaki, *Gene* **2005**, *364*, 130–138.
- [12] T. C. Marsh, J. Vesenska, E. Henderson, *Nucleic Acids Res.* **1995**, *23*, 696–700.
- [13] D. P. N. Gonçalves, T. L. Schmidt, M. B. Koeppel, A. Heckel, *Small* **2010**, *6*, 1347–1352.
- [14] F. Wang, X. Liu, I. Willner, *Angew. Chem. Int. Ed.* **2015**, *54*, 1098–1129; *Angew. Chem.* **2015**, *127*, 1112–1144.
- [15] R. H. Kramer, D. L. Fortin, D. Trauner, *Curr. Opin. Neurobiol.* **2009**, *19*, 544–552.
- [16] A. A. Beharry, G. A. Woolley, *Chem. Soc. Rev.* **2011**, *40*, 4422–4437.
- [17] C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer, A. Heckel, *Angew. Chem. Int. Ed.* **2012**, *51*, 8446–8476; *Angew. Chem.* **2012**, *124*, 8572–8604.
- [18] L. Wu, K. Koumoto, N. Sugimoto, *Chem. Commun.* **2009**, 1915–1917.
- [19] F. Rohrbach, F. Schäfer, M. A. H. Fichte, F. Pfeiffer, J. Müller, B. Pötzsch, A. Heckel, G. Mayer, *Angew. Chem. Int. Ed.* **2013**, *52*, 11912–11915; *Angew. Chem.* **2013**, *125*, 12129–12132.
- [20] F. Schäfer, J. Wagner, A. Knau, S. Dimmeler, A. Heckel, *Angew. Chem. Int. Ed.* **2013**, *52*, 13558–13561; *Angew. Chem.* **2013**, *125*, 13801–13805.
- [21] G. Mayer, L. Kröck, V. Mikat, M. Engeser, A. Heckel, *ChemBioChem* **2005**, *6*, 1966–1970.
- [22] A. Heckel, M. C. R. Buff, M. L. Raddatz, J. Müller, B. Pötzsch, G. Mayer, *Angew. Chem. Int. Ed.* **2006**, *45*, 6748–6750; *Angew. Chem.* **2006**, *118*, 6900–6902.
- [23] T. L. Schmidt, M. B. Koeppel, J. Thevarpadam, D. P. N. Gonçalves, A. Heckel, *Small* **2011**, *7*, 2163–2167.
- [24] H. Ito, X. Liang, H. Nishioka, H. Asanuma, *Org. Biomol. Chem.* **2010**, *8*, 5519–5524.
- [25] H. Nishioka, X. Liang, T. Kato, H. Asanuma, *Angew. Chem. Int. Ed.* **2012**, *51*, 1165–1168; *Angew. Chem.* **2012**, *124*, 1191–1194.
- [26] S. Ogasawara, M. Maeda, *Angew. Chem. Int. Ed.* **2009**, *48*, 6671–6674; *Angew. Chem.* **2009**, *121*, 6799–6802.
- [27] S. Lena, P. Neviani, S. Masiero, S. Pieraccini, G. P. Spada, *Angew. Chem. Int. Ed.* **2010**, *49*, 3657–3660; *Angew. Chem.* **2010**, *122*, 3739–3742.
- [28] D. P. N. Gonçalves, R. Rodriguez, S. Balasubramanian, J. K. M. Sanders, *Chem. Commun.* **2006**, 4685–4687.
- [29] D. P. N. Gonçalves, S. Ladame, S. Balasubramanian, J. K. M. Sanders, *Org. Biomol. Chem.* **2006**, *4*, 3337–3342.
- [30] R. Rodriguez, G. D. Pantoş, D. P. N. Gonçalves, J. K. M. Sanders, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2007**, *46*, 5405–5407; *Angew. Chem.* **2007**, *119*, 5501–5503.
- [31] X. Wang, J. Huang, Y. Zhou, S. Yan, X. Weng, X. Wu, M. Deng, X. Zhou, *Angew. Chem. Int. Ed.* **2010**, *49*, 5305–5309; *Angew. Chem.* **2010**, *122*, 5433–5437.
- [32] Y. Kim, J. A. Phillips, H. Liu, H. Kang, W. Tan, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6489–6494.
- [33] S. Paramasivan, I. Rujan, P. H. Bolton, *Methods* **2007**, *43*, 324–331.
- [34] A. Randazzo, G. P. Spada, M. Webba da Silva, *Top. Curr. Chem.* **2012**, *330*, 67–86.
- [35] C. Slavov, H. Hartmann, J. Wachtveitl, *Anal. Chem.* **2015**, *87*, 2328–2336.
- [36] L. Olejko, P. J. Cywinski, I. Bald, *Angew. Chem. Int. Ed.* **2015**, *54*, 673–677; *Angew. Chem.* **2015**, *127*, 683–687.
- [37] I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella, F. Sica, *Nucleic Acids Res.* **2012**, *40*, 8119–8128.
- [38] J.-L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, e138.
- [39] H. T. Miles, J. Frazier, *Biochem. Biophys. Res. Commun.* **1972**, *49*, 199–204.
- [40] M. R. Guzmán, J. Liquier, S. K. Brahmachari, E. Taillandier, *Spectrochim. Acta Part A* **2006**, *64*, 495–503.
- [41] J. A. Walmsley, M. L. Schneider, P. J. Farmer, J. R. Cave, C. R. Toth, R. M. Wilson, *J. Biomol. Struct. Dyn.* **1992**, *10*, 619–638.
- [42] K. Neumann, M.-K. Verhoefen, I. Weber, C. Glaubitz, J. Wachtveitl, *Biophys. J.* **2008**, *94*, 4796–4807.
- [43] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631–644.
- [44] P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532–1547.
- [45] D. Wei, A. K. Todd, M. Zloh, M. Gunaratnam, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2013**, *135*, 19319–19329.
- [46] M. McCullagh, I. Franco, M. A. Ratner, G. C. Schatz, *J. Am. Chem. Soc.* **2011**, *133*, 3452–3459.

Received: November 4, 2015

Revised: December 11, 2015

Published online: January 25, 2016

Supporting Information

Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif

Julie Thevarpadam⁺, Irene Bessi⁺, Oliver Binas⁺, Diana P. N. Gonçalves, Chavdar Slavov, Hendrik R. A. Jonker, Christian Richter, Josef Wachtveitl, Harald Schwalbe, and Alexander Heckel**

anie_201510269_sm_miscellaneous_information.pdf

Supporting Information

Table of Contents

Synthesis of azobenzene-derived phosphoramidites	1
Oligonucleotide synthesis and purification.....	5
Characterization of oligonucleotides	5
Photoisomerization	6
CD measurements	7
UV/vis measurements	8
Native gel electrophoresis.....	10
NMR-spectra of synthesized compounds.....	11
NMR measurements.....	25
Screening of the azobenzene derivatives GG-Az1-GG, GG-Az2-GG and GG-Az3-GG using 1D ¹ H-NMR spectra	26
GG-Az1-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl.....	26
GG-Az2-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl.....	27
GG-Az3-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl.....	28
Ion selectivity of GG-Az1-GG: titration with NaCl and KCl followed by UV/vis irradiation at (100 mM NaCl + 100 mM KCl).....	29
Screening of the azobenzene derivatives GG-Az1-GGG, GGG-Az1-GG, G-Az1-GG and GG-Az1-G using 1D ¹ H-NMR spectra	30
Assignment of the 1D ¹ H-NMR spectrum of GG-Az1-GG	32
Full assignment of the 2D ¹ H, ¹ H-NOESY spectrum of GG-Az1-GG	33
2D ¹ H, ¹³ C-HMBC spectrum of GG-Az1-GG: C5 region.....	34
2D ¹ H, ¹ H-NOESY spectrum: H1-H8 and H8-H1' region	34
H-D exchange experiment: imino region of the 1D ¹ H-NMR spectrum of GG-Az1-GG before and after exchange with D ₂ O	35
DOSY spectrum of GG-Az1-GG	36

2D ^1H , ^1H -P.E.COSY of GG-Az1-GG: H1'-H2' and H1'-H2'' region	37
2D ^1H , ^{13}C -HSQC of GG-Az1-GG: Zoom on a H5'-C5' peak.....	37
2D ^1H , ^{31}P -hetero TOCSY-NOESY of GG-Az1-GG:.....	38
2D ^1H , ^{13}C -HSQC of GG-Az1-GG:	39
C3'-H3', C5'-H5'/H5'', C4'-H4' and C2'-H2'/H2'' region	39
Structure calculations.....	40
Chemical Shift Resonances.....	41
3J coupling constants	42
FTIR measurements.....	42
FTIR absorption spectra of GG-Az1-GG	43
Ultrafast UV-pump/mid-IR-probe experiments	44
Investigation of the azobenzene derivative GGG-Az1-GGG using 1D ^1H -NMR- and CD-spectroscopy and native gel electrophoresis	46
Native gel electrophoresis of GG-Az1-GG, GG-Az2-GG and GG-Az3-GG.....	47
Structural comparison between the thrombin-binding aptamer and GG-Az1-GG	48
Investigation of photo-fatigue by ^1H -NMR spectroscopy	49

Synthesis of azobenzene-derived phosphoramidites

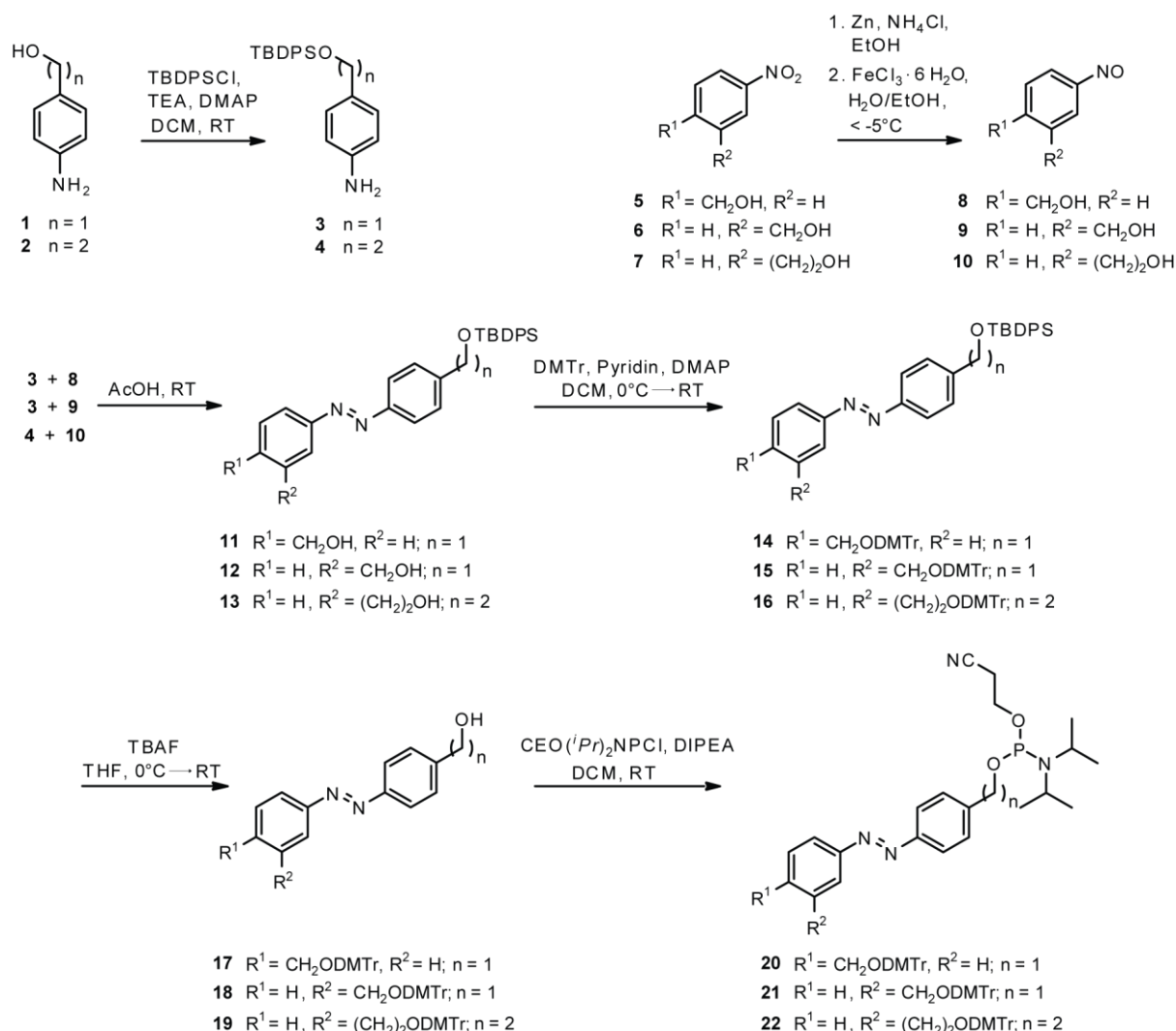


Figure S1. Synthesis scheme of azobenzene-derived phosphoramidites.

General: Compounds **3** and **4** were synthesized according to literature procedure.^[1] All reactions were performed under argon atmosphere using solvents of p.a. quality or higher. Dry solvents were purchased in sealed bottles over molecular sieves. For flash chromatography silica gel 60 by Macherey Nagel was used. TLC analysis was performed on aluminium plates coated with silica gel60F 254 (Merck). NMR spectra were recorded on BRUKER AV 300, AV 400 MHz or AV 500 MHz instruments. HRMS spectra were recorded on Thermo Scientific MALDI LTQ Orbitrap.

General Procedure I: Nitrobenzyl alcohol was dissolved in ethanol and ammonium chloride in water and zinc powder were added. After being stirred for 30 min the insoluble matter was filtered off and the solution was added dropwise to another solution of iron chloride hexahydrate in water and ethanol at -5°C . The solution turned green and it was stirred for another 30 min keeping the temperature below -5°C . After adding brine (50 mL) the solution was extracted with dichloromethane. The organic layer was washed with water, dried over NaSO_4 and filtered. After removing the solvent under reduced pressure the crude product was obtained which was used without further purification.

General Procedure II: The nitroso compound and the amino compound were mixed in acetic acid and stirred overnight at room temperature. The solvent was evaporated. Flash chromatography (cyclohexane/ ethyl acetate 8:2) of the remaining residue afforded the corresponding azobenzene derivative.

General Procedure III: The azobenzene derivative was dissolved in dichloromethane and cooled to 0°C. Pyridine, 4-dimethylaminopyridine and dimethoxytritylchloride were added and the solution was stirred overnight without further cooling. Dichloromethane and pyridine were removed under reduced pressure. After flash chromatography (cyclohexane/ethyl acetate 9:1, for column packing 1% NEt₃ was used) the product was obtained.

General Procedure IV: The azobenzene derivative was dissolved in tetrahydrofuran and tetrabutylammonium fluoride was added at 0°C. The solution was stirred for 2 h at rt. Afterwards, the solvent was evaporated and the remaining residue was purified by flash chromatography (cyclohexane/ethyl acetate 7:3, for column packing 1% NEt₃ was used).

General Procedure V: The azobenzene derivative was dissolved in dry dichloromethane, *N,N*-diisopropylethylamine was added and the reaction mixture was stirred for 5 min. Then cyanoethoxy-*N,N*-diisopropylamine was added via a syringe and the reaction was stirred for 2 h. After diluting the reaction mixture with dichloromethane (15 mL), the solution was washed with saturated NaHCO₃ solution and dried over NaSO₄. After filtration the solvent was evaporated under reduced pressure and the residue was purified by flash chromatography (cyclohexane/ethyl acetate 8:2, for column packing 1% NEt₃ was used).

Synthesis of compound 8: According to General Procedure I 4-nitrobenzyl alcohol **5** (3.5 g, 21.0 mmol, 1 equiv) in ethanol (40 mL) and ammonium chloride (1.6 g, 29.4 mmol, 1.4 equiv) in water (10 mL) and zinc powder (4.1 g, 63.0 mmol, 3 equiv) were used with an iron chloride hexahydrate solution (6.8 g, 25.2 mmol, 1.2 equiv) in water (20 mL) and ethanol (8 mL). The crude product was obtained as a green oil which was used without further purification (1.6 g).

Synthesis of compound 9: According to General Procedure I 3-nitrobenzyl alcohol **6** (3.5 g, 2.7 mL, 21.0 mmol, 1 equiv) in ethanol (40 mL) and ammonium chloride (1.6 g, 29.4 mmol, 1.4 equiv) in water (10 mL) and zinc powder (4.1 g, 63.0 mmol, 3 equiv) were used with an iron chloride hexahydrate solution (6.8 g, 25.2 mmol, 1.2 equiv) in water (20 mL) and ethanol (8 mL). The crude product was obtained as a green oil which was used without further purification (1.2 g).

Synthesis of compound 10: According to General Procedure I 3-nitrophenethyl alcohol **7** (3.5 g, 20.9 mmol, 1 equiv) in ethanol (40 mL) and ammonium chloride (1.6 g, 29.3 mmol, 1.4 equiv) in water (10 mL) and zinc powder (4.1 g, 62.7 mmol, 3 equiv) were used with an iron chloride hexahydrate solution (6.8 g, 25.1 mmol, 1.2 equiv) in water (20 mL) and ethanol (8 mL). The crude product was obtained as a green oil which was used without further purification (2.6 g).

Synthesis of compound 11: According to General Procedure II compounds **3** (4.2 g, 11.6 mmol) and **8** (1.6 g) in acetic acid (40 mL) were used. Flash chromatography afforded compound **11** as an orange solid (2.1 g, 4.5 mmol, 56%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.90-7.84 (m, 4H, Ar-H), 7.69-7.64 (m, 4H, Ar-H), 7.55 (t, 4H, ³J(H,H)=8.3 Hz, Ar-H), 7.48-7.39 (m, 6H, Ar-H), 5.38 (t, ³J(H,H)=5.7 Hz, 1H, OH), 4.88 (s, 2H, CH₂), 4.62 (d, ³J(H,H)=5.7 Hz, 2H, CH₂), 1.06 ppm (s, 9H, 3xCH₃); ¹³C-NMR (101 MHz, [D₆]DMSO): δ=151.1, 150.9, 146.3, 144.0, 135.0, 132.7, 130.0, 128.0, 127.1, 126.7, 126.6, 122.5, 122.4, 64.7, 62.5, 26.7, 18.9 ppm. HRMS (MALDI): *m/z* calcd for C₃₀H₃₃N₂O₂Si [*M* + H]⁺ 481.23058, found 481.23006 (Δ*m*=0.00052, error 1.1 ppm).

Synthesis of compound 12: According to General Procedure II compounds **3** (3.2 g, 8.8 mmol) and **9** (1.2 g) in acetic acid (40 mL) were used. Flash chromatography afforded compound **12** as an orange solid (1.8 g, 3.7 mmol, 47%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.92-7.87 (m, 2H, Ar-H), 7.85 (s, 1H, Ar-H), 7.80-7.75 (m, 1H, Ar-H), 7.70-7.64 (m, 4H, Ar-H), 7.62-7.50 (m, 4H, Ar-H), 7.50-7.41 (m, 6H, Ar-H), 5.37 (t, ³J(H,H)=5.8 Hz, 1H, OH), 4.89 (s, 2H, CH₂), 4.63 (d, ³J(H,H)=5.8 Hz, 2H, CH₂), 1.06 ppm (s, 9H, 3xCH₃); ¹³C-NMR (100 MHz, [D₆]DMSO): δ=151.9, 151.1, 144.2, 144.2, 135.0, 132.7, 130.0, 129.3, 129.1, 128.0, 126.6, 122.6, 121.7, 119.5, 64.7, 62.4, 26.7, 18.9 ppm. HRMS (MALDI): *m/z* calcd for C₃₀H₃₃N₂O₂Si [*M* + H]⁺ 481.23058, found 481.23010 (Δ*m*=0.00048, error 1.0 ppm).

Synthesis of compound 13: According to General Procedure II compounds **4** (5.0 g, 13.3 mmol) and **10** (2.6 g) in acetic acid (40 mL) were used. Flash chromatography afforded compound **13** as an orange solid (3.9 g, 7.7 mmol, 58%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.84-7.79 (m, 2H, Ar-H), 7.77-7.74 (m, 1H, Ar-H), 7.74-7.70 (m, 1H, Ar-H), 7.55-7.50 (m, 4H, Ar-H), 7.50-7.43 (m, 4H, Ar-H), 7.43-7.35 (m, 6H, Ar-H), 4.69 (t, ³J(H,H)=5.2 Hz, 1H, OH), 3.88 (t, ³J(H,H)=6.3 Hz, 2H, CH₂), 3.72-3.65 (m, 2H, CH₂), 2.92 (t, ³J(H,H)=6.3 Hz, 2H, CH₂), 2.85 (t, ³J(H,H)=6.6 Hz, 2H, CH₂), 0.94 ppm (s, 9H, 3xCH₃); ¹³C-NMR (101 MHz, [D₆]DMSO): δ=151.9, 150.6, 143.0, 141.2, 135.0, 133.0, 132.0, 130.1, 129.8, 129.4, 127.8, 122.7, 122.3, 120.3, 120.0, 64.2, 61.9, 38.3, 26.6, 18.7 ppm. HRMS (MALDI): *m/z* calcd for C₃₂H₃₇N₂O₂Si [*M* + H]⁺ 509.26188, found 509.26102 (Δ*m*=0.00086, error 1.7 ppm).

Synthesis of compound 14: According to General Procedure III compound **11** (2.1 g, 4.5 mmol, 1 equiv) dissolved in dichloromethane (10 mL), pyridine (8 mL), 4-dimethylaminopyridine (135 mg, 1.1 mmol, 0.25 equiv) and dimethoxytritylchloride (4.6 g, 13.5 mmol, 3 equiv) were used. After flash chromatography the product was obtained as an orange foam (3.5 g, 4.5 mmol, 99%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.91-7.84 (m, 4H, Ar-H), 7.69-7.64 (m, 4H, Ar-H), 7.57 (d, ³J(H,H)=8.2 Hz, 4H, Ar-H), 7.50-7.43 (m, 8H, Ar-H), 7.36-7.32 (m, 6H, Ar-H), 7.26-7.21 (m, 1H, Ar-H), 6.95-6.90 (m, 4H, Ar-H), 4.88 (s, 2H, CH₂), 4.22 (s, 2H, CH₂), 3.76 (s, 6H, 2xCH₃), 1.08 ppm (s, 9H, 3xCH₃); ¹³C-NMR (101 MHz, [D₆]DMSO): δ=158.2, 151.1, 151.1, 144.8, 144.2, 142.3, 135.5, 135.0, 132.7, 130.0, 129.6, 128.0, 128.0, 127.6, 127.5, 126.8, 126.6, 122.6, 122.5, 113.3, 86.1, 64.7, 64.5, 55.0, 26.7, 18.9 ppm. HRMS (MALDI): *m/z* calcd for C₅₁H₅₁N₂O₄Si [*M* + H]⁺ 783.36126, found 783.36188 (Δ*m*=0.00062, error 0.8 ppm).

Synthesis of compound 15: According to General Procedure III compound **12** (1.8 g, 3.7 mmol, 1 equiv) dissolved in dichloromethane (10 mL), pyridine (8 mL), 4-dimethylaminopyridine (114 mg, 0.9 mmol, 0.25 equiv) and dimethoxytritylchloride (3.8 g, 11.2 mmol, 3 equiv) were used. After flash chromatography the product was obtained as an orange foam (2.7 g, 3.4 mmol, 92%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.92-7.86 (m, 2H, Ar-H), 7.82 (s, 1H, Ar-H), 7.81-7.77 (m, 1H, Ar-H), 7.69-7.64 (m, 4H, Ar-H), 7.60-7.53 (m, 4H, Ar-H), 7.50-7.42 (m, 8H, Ar-H), 7.36-7.32 (m, 6H, Ar-H), 7.26-7.21 (m, 1H, Ar-H), 6.95-6.89 (m, 4H, Ar-H), 4.88 (s, 2H, CH₂), 4.23 (s, 2H, CH₂), 3.73 (s, 6H, 2xCH₃), 1.06 ppm (s, 9H, 3xCH₃); ¹³C-NMR (100 MHz, [D₆]DMSO): δ=158.2, 152.0, 151.0, 144.9, 144.3, 140.3, 135.6, 135.0, 132.7, 130.0, 129.6, 129.6, 129.4, 128.0, 128.0, 127.6, 126.8, 126.6, 122.6, 121.3, 120.6, 113.3, 86.1, 64.7, 64.6, 55.0, 26.7, 18.9 ppm. HRMS (MALDI): *m/z* calcd for C₅₁H₅₁N₂O₄Si [*M* + H]⁺ 783.36126, found 783.36532 (Δ*m*=0.00406, error 5.2 ppm).

Synthesis of compound 16: According to General Procedure III compound **13** (3.5 g, 6.8 mmol, 1 equiv) dissolved in dichloromethane (10 mL), pyridine (8 mL), 4-dimethylaminopyridine (0.2 g, 1.7 mmol, 0.25 equiv) and dimethoxytritylchloride (5.0 g, 14.8 mmol, 2 equiv) were used. After flash chromatography the product was obtained as an orange foam (3.6 g, 3.4 mmol, 76%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.84-7.79 (m, 2H, Ar-H), 7.79-7.74 (m, 2H, Ar-H), 7.56-7.50 (m, 4H, Ar-H), 7.45-7.35 (m, 10H, Ar-H), 7.31-7.26 (m, 2H, Ar-H), 7.26-7.20 (m, 2H, Ar-H), 7.19-7.13 (m, 5H, Ar-H), 6.84-6.78 (m, 4H, Ar-H), 3.89 (t, ³J(H,H)=6.2 Hz, 2H, CH₂), 3.68 (s, 6H, 2xCH₃), 3.22 (t, ³J(H,H)=6.2 Hz, 2H, CH₂), 2.98-2.88 (m, 4H, 2xCH₂), 0.95 ppm (s, 9H, 3xCH₃); ¹³C-NMR (101 MHz, [D₆]DMSO): δ=157.9, 151.9, 150.6, 145.0, 143.2, 140.9, 135.8, 135.0, 133.0, 130.2, 129.7, 129.6, 129.1, 127.8, 127.7, 127.6, 126.5, 122.5, 122.3, 121.0, 113.0, 85.4, 64.4, 63.9, 54.9, 38.2, 35.6, 26.6, 18.7 ppm. HRMS (MALDI): *m/z* calcd for C₅₃H₅₅N₂O₄Si [*M* + H]⁺ 811.39256, found 811.39093 (Δ*m*=0.00163, error 2.0 ppm).

Synthesis of compound 17: According to General Procedure IV a solution of compound **14** (2.6 g, 3.4 mmol, 1 equiv) in tetrahydrofuran (20 mL) and tetrabutylammonium fluoride (3.4 mL of a 1.0 M solution in THF, 3.4 mmol, 1 equiv) were used. Flash chromatography afforded an orange foam (1.7 g, 3.2 mmol, 93%).

¹H-NMR (400 MHz, [D₆]DMSO): δ=7.90-7.84 (m, 4H, Ar-H), 7.59-7.51 (m, 4H, Ar-H), 7.49-7.44 (m, 2H, Ar-H), 7.36-7.31 (m, 6H, Ar-H), 7.27-7.23 (m, 1H, Ar-H), 6.96-6.90 (m, 4H, Ar-H), 5.39 (t, ³J(H,H)=4.6 Hz, 1H, OH), 4.61 (d, ³J(H,H)=4.5, 2H, CH₂), 4.20 (s, 2H, CH₂), 3.74 ppm (s, 6H, 2xCH₃); ¹³C-NMR (126 MHz, [D₆]DMSO): δ=158.6, 151.6, 151.3, 146.9, 145.3, 142.7, 136.0, 130.1, 128.5, 128.1, 127.9,

127.9, 127.6, 127.3, 123.0, 122.9, 113.8, 86.6, 65.0, 62.9, 55.5 ppm. HRMS (MALDI): m/z calcd for $C_{35}H_{33}N_2O_4$ [$M + H$]⁺ 545.24348, found 545.24126 ($\Delta m=0.00222$, error 4.1 ppm).

Synthesis of compound 18: According to General Procedure IV a solution of compound **15** (1.7 g, 2.2 mmol, 1 equiv) in tetrahydrofuran (20 mL) and tetrabutylammonium fluoride (2.2 mL of a 1.0 M solution in THF, 2.2 mmol, 1 equiv) were used. Flash chromatography afforded an orange foam (1.1 g, 2.1 mmol, 95%).

¹H-NMR (400 MHz, [D₆]DMSO): δ =7.84-7.80 (m, 2H, Ar-H), 7.83-7.77 (m, 2H, Ar-H), 7.60-7.51 (m, 4H, Ar-H), 7.49-7.44 (m, 2H, Ar-H), 7.38-7.31 (m, 6H, Ar-H), 7.28-7.22 (m, 1H, Ar-H), 6.96-6.90 (m, 4H, Ar-H), 5.36 (t, ³J(H,H)=5.7 Hz, 1H, OH), 4.61 (d, ³J(H,H)=5.4, 2H, CH₂), 4.23 (s, 2H, CH₂), 3.74 ppm (s, 6H, 2xCH₃); ¹³C-NMR (100 MHz, [D₆]DMSO): δ =158.2, 152.0, 150.8, 146.5, 144.9, 140.3, 135.6, 129.7, 129.5, 129.4, 128.0, 127.6, 127.1, 126.8, 122.5, 121.4, 120.5, 113.3, 86.1, 64.6, 62.5, 55.0 ppm. HRMS (MALDI): m/z calcd for $C_{35}H_{33}N_2O_4$ [$M + H$]⁺ 545.24348, found 545.24397 ($\Delta m=0.00049$, error 0.9 ppm).

Synthesis of compound 19: According to General Procedure IV a solution of compound **16** (3.6 g, 4.5 mmol, 1 equiv) in tetrahydrofuran (20 mL) and tetrabutylammonium fluoride (4.5 mL of a 1.0 M solution in THF, 4.5 mmol, 1 equiv) were used. Flash chromatography afforded an orange foam (2.4 g, 4.2 mmol, 93%).

¹H-NMR (500 MHz, [D₆]DMSO): δ =7.84-7.79 (m, 2H, Ar-H), 7.78-7.73 (m, 2H, Ar-H), 7.51 (t, ³J(H,H)=7.7 Hz, 1H, Ar-H), 7.47-7.42 (m, 2H, Ar-H), 7.42-7.37 (m, 1H, Ar-H), 7.30-7.26 (m, 2H, Ar-H), 7.25-7.21 (m, 2H, Ar-H), 7.19-7.13 (m, 5H, Ar-H), 6.83-6.79 (m, 4H, Ar-H), 4.73 (t, ³J(H,H)=5.2 Hz, 1H, OH), 3.71-3.65 (m, 8H, 2xCH₃, CH₂), 3.21 (t, ³J(H,H)=6.3 Hz, 2H, CH₂), 2.94 (t, ³J(H,H)=6.3 Hz, 2H, CH₂), 2.83 ppm (t, ³J(H,H)=6.8 Hz, 2H, CH₂); ¹³C-NMR (126 MHz, [D₆]DMSO): δ =158.4, 152.4, 150.9, 145.5, 144.3, 141.3, 136.2, 132.5, 130.4, 130.0, 129.7, 128.2, 128.1, 127.0, 123.0, 122.9, 121.5, 113.5, 85.9, 64.5, 62.3, 55.4, 36.1 ppm. HRMS (MALDI): m/z calcd for $C_{37}H_{37}N_2O_4$ [$M + H$]⁺ 573.27478, found 573.27405 ($\Delta m=0.00073$, error 1.3 ppm).

Synthesis of compound 20: According to General Procedure V compound **17** (225 mg, 413 μ mol, 1 equiv) was in dry dichloromethane (5 mL), *N,N*-diisopropylethylamine (351 μ L, 2.1 mmol, 5 equiv) and cyanoethoxy-*N,N*-diisopropylamine (184 μ L, 826 μ mol, 2 equiv) were used. Flash chromatography afforded the product as an orange foam (210 mg, 282 μ mol, 68%).

¹H-NMR (500 MHz, [D₆]DMSO): δ =7.92-7.85 (m, 4H, Ar-H), 7.61-7.55 (m, 4H, Ar-H), 7.48-7.44 (m, 2H, Ar-H), 7.38-7.30 (m, 6H, Ar-H), 7.29-7.23 (m, 1H, Ar-H), 6.96-6.90 (m, 4H, Ar-H), 4.86-4.73 (m, 2H, CH₂), 4.20 (s, 2H, CH₂), 3.86-3.75 (m, 2H, CH₂), 3.76 (s, 6H, 2xCH₃), 3.68-3.59 (m, 2H, 2xCH), 2.80 (t, ³J(H,H)=5.8 Hz, 2H, CH₂), 1.20-1.14 ppm (m, 12H, 4xCH₃); ¹³C-NMR (126 MHz, [D₆]DMSO): δ =158.6, 151.7, 151.5, 145.3, 143.2, 143.2, 142.9, 136.0, 130.1, 130.0, 128.5, 128.2, 128.1, 127.9, 127.3, 123.1, 123.0, 119.5, 113.8, 86.6, 65.0, 64.8, 64.7, 59.0, 58.8, 55.5, 42.5, 43.1, 43.0, 24.9, 24.8, 20.4, 20.3 ppm; ³¹P-NMR (202 MHz, [D₆]DMSO): δ =147.77 ppm. HRMS (MALDI): m/z calcd for $C_{44}H_{50}N_4O_5P$ [$M + H$]⁺ 745.35133, found 745.35196 ($\Delta m=0.00063$, error 0.8 ppm).

Synthesis of compound 21: According to General Procedure V compound **18** (200 mg, 367 μ mol, 1 equiv) was in dry dichloromethane (5 mL), *N,N*-diisopropylethylamine (312 μ L, 1.8 mmol, 5 equiv) and cyanoethoxy-*N,N*-diisopropylamine (164 μ L, 734 μ mol, 2 equiv) were used. Flash chromatography afforded the product as an orange foam (176 mg, 237 μ mol, 64%).

¹H-NMR (300 MHz, [D₆]DMSO): δ =7.85-7.79 (m, 2H, Ar-H), 7.78-7.73 (m, 2H, Ar-H), 7.61-7.53 (m, 4H, Ar-H), 7.50-7.43 (m, 2H, Ar-H), 7.39-7.30 (m, 6H, Ar-H), 7.28-7.23 (m, 1H, Ar-H), 6.96-6.89 (m, 4H, Ar-H), 4.88-4.71 (m, 2H, CH₂), 4.23 (s, 2H, CH₂), 3.88-3.75 (m, 2H, CH₂), 3.74 (s, 6H, 2xCH₃), 3.69-3.56 (m, 2H, 2xCH), 2.80 (t, ³J(H,H)=5.8 Hz, 2H, CH₂), 1.21-1.13 ppm (m, 12H, 4xCH₃); ¹³C-NMR (75 MHz, [D₆]DMSO): δ =158.1, 151.9, 151.1, 144.8, 142.8, 140.2, 135.5, 129.6, 129.4, 127.9, 127.6, 127.6, 126.7, 122.6, 121.4, 120.5, 119.0, 113.3, 86.1, 64.5, 58.5, 58.2, 55.0, 42.7, 42.5, 24.4, 24.3, 19.9, 19.8 ppm; ³¹P-NMR (121 MHz, [D₆]DMSO): δ =147.74 ppm. HRMS (MALDI): m/z calcd for $C_{44}H_{50}N_4O_5P$ [$M + H$]⁺ 745.35133, found 745.35297 ($\Delta m=0.00164$, error 2.2 ppm).

Synthesis of compound 22: According to General Procedure V compound **19** (150 mg, 262 μ mol, 1 equiv) was in dry dichloromethane (5 mL), *N,N*-diisopropylethylamine (223 μ L, 1.3 mmol, 5 equiv) and

cynoethoxy-*N,N'*-diisopropylamine (117 μ L, 524 μ mol, 2 equiv) were used. Flash chromatography afforded the product as an orange foam (138 mg, 179 μ mol, 69%).

$^1\text{H-NMR}$ (400 MHz, $[\text{D}_6]\text{DMSO}$): δ =7.85-7.79 (m, 2H, Ar-H), 7.78-7.73 (m, 2H, Ar-H), 7.55-7.45 (m, 3H, Ar-H), 7.43-7.37 (m, 1H, Ar-H), 7.31-7.20 (m, 4H, Ar-H), 7.19-7.11 (m, 5H, Ar-H), 6.84-6.78 (m, 4H, Ar-H), 3.95-3.77 (m, 2H, CH_2), 3.71-3.60 (m, 8H, $2\times\text{CH}_3$, CH_2), 3.59-3.45 (m, 2H, $2\times\text{CH}$), 3.22 (t, $^3\text{J}(\text{H,H})=6.2$ Hz, 2H, CH_2), 3.00-2.90 (m, 4H, $2\times\text{CH}_2$), 2.74 (t, $^3\text{J}(\text{H,H})=5.8$ Hz, 2H, CH_2), 1.15-1.05 ppm (m, 12H); $^{13}\text{C-NMR}$ (101 MHz, $[\text{D}_6]\text{DMSO}$): δ =157.9, 151.9, 150.6, 144.9, 142.9, 140.9, 135.8, 130.0, 129.6, 129.2, 127.7, 127.6, 126.5, 122.5, 122.4, 121.0, 119.0, 113.0, 85.4, 63.9, 58.2, 58.0, 54.9, 42.5, 42.4, 24.4, 24.3, 24.3, 24.2, 19.9, 19.8 ppm; $^{31}\text{P-NMR}$ (162 MHz, $[\text{D}_6]\text{DMSO}$): δ =146.54 ppm. HRMS (MALDI): m/z calcd for $\text{C}_{46}\text{H}_{54}\text{N}_4\text{O}_5\text{P}$ [$M + \text{Na}$] $^+$ 773.38263, found 773.38246 ($\Delta m=0.00017$, error 0.2 ppm).

Oligonucleotide synthesis and purification

Solid phase synthesis of the oligonucleotides was accomplished by an ABI 392 DNA/RNA synthesizer using the standard phosphoramidite chemistry. A 0.25 M solution of 5-(ethylthio)-1-*H*-tetrazole in dry acetonitrile was used as activator. The modified oligonucleotides were synthesized in DMTr-on mode. Deprotection was performed incubating the support with NH_3 for 15 hours at room temperature. Purification of the crude product was performed by reversed-phase HPLC on a Nucleosil phenomex Jupiter 4 u Proteo 90 A (250 x 15 mm 4 micron) with gradient A (eluent A: 0.1 M triethylammonium acetat buffer pH 7, eluent B: acetonitrile, 100% A for 1 min, 100% to 45% A in 34 min, flow rate: 4 $\text{mL}\cdot\text{min}^{-1}$) or on a Nucleosil 100-5 C18 (250 x 4.6 mm) with gradient B (eluent A: 0.1 M triethylammonium acetate buffer pH 7, eluent B: acetonitrile, 95% A for 2 min, 95% to 60% A in 33 min, flow rate: 1 $\text{mL}\cdot\text{min}^{-1}$). The solvent was evaporated and the dimethoxytrityl group was cleaved off over 30 min with 80% aqueous acetic acid. After removing the acid the resulting oligonucleotide was purified again by RP-HPLC using gradient A or B. Purity and identity were confirmed by HPLC-ESI-MS (Bruker microTOF-Q).

Characterization of oligonucleotides

Table S1. Mass analysis of synthesized oligonucleotides

sequence	calculated	found
GG- Az1 -GG	1559.1	1558.6
GG- Az1 -G	1229.9	1229.5
GGG- Az1 -GG	1888.3	1887.7
G- Az1 -GG	1229.9	1229.5
GG- Az1 -GGG	1888.3	1887.7
GG- Az2 -GG	1559.1	1558.6
GG- Az3 -GG	1587.2	1586.6

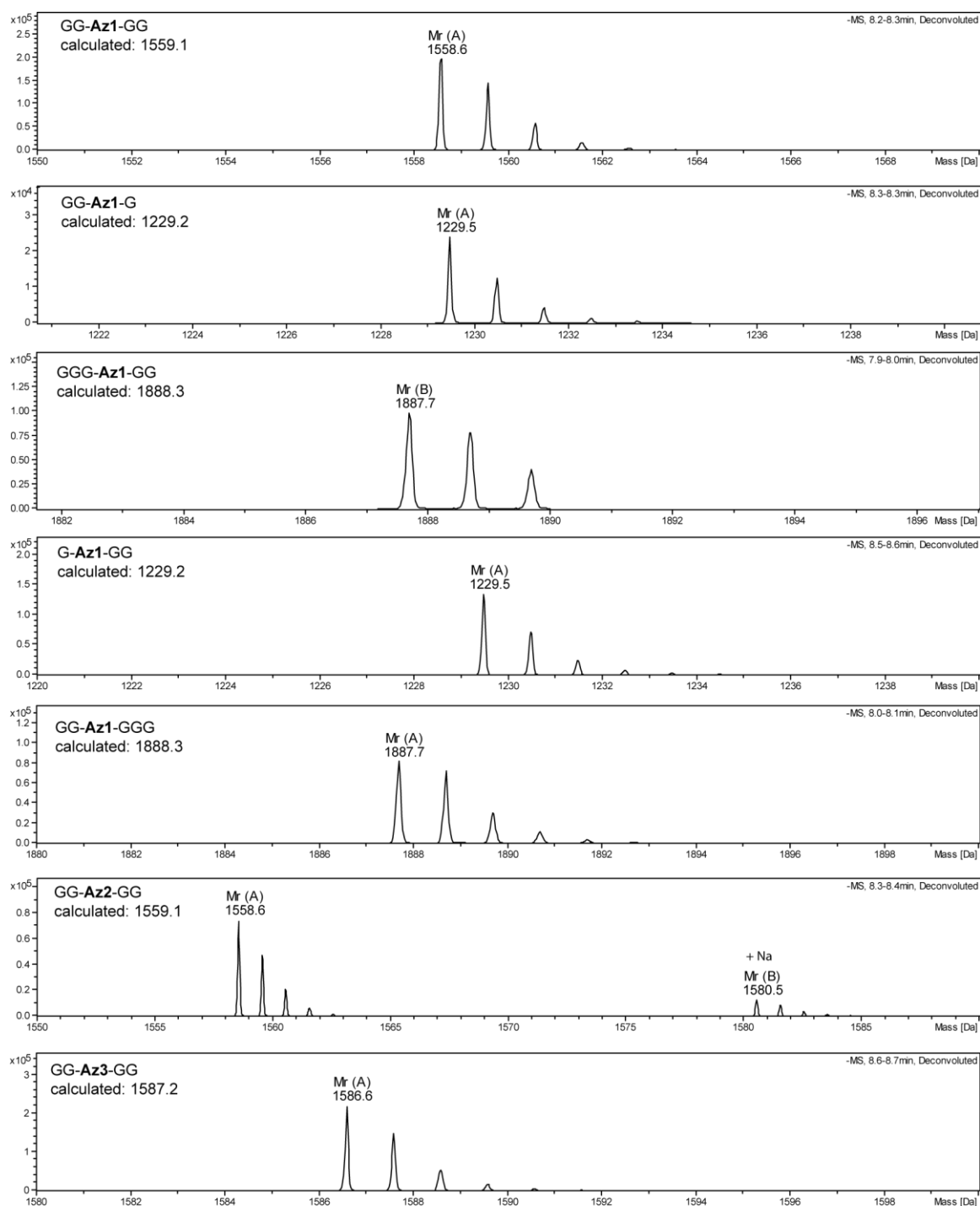


Figure S2. Deconvoluted mass spectra of synthesized azobenzene-containing oligonucleotides

Photoisomerization

UV irradiation ($\lambda_{\max} = 365 \text{ nm}$) of the sample was performed in a custom-built irradiation setup containing three LEDs with 100 mW power each (Nichia NCCU033(T)).

For vis light irradiation a light source from Müller Elektronik Optik was used with a xenon arc lamp.

CD measurements

CD spectra were recorded on JASCO J-710 spectropolarimeter at 25°C (JASCO peltier control) over the wavelength range 210-320 nm in a 1 cm path length cuvette. Each trace is the average of 3 scans at a scanning speed of 100 nm·min⁻¹, with a 1 s time constant, 1.0 nm data pitch and 1.0 nm bandwidth. All CD spectra were baseline-corrected for signal contributions of the buffer.

Sample concentrations of 50 μM GG-**Az1**-GG, GG-**Az2**-GG or GG-**Az3**-GG in 50 mM Tris-HCl (pH 7.4) were used for CD experiments. Titration experiments with K⁺ and Na⁺ were performed such that finally the added amount never exceeded 6% of the total initial volume. After UV irradiation of the sample CD spectra were recorded until the sample was thermally equilibrated (red data points).

CD annealing and melting curves of GG-**Az1**-GG, GG-**Az2**-GG and GG-**Az3**-GG in 10 mM lithium cacodylate buffer (pH 7.4) were monitored at 295 nm with a temperature gradient of 0.2°C·min⁻¹.

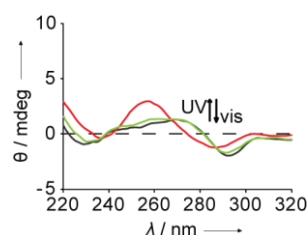


Figure S3. CD spectra of GG-**Az1**-GG with photoirradiation (5 min UV, 2 min vis) without K⁺.

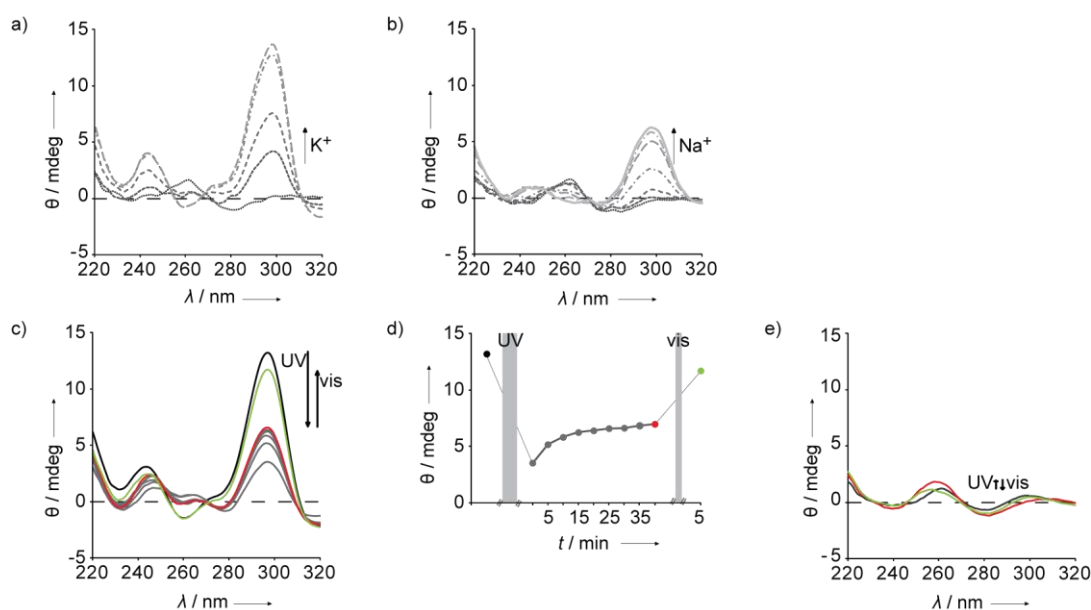


Figure S4. CD spectra of GG-**Az2**-GG. a) Series of CD spectra with increasing concentration (0, 5, 25, 50, 100 mM) of K⁺. b) Series of CD spectra with increasing concentration (0, 50, 100, 200, 300, 400, 500 mM) of Na⁺. c) CD spectra of GG-**Az2**-GG with photoirradiation (30 min UV, 2 min vis) with K⁺. d) Time course of the signal at 295 nm of GG-**Az2**-GG with K⁺ under the indicated treatment. e) CD spectra of GG-**Az2**-GG photoirradiation (5 min UV, 2 min vis) without K⁺.

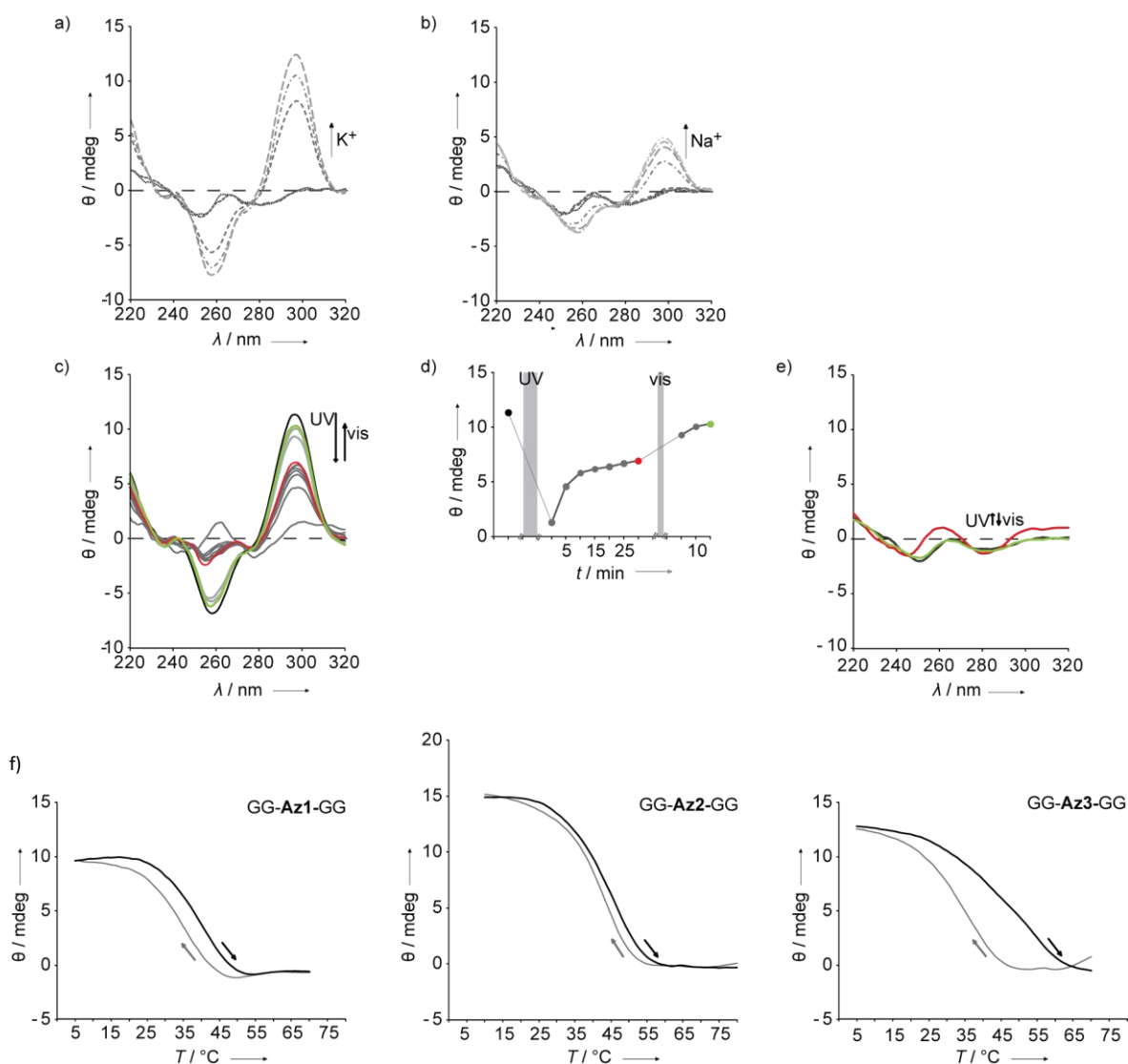


Figure S5. CD spectra of GG-Az3-GG. a) Series of CD spectra with increasing concentration (0, 5, 25, 50, 100 mM) of K⁺. b) Series of CD spectra with increasing concentration (0, 50, 100, 200, 300, 400, 500 mM) of Na⁺. c) CD spectra of GG-Az3-GG with photoirradiation (10 min UV, 2 min vis) with K⁺. d) Time course of the signal at 295 nm of GG-Az3-GG with K⁺ under the indicated treatment. e) CD spectra of GG-Az3-GG photoirradiation (5 min UV, 2 min vis) without K⁺. f) CD annealing and melting profiles for GG-Az1-GG, GG-Az2-GG and GG-Az3-GG in (E)-conformation.

UV/vis measurements

UV/vis spectra were recorded on Specord S600 (analytikjena) spectrometer equipped with an external thermostat (Haake K20) at 25°C in a 1 cm path length cuvette. All CD spectra were baseline-corrected for signal contributions of the buffer.

After UV irradiation, experiments of GG-Az1-GG, GG-Az2-GG and GG-Az3-GG GG, 50 μm in 50 mM Tris-HCl (pH 7.4), were conducted in the same manner (sample concentration, irradiation time and time intervals) as for CD experiments. The UV/vis difference spectra of the photoswitched sample were obtained by subtracting the UV/vis spectrum of the folded sample in K⁺ (light grey) from the UV/vis spectrum of the UV irradiated sample after thermal equilibration (red).

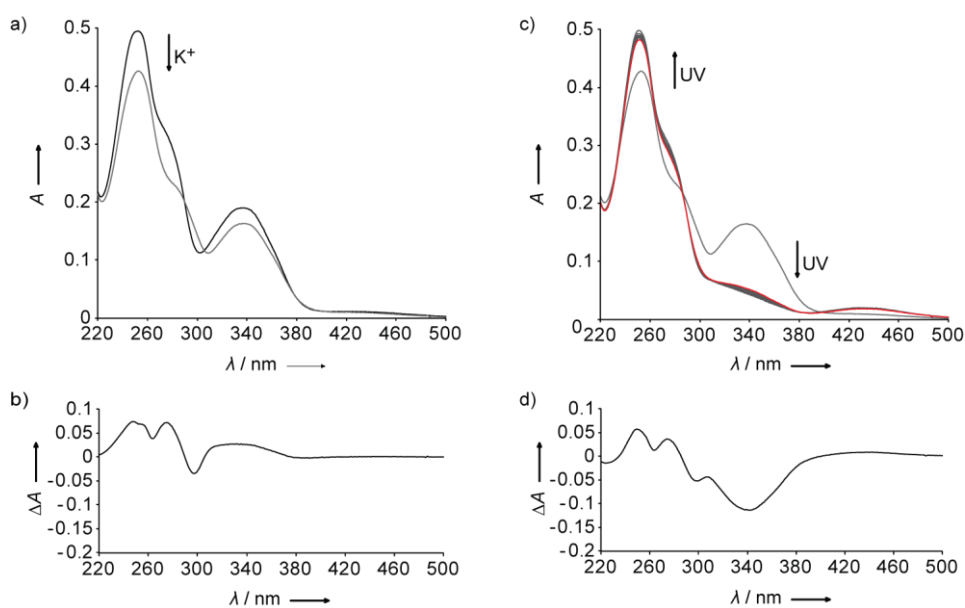


Figure S6. UV/vis spectra of GG-Az1-GG. a) UV/vis spectra of GG-Az1-GG before and after addition of K^+ to an end concentration of 100 mM and b) corresponding UV/vis difference spectrum. c) UV/vis spectra of GG-Az1-GG with K^+ and after UV irradiation and d) corresponding UV/vis difference spectrum.

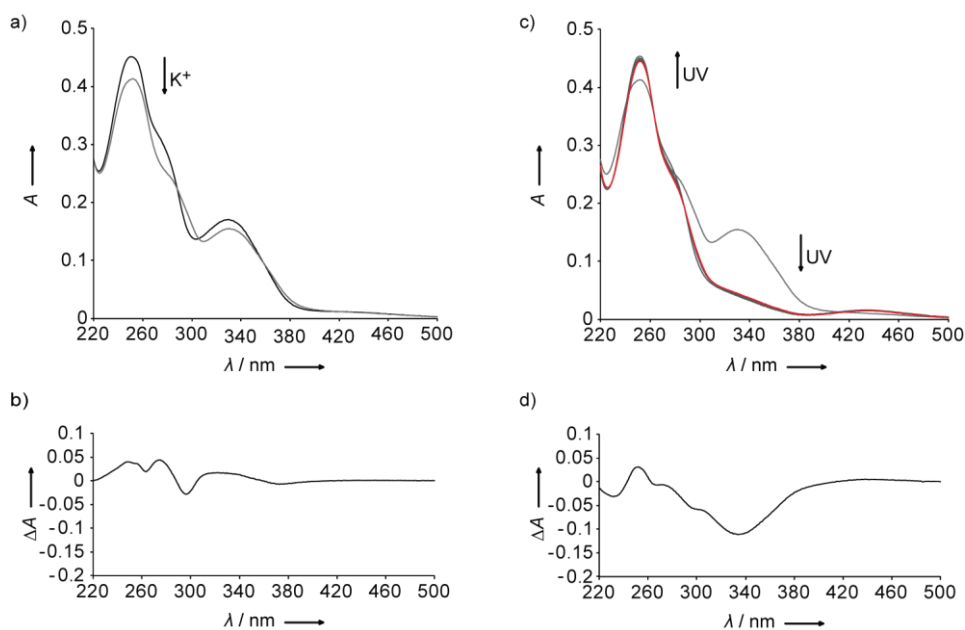


Figure S7. UV/vis spectra of GG-Az2-GG. a) UV/vis spectra of GG-Az2-GG before and after addition of K^+ to an end concentration of 100 mM and b) corresponding UV/vis difference spectrum. c) UV/vis spectra of GG-Az2-GG with K^+ and after UV irradiation and d) corresponding UV/vis difference spectrum.

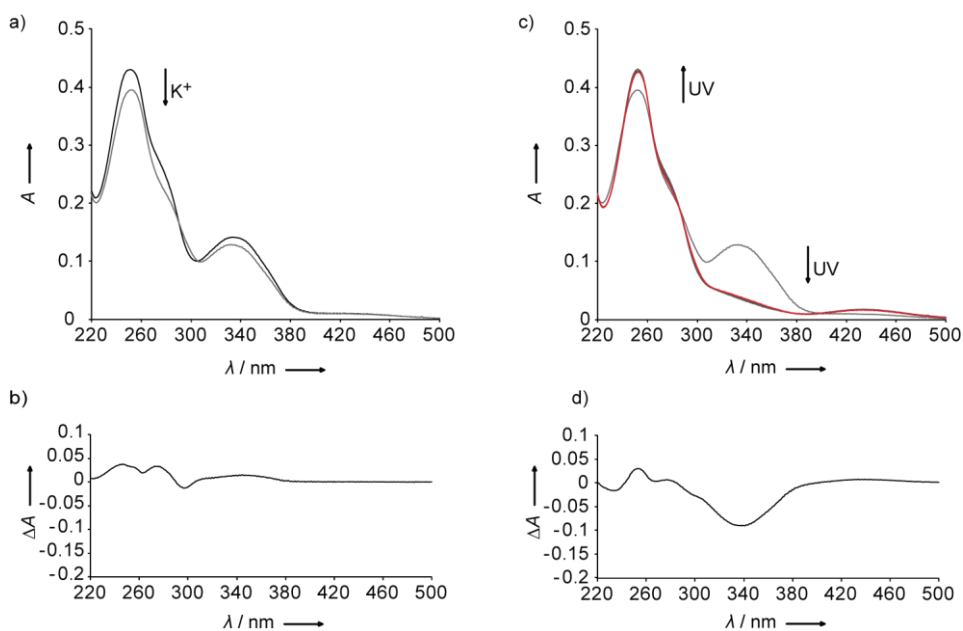


Figure S8. UV/vis spectra of GG-**Az3**-GG. a) UV/vis spectra of GG-**Az3**-GG before and after addition of K^+ to an end concentration of 100 mM and b) corresponding UV/vis difference spectrum. c) UV/vis spectra of GG-**Az3**-GG with K^+ and after UV irradiation and d) corresponding UV/vis difference spectrum.

Concentrations of GG-**Az1**-GG, GG-**Az2**-GG and GG-**Az3**-GG were determined using extinction coefficients $\epsilon = 41800 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ (IDT OligoAnalyzer 3.1) for GGGG and $\epsilon = 4100 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ for the contribution of the azobenzene at 260 nm.

NMR-spectra of synthesized compounds

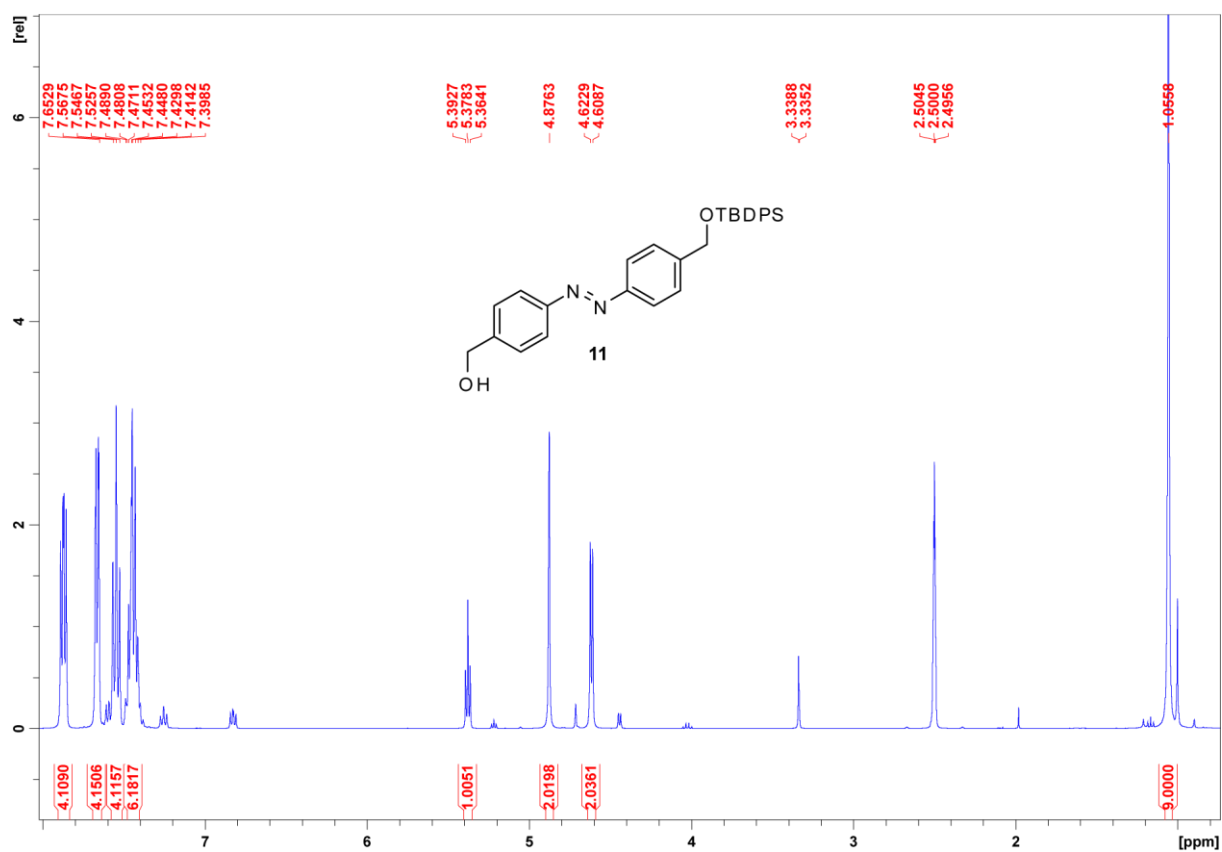


Figure S9. $^1\text{H-NMR}$ spectrum of **11** in $[\text{D}_6]\text{DMSO}$.

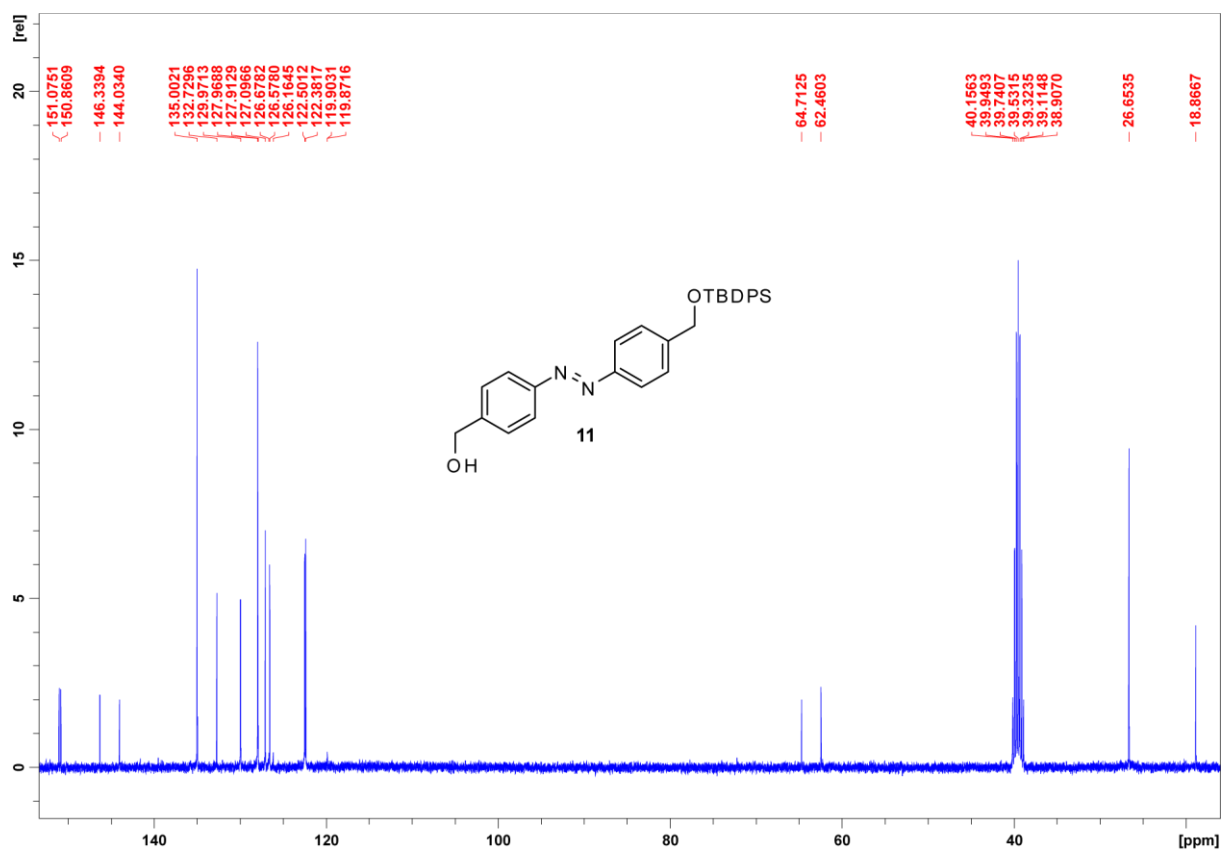


Figure S10. $^{13}\text{C-NMR}$ spectrum of **11** in $[\text{D}_6]\text{DMSO}$.

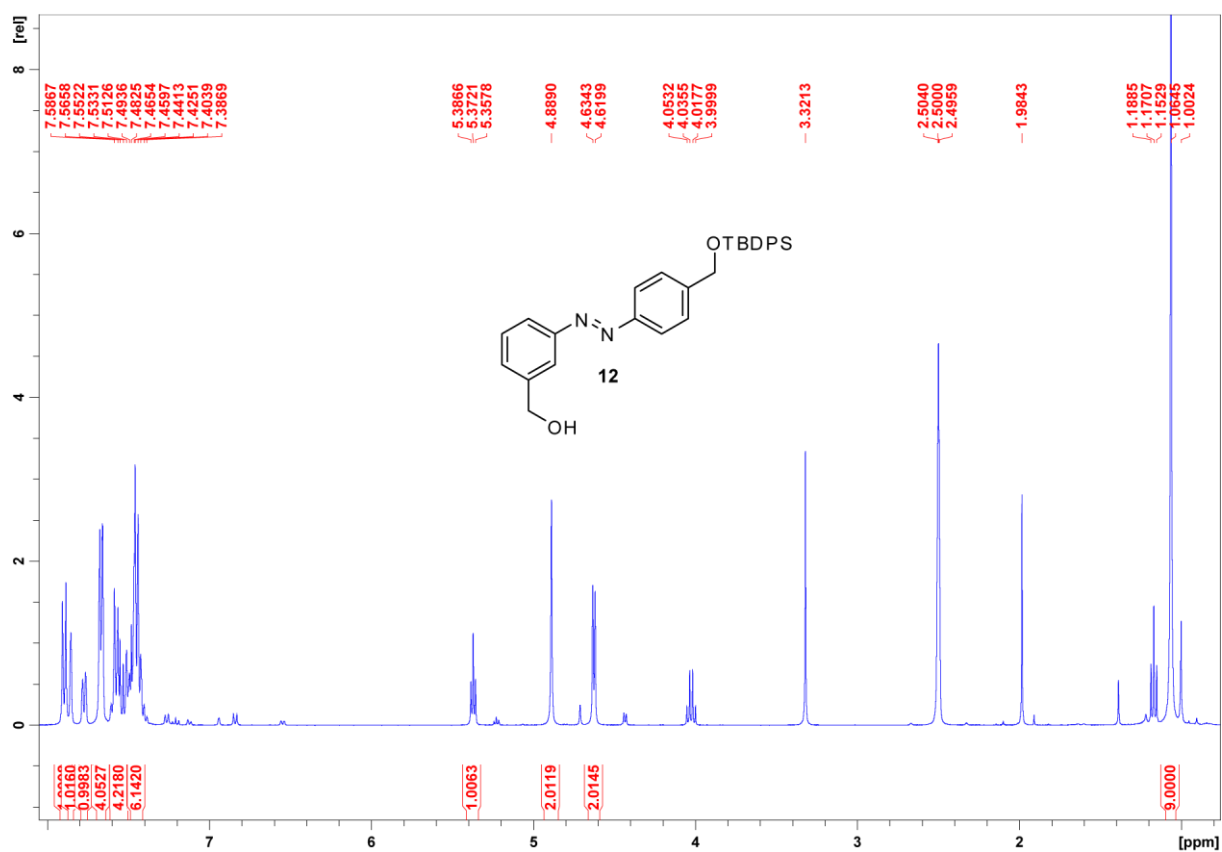


Figure S11. $^1\text{H-NMR}$ spectrum of **12** in $[\text{D}_6]\text{DMSO}$.

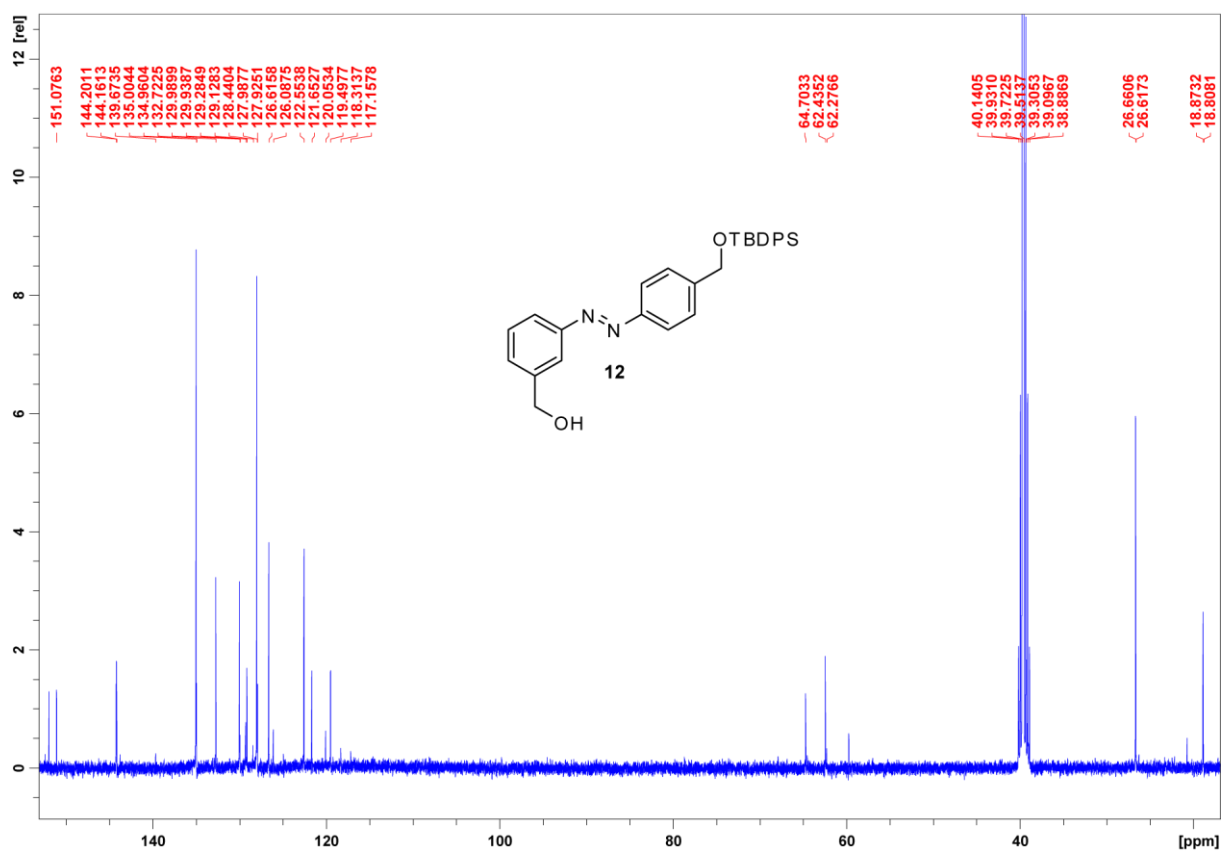


Figure S12. $^{13}\text{C-NMR}$ spectrum of **12** in $[\text{D}_6]\text{DMSO}$.

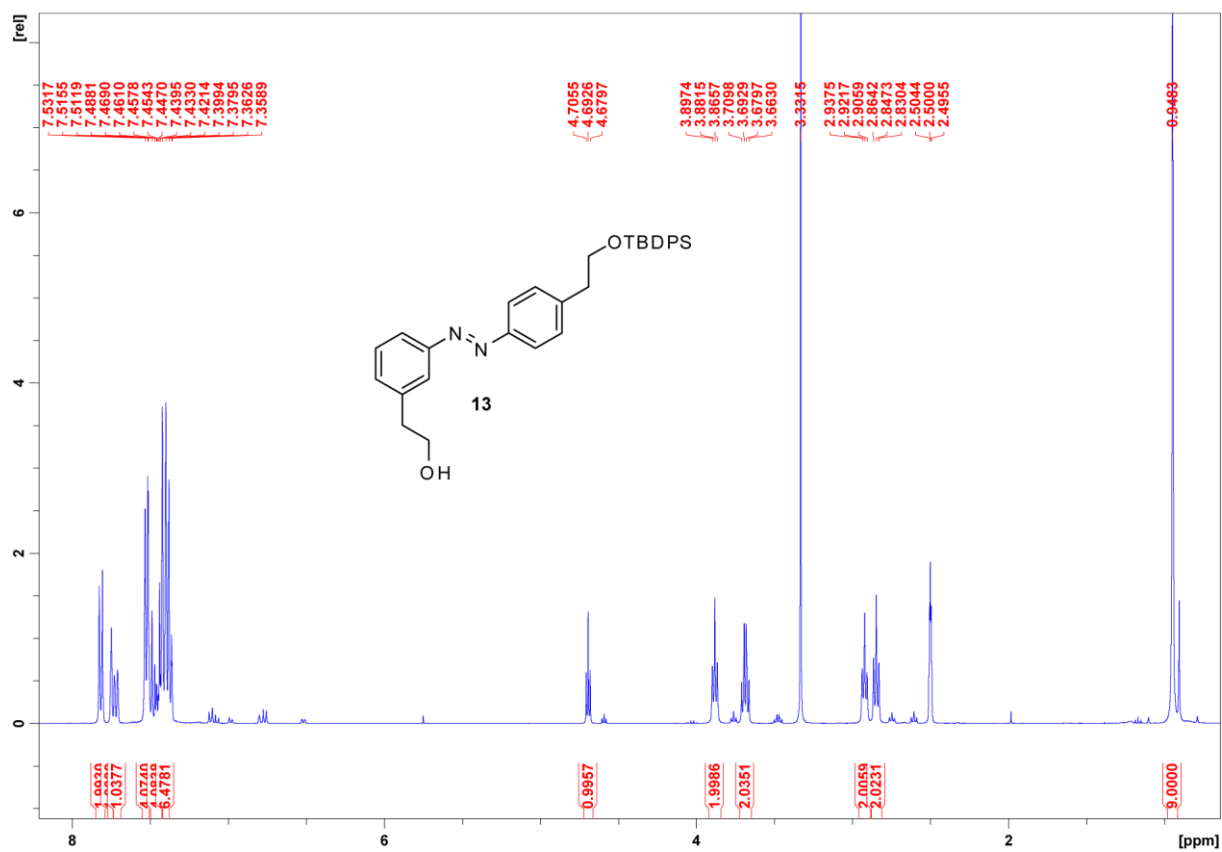


Figure S13. $^1\text{H-NMR}$ spectrum of 13 in $[\text{D}_6]\text{DMSO}$.

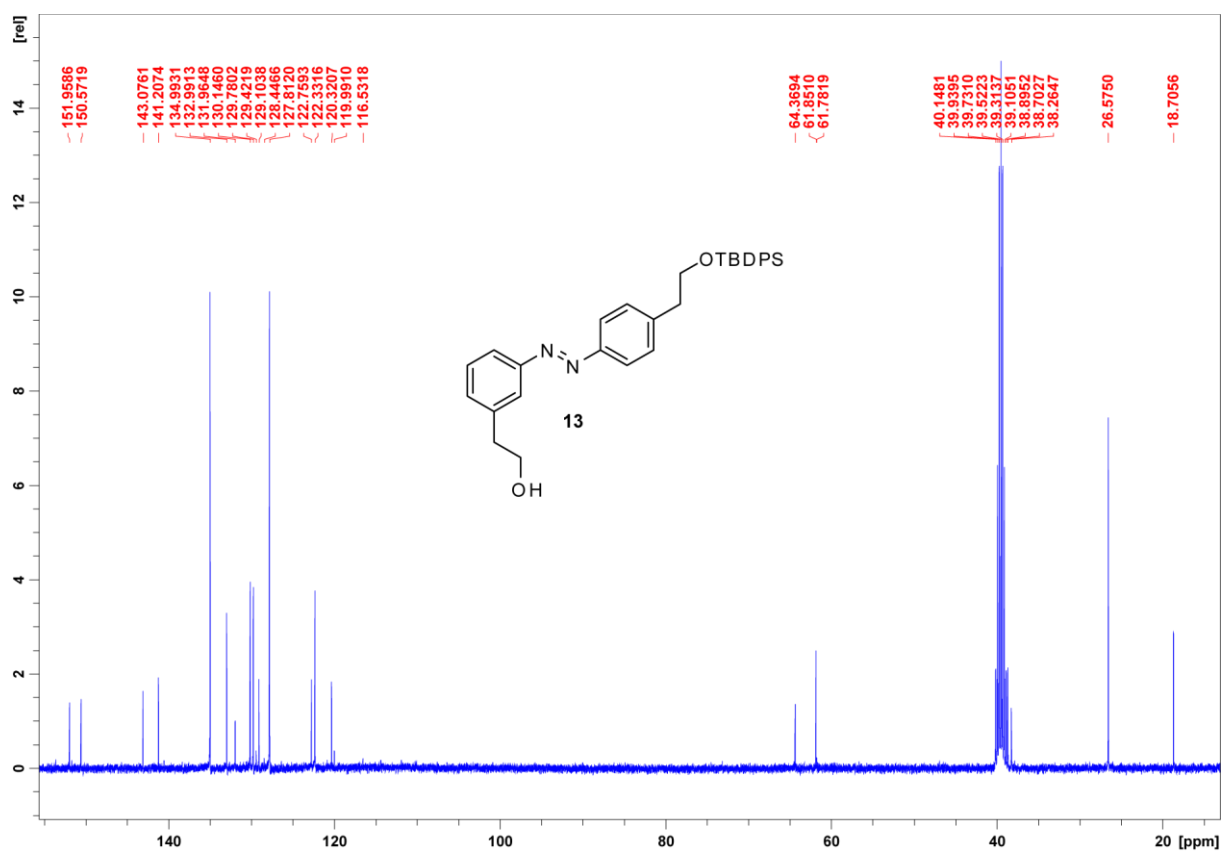


Figure S14. $^{13}\text{C-NMR}$ spectrum of 13 in $[\text{D}_6]\text{DMSO}$.

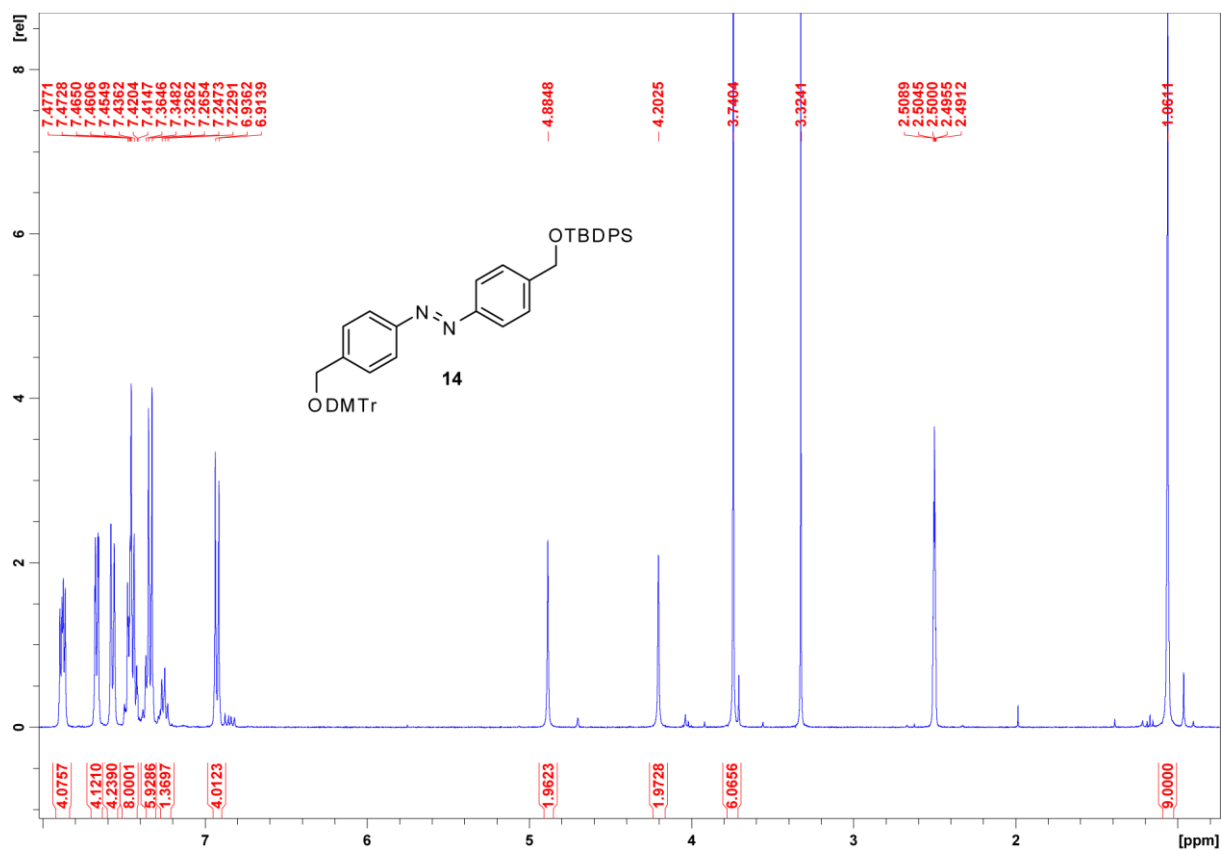


Figure S15. $^1\text{H-NMR}$ spectrum of **14** in $[\text{D}_6]\text{DMSO}$.

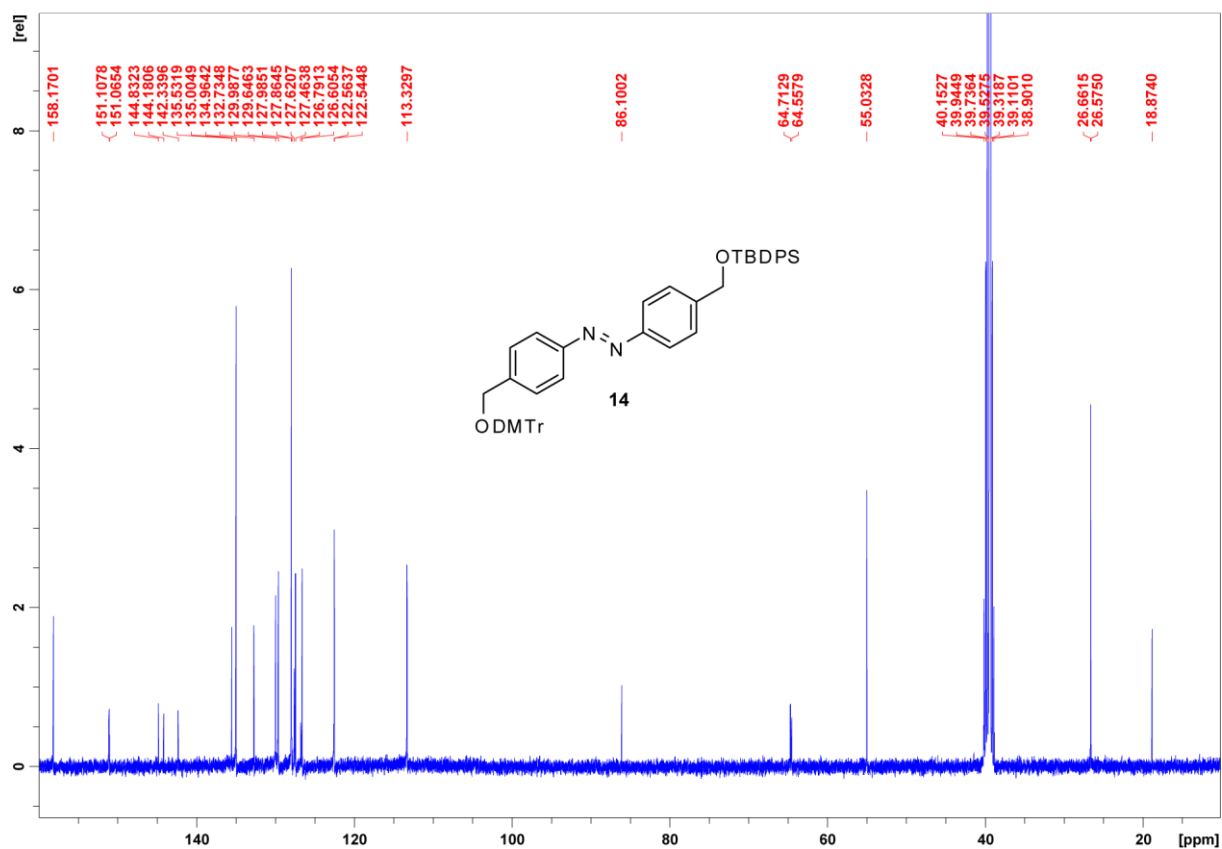


Figure S16. $^{13}\text{C-NMR}$ spectrum of **14** in $[\text{D}_6]\text{DMSO}$.

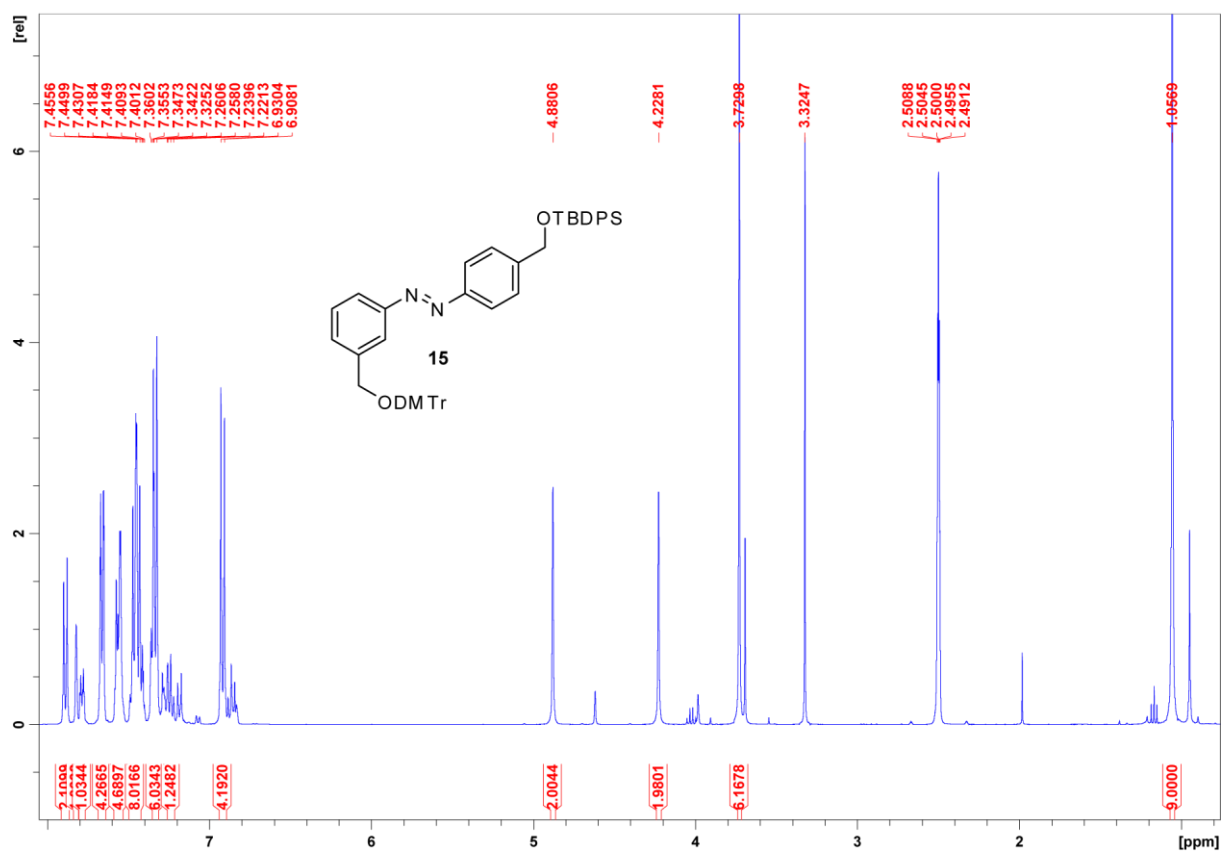


Figure S17. $^1\text{H-NMR}$ spectrum of **15** in $[\text{D}_6]\text{DMSO}$.

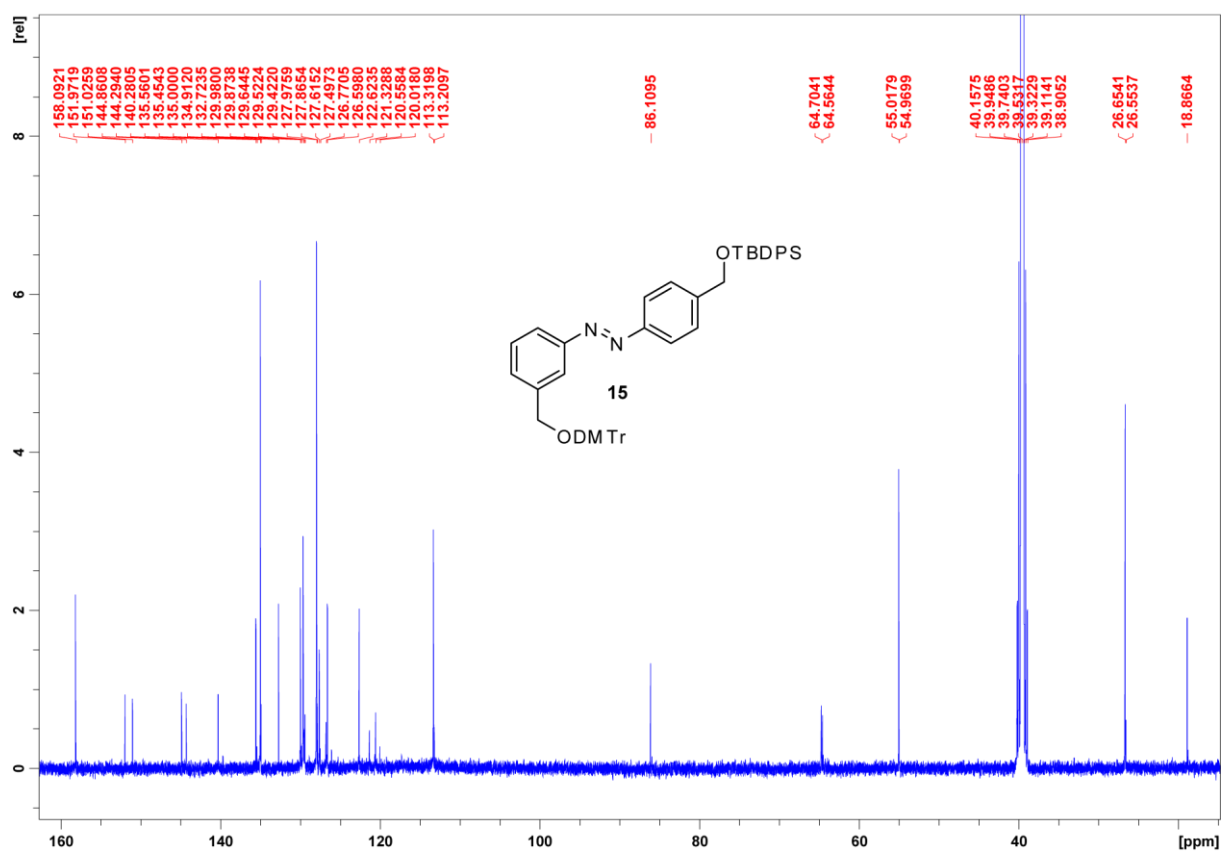


Figure S18. $^{13}\text{C-NMR}$ spectrum of **15** in $[\text{D}_6]\text{DMSO}$.

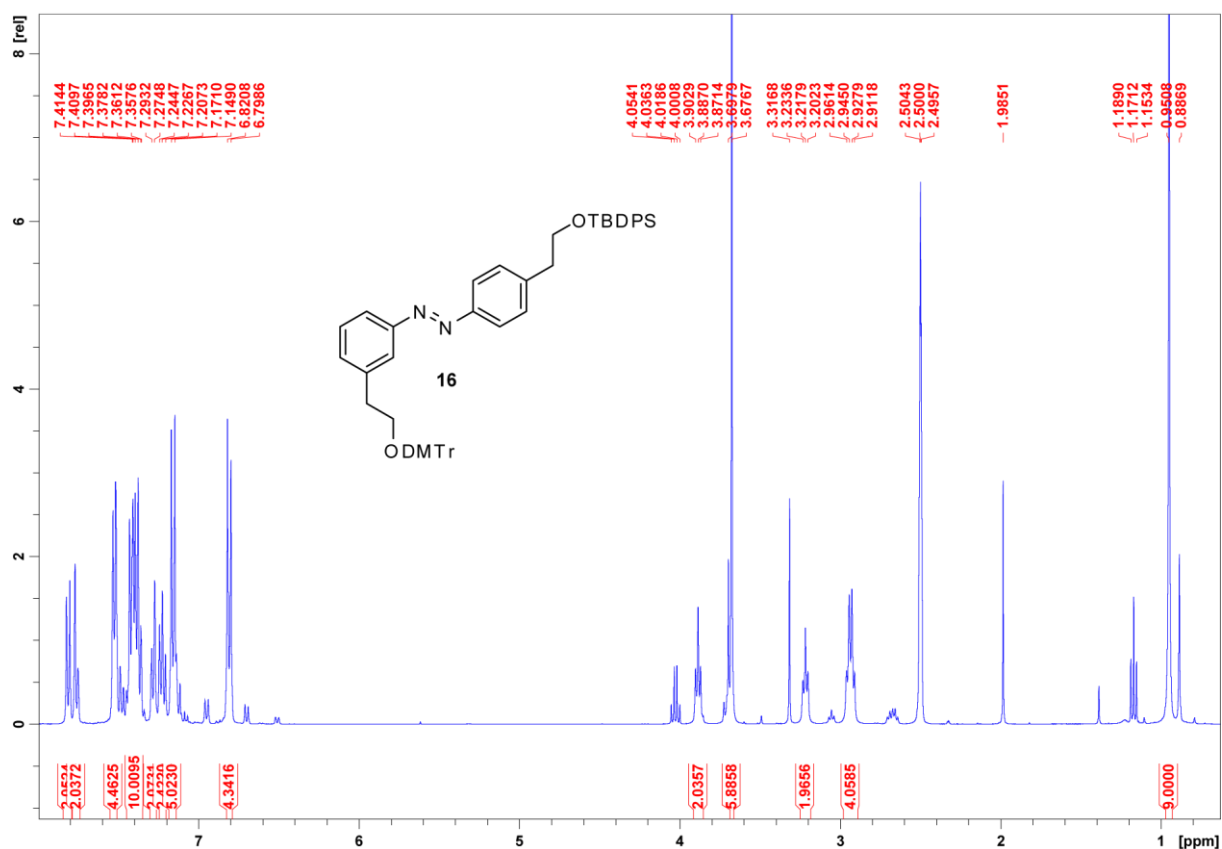


Figure S19. $^1\text{H-NMR}$ spectrum of **16** in $[\text{D}_6]\text{DMSO}$.

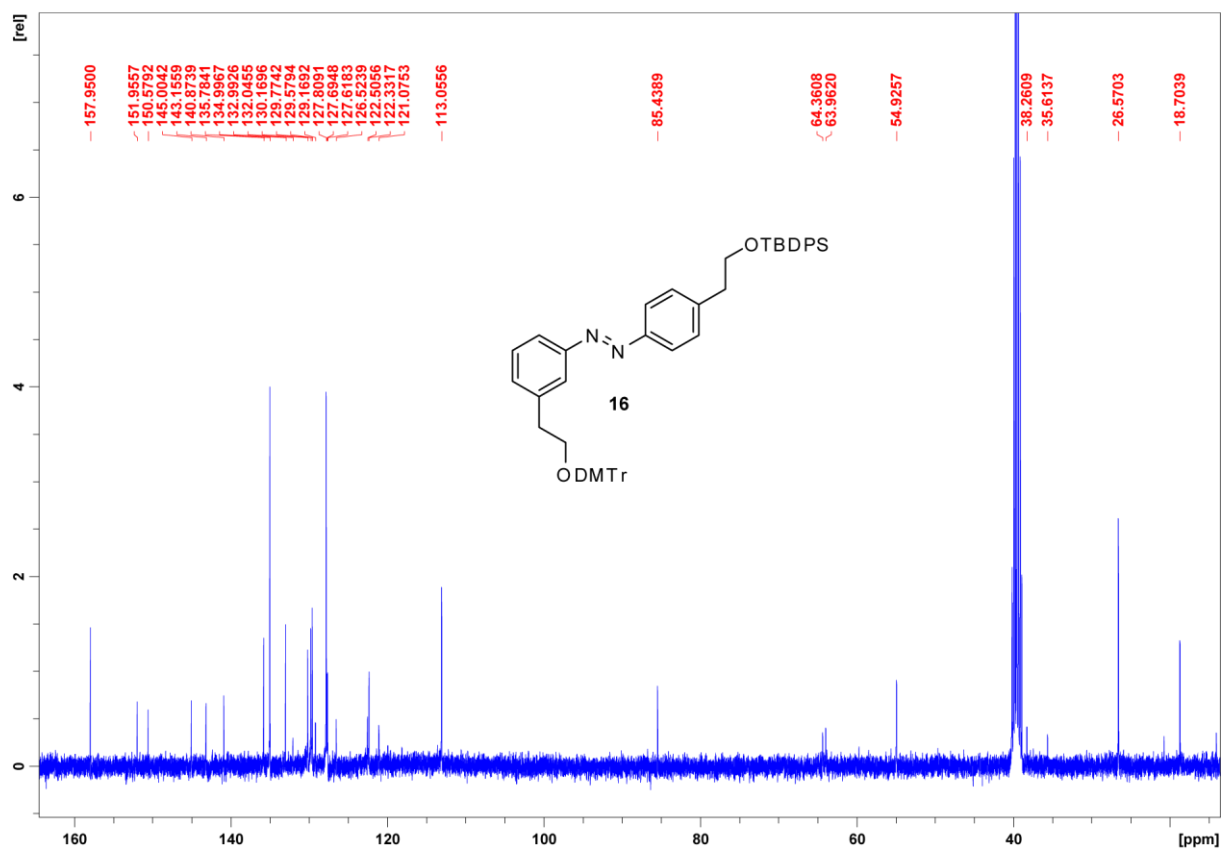


Figure S20. $^{13}\text{C-NMR}$ spectrum of **16** in $[\text{D}_6]\text{DMSO}$.

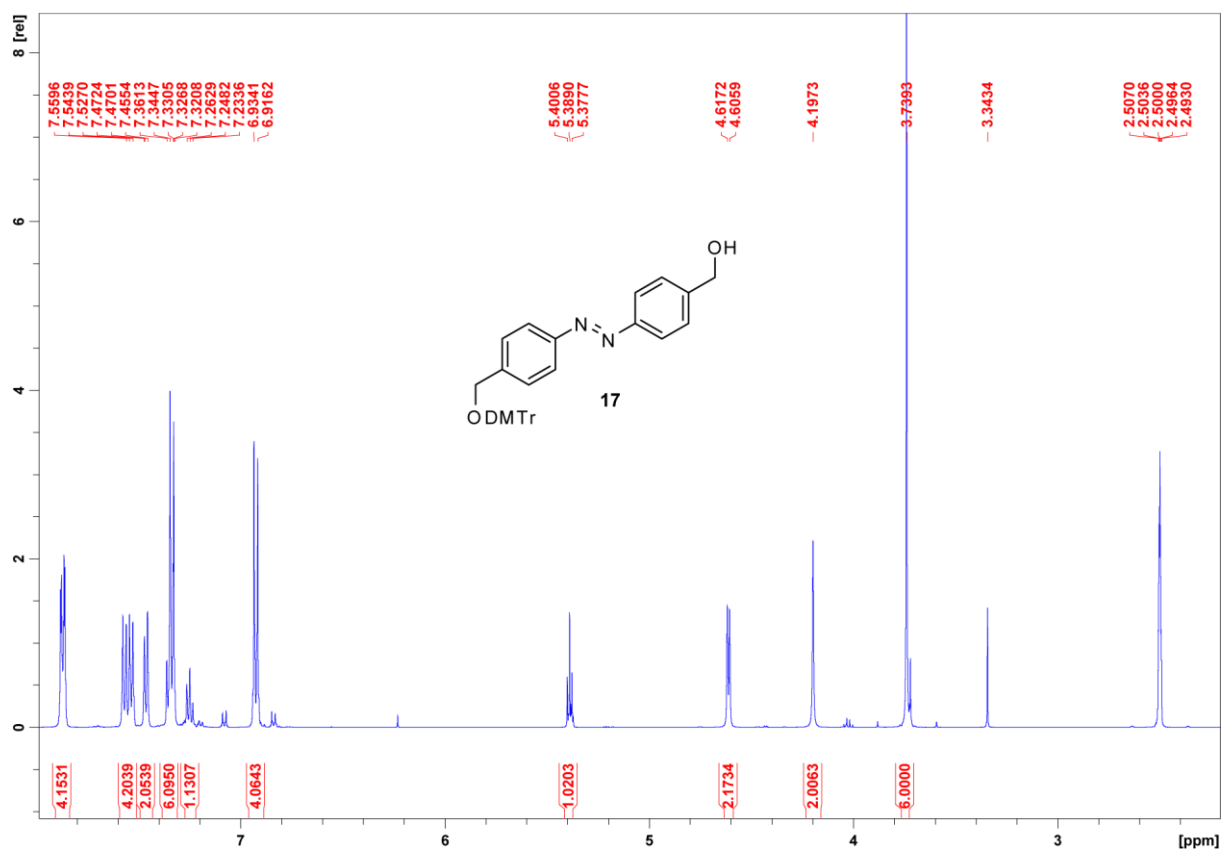


Figure S21. ¹H-NMR spectrum of 17 in [D₆]DMSO.

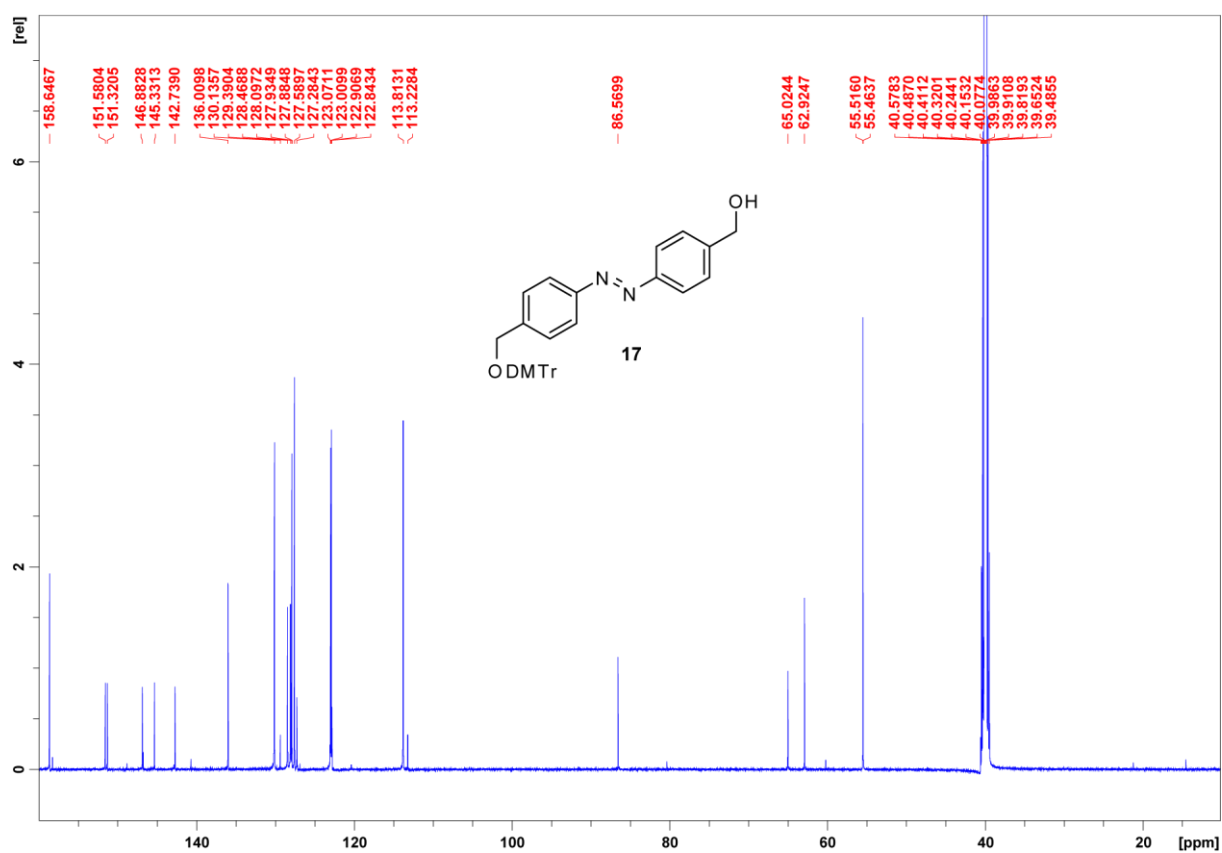


Figure S22. ¹³C-NMR spectrum of 17 in [D₆]DMSO.

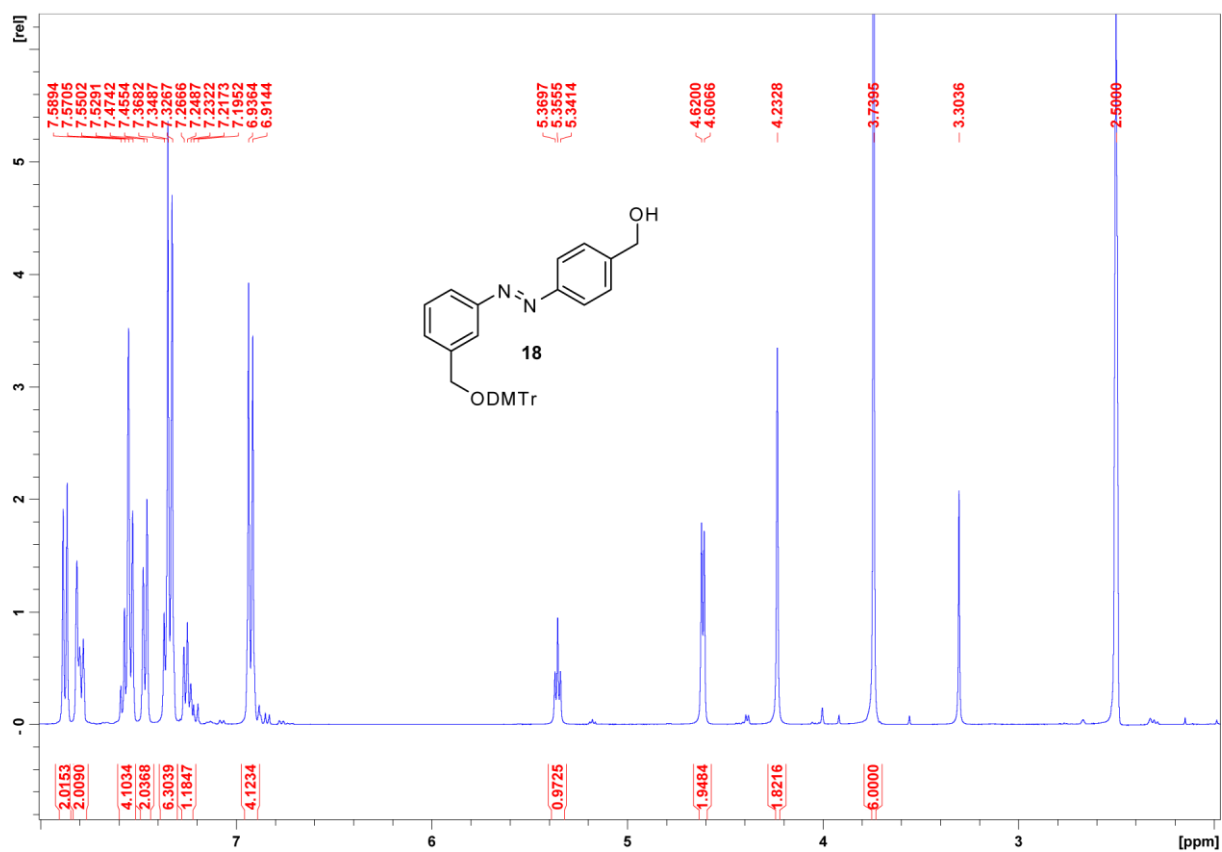


Figure S23. ¹H-NMR spectrum of 18 in [D₆]DMSO.

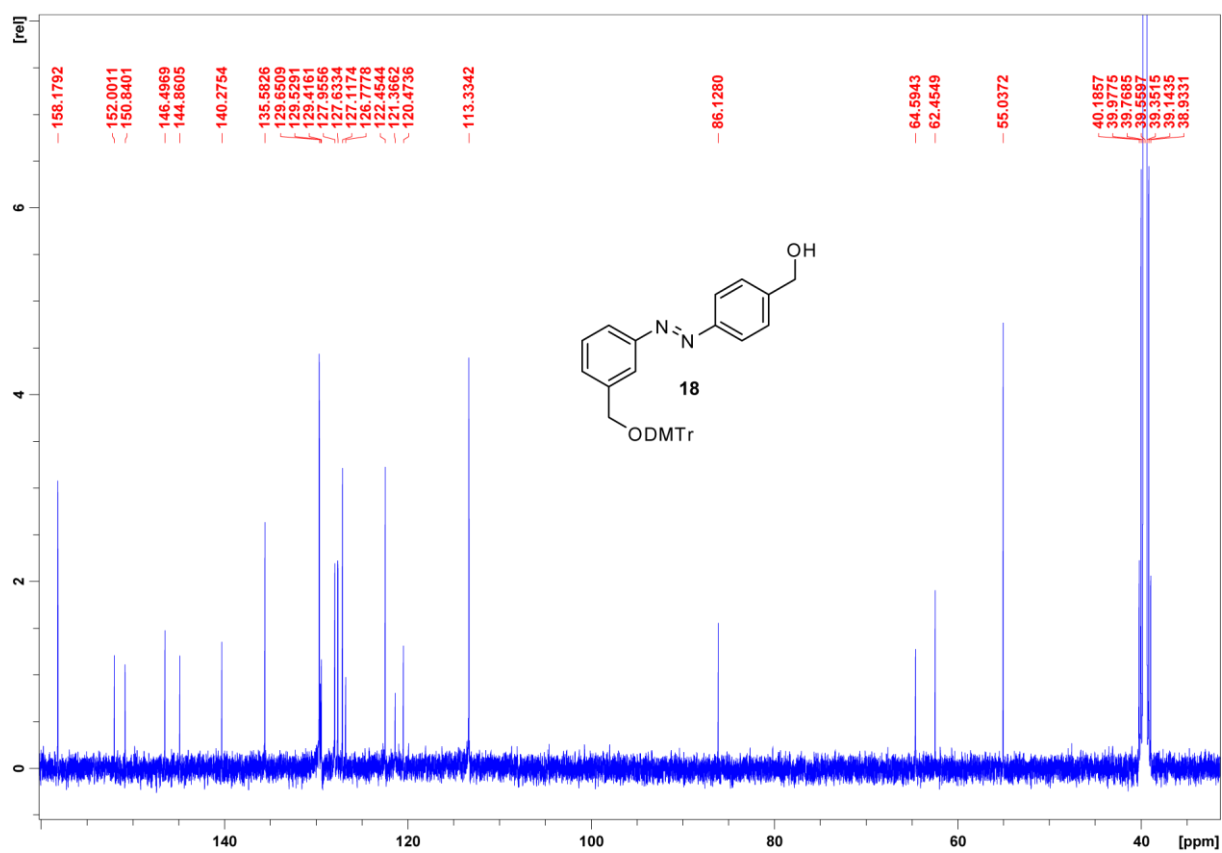


Figure S24. ¹³C-NMR spectrum of 18 in [D₆]DMSO.

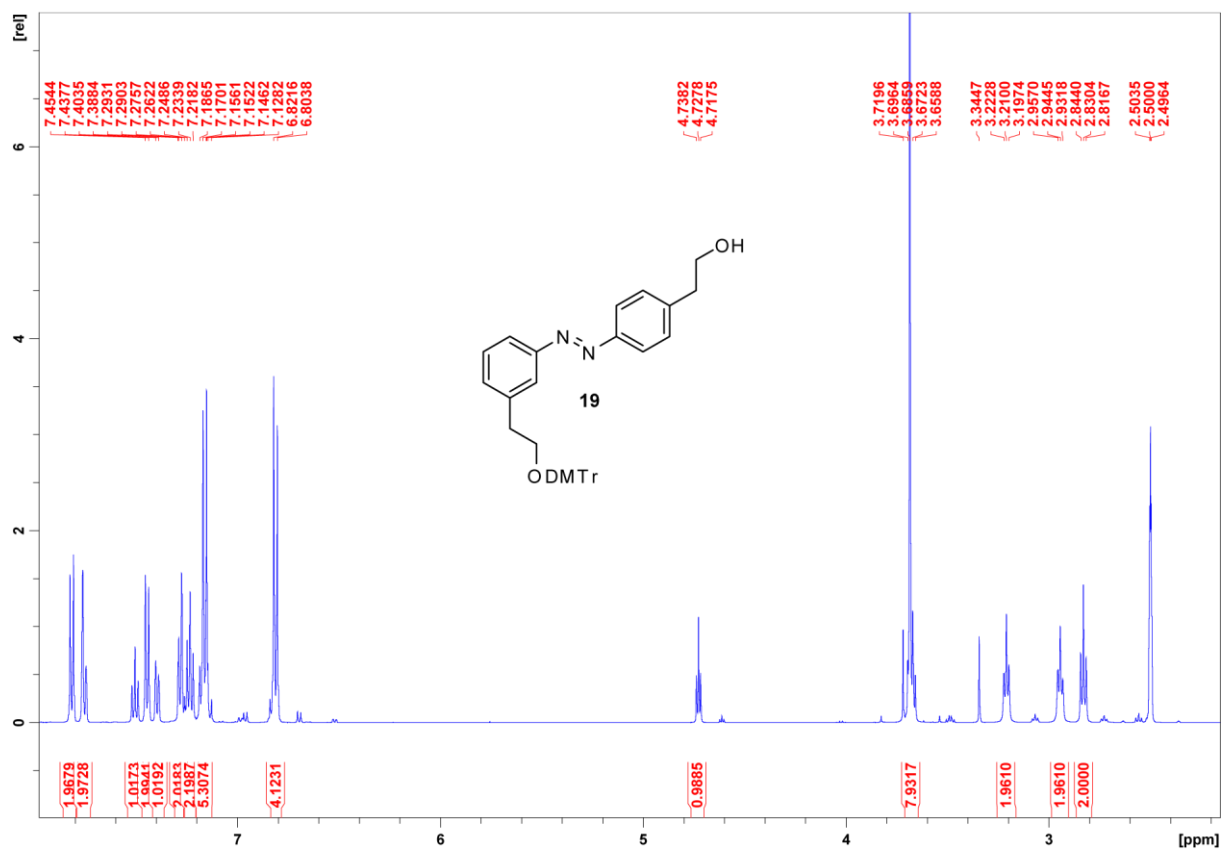


Figure S25. $^1\text{H-NMR}$ spectrum of **19** in $[\text{D}_6]\text{DMSO}$.

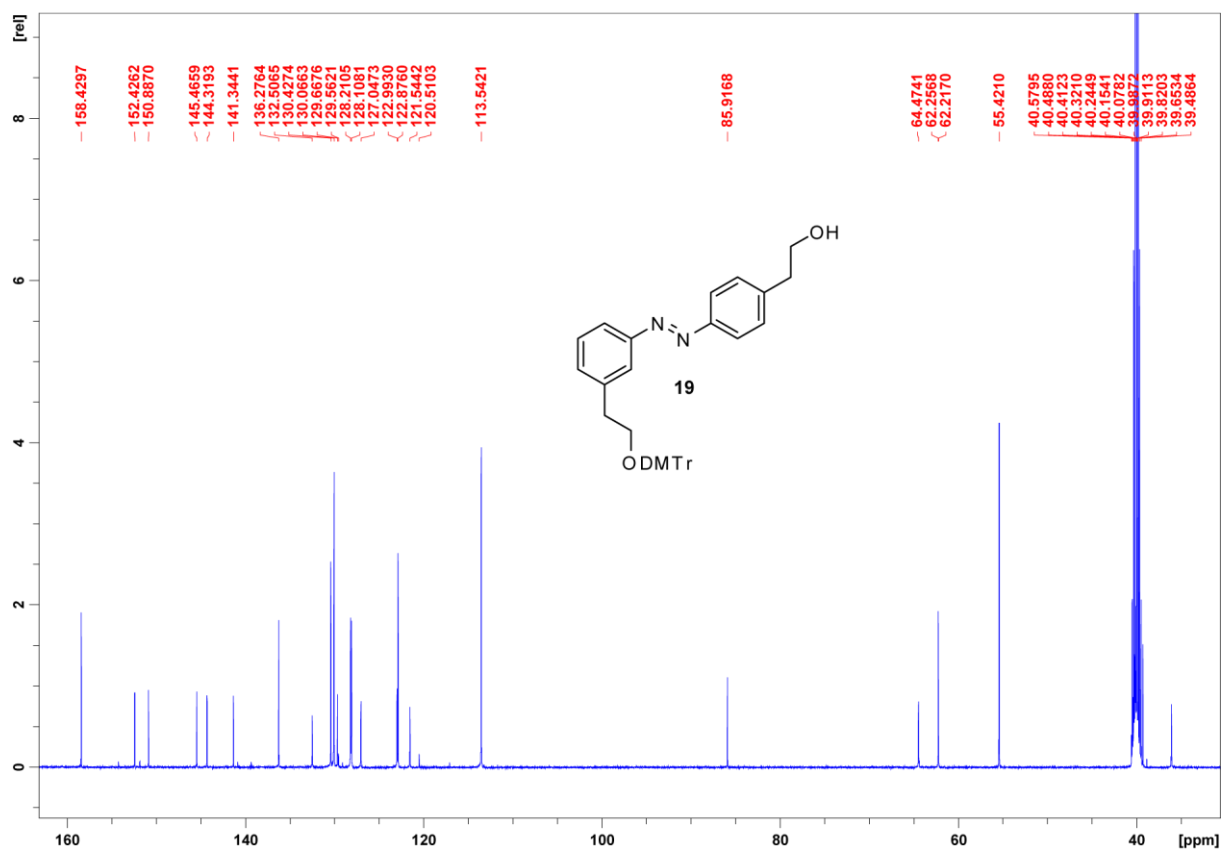


Figure S26. $^{13}\text{C-NMR}$ spectrum of **19** in $[\text{D}_6]\text{DMSO}$.

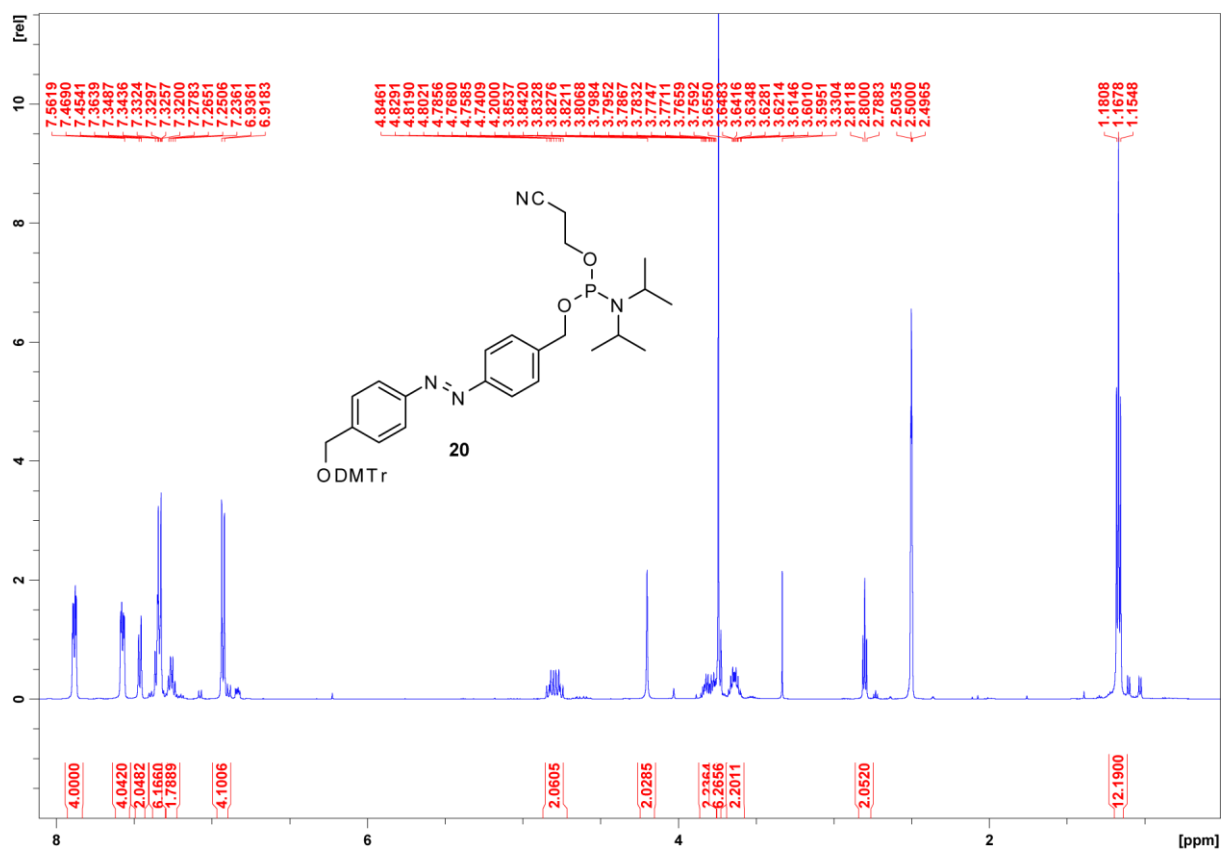


Figure S27. ¹H-NMR spectrum of **20** in [D₆]DMSO.

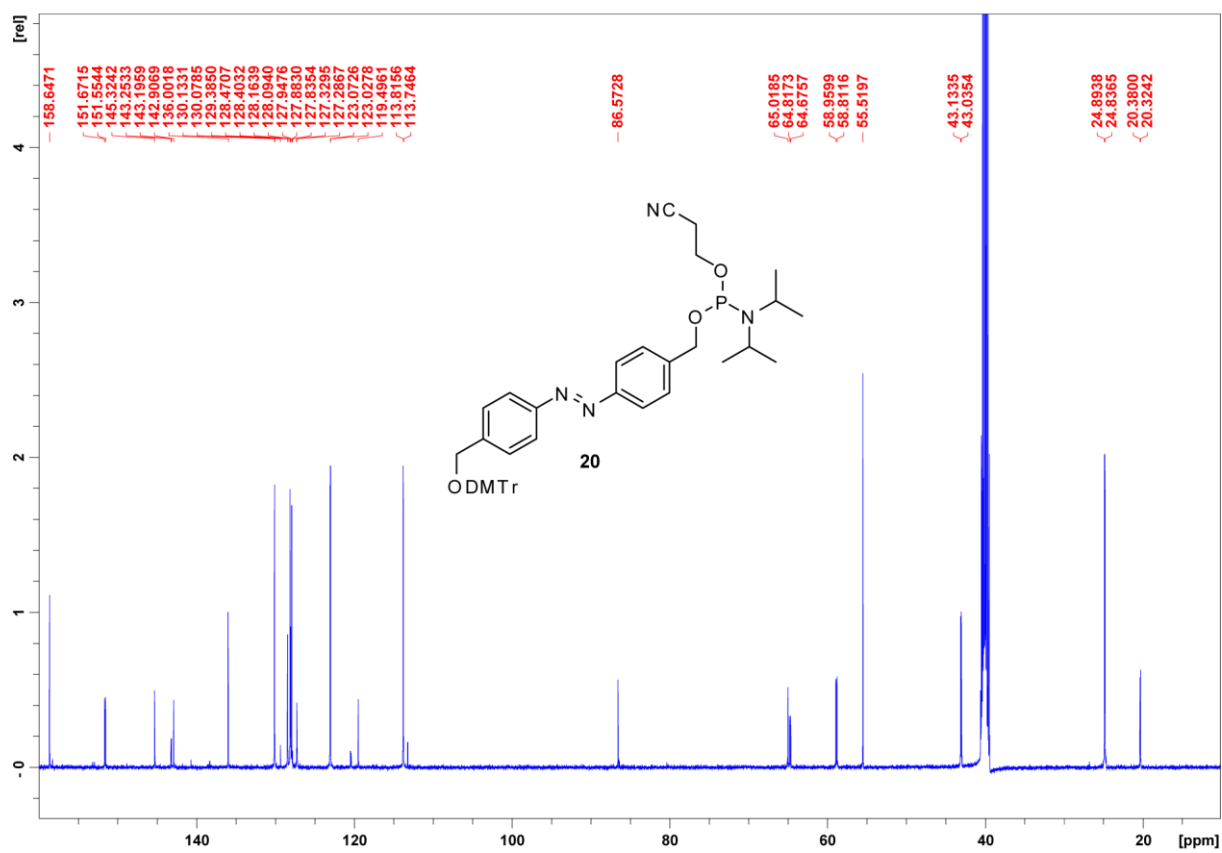


Figure S28. ¹³C-NMR spectrum of **20** in [D₆]DMSO.

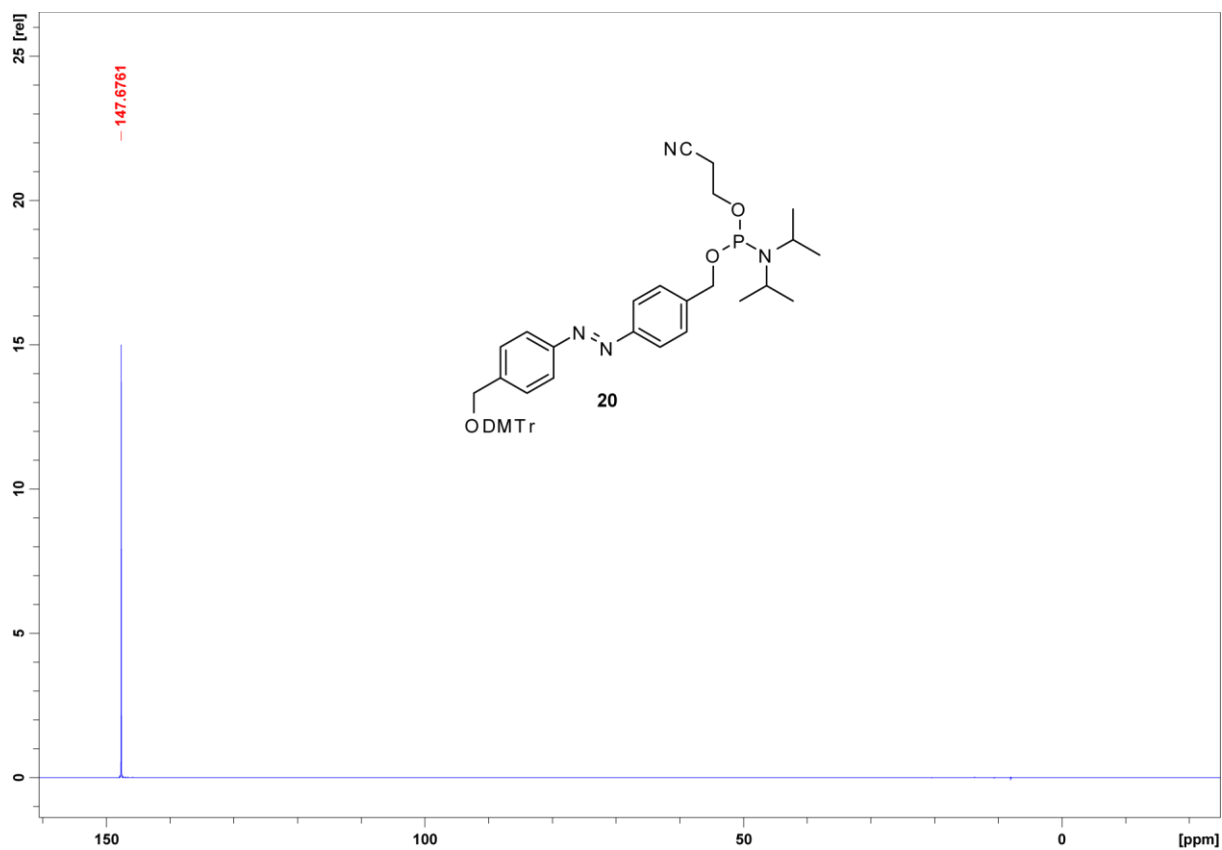


Figure S29. ^{31}P -NMR spectrum of **20** in $[\text{D}_6]\text{DMSO}$.

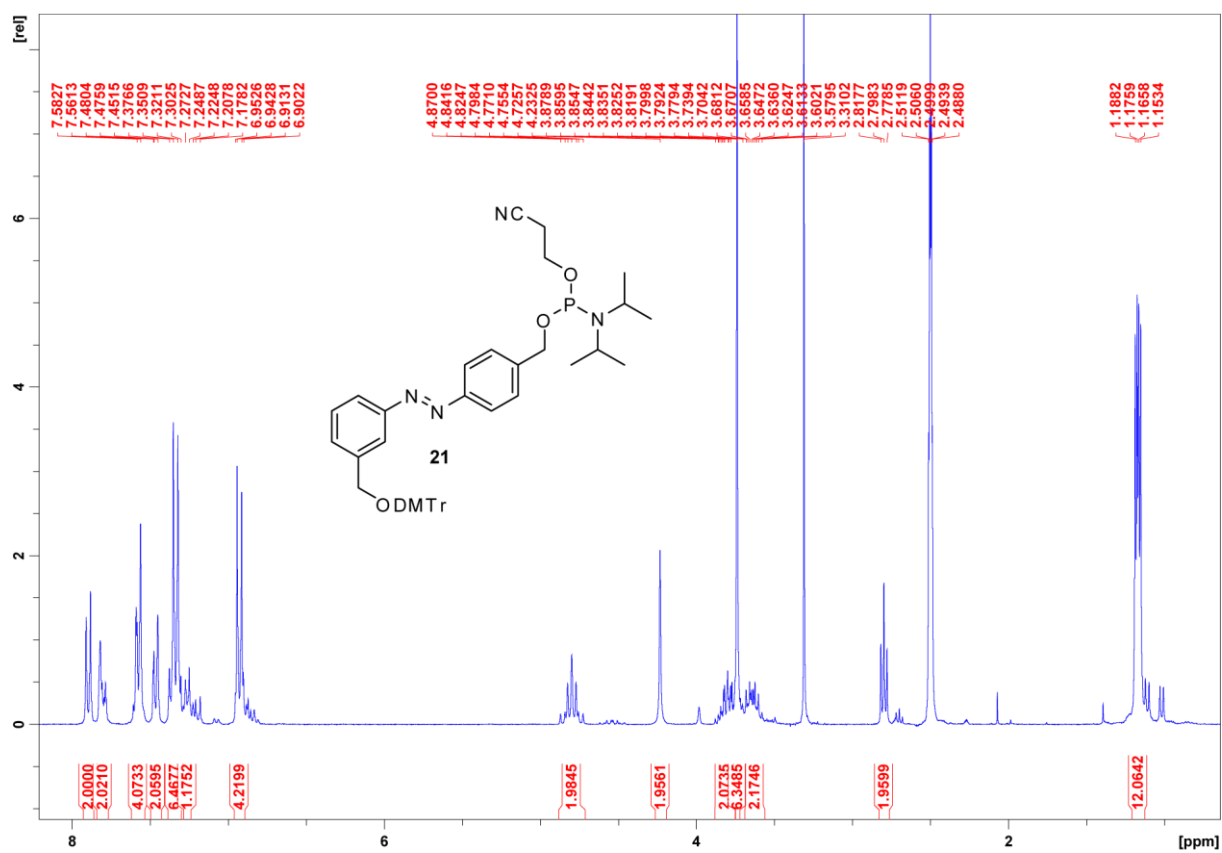


Figure S30. ^1H -NMR spectrum of **21** in $[\text{D}_6]\text{DMSO}$.

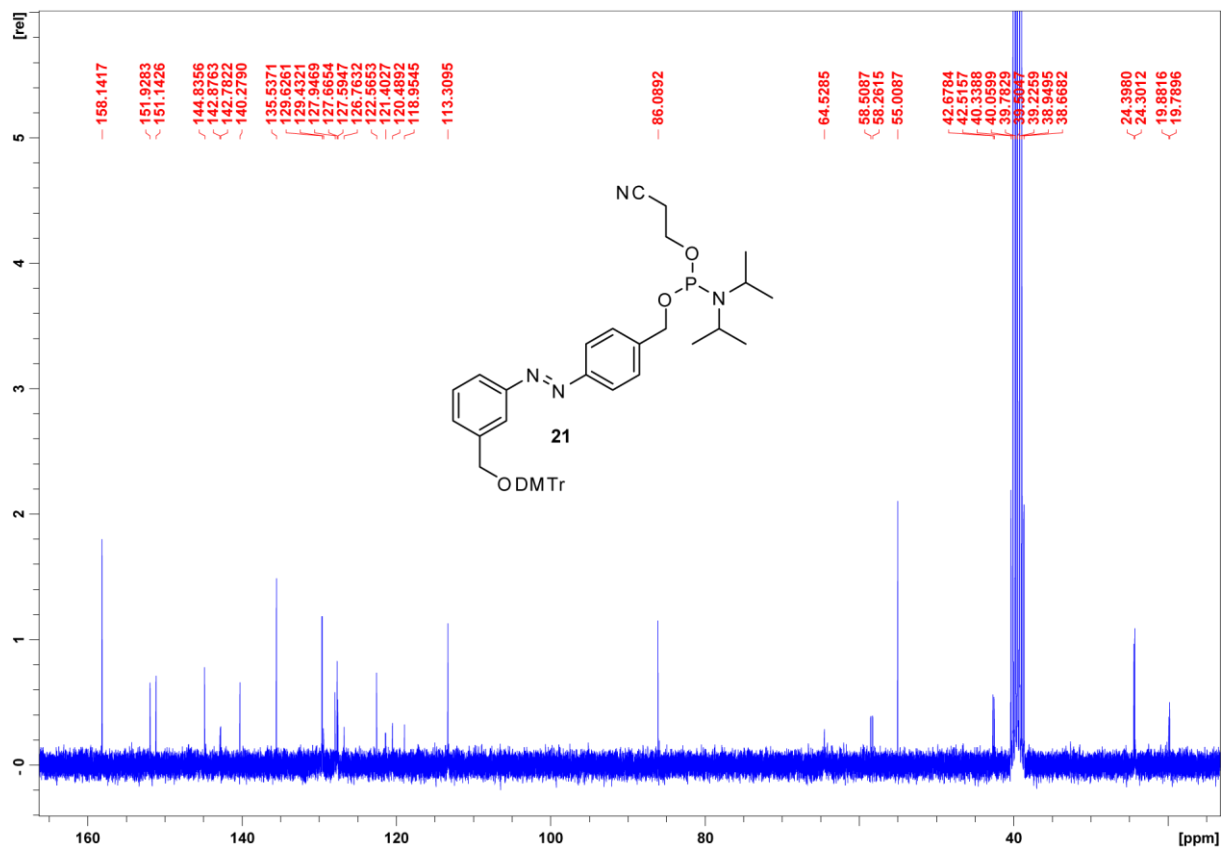


Figure S31. ^{13}C -NMR spectrum of **21** in $[\text{D}_6]\text{DMSO}$.

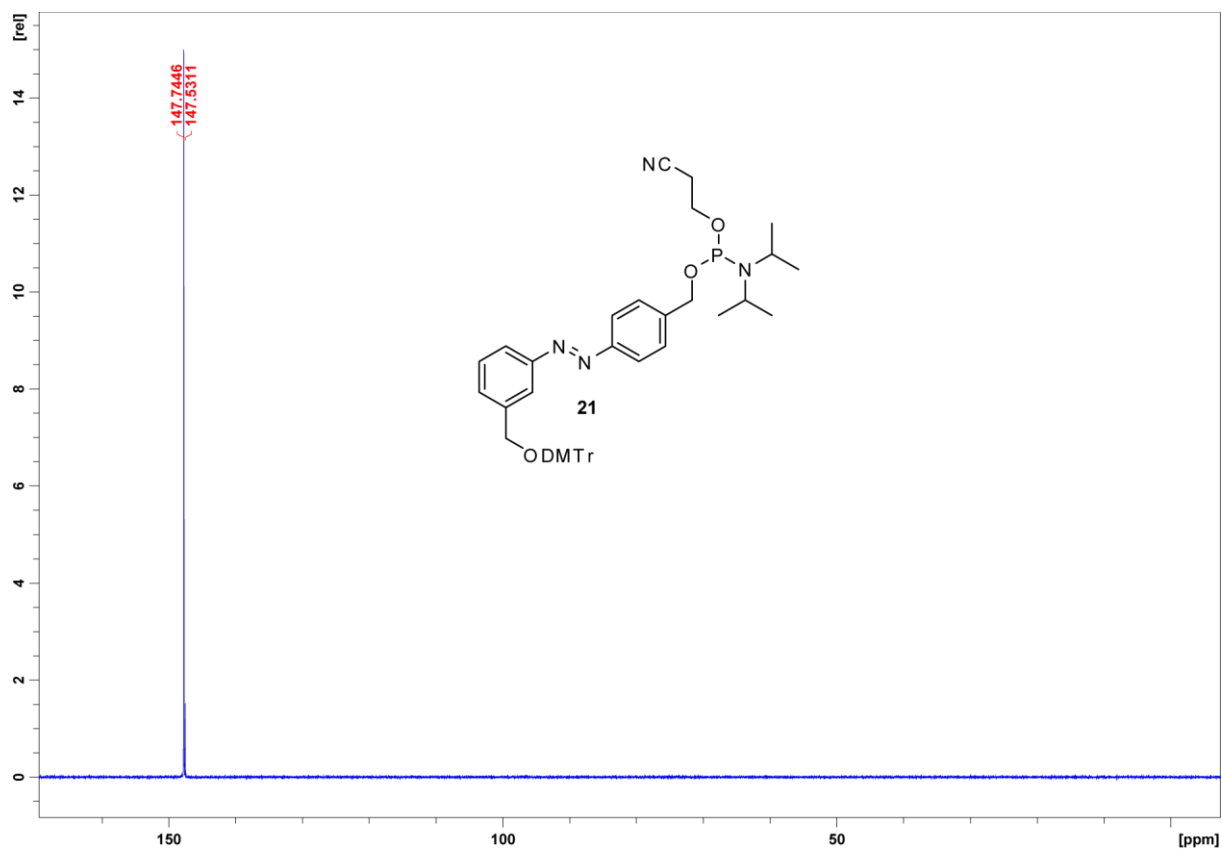


Figure S32. ^{31}P -NMR spectrum of **21** in $[\text{D}_6]\text{DMSO}$.

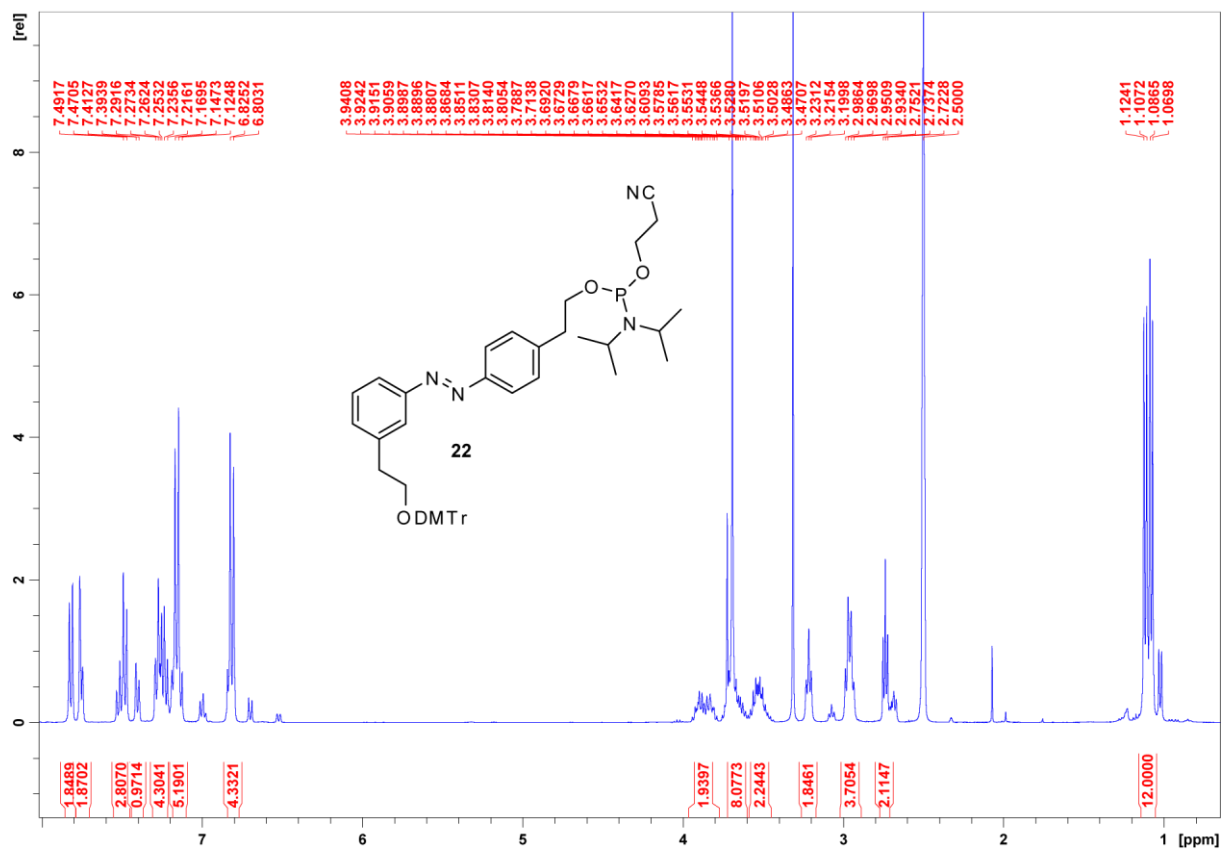


Figure S33. ¹H-NMR spectrum of **22** in [D₆]DMSO.

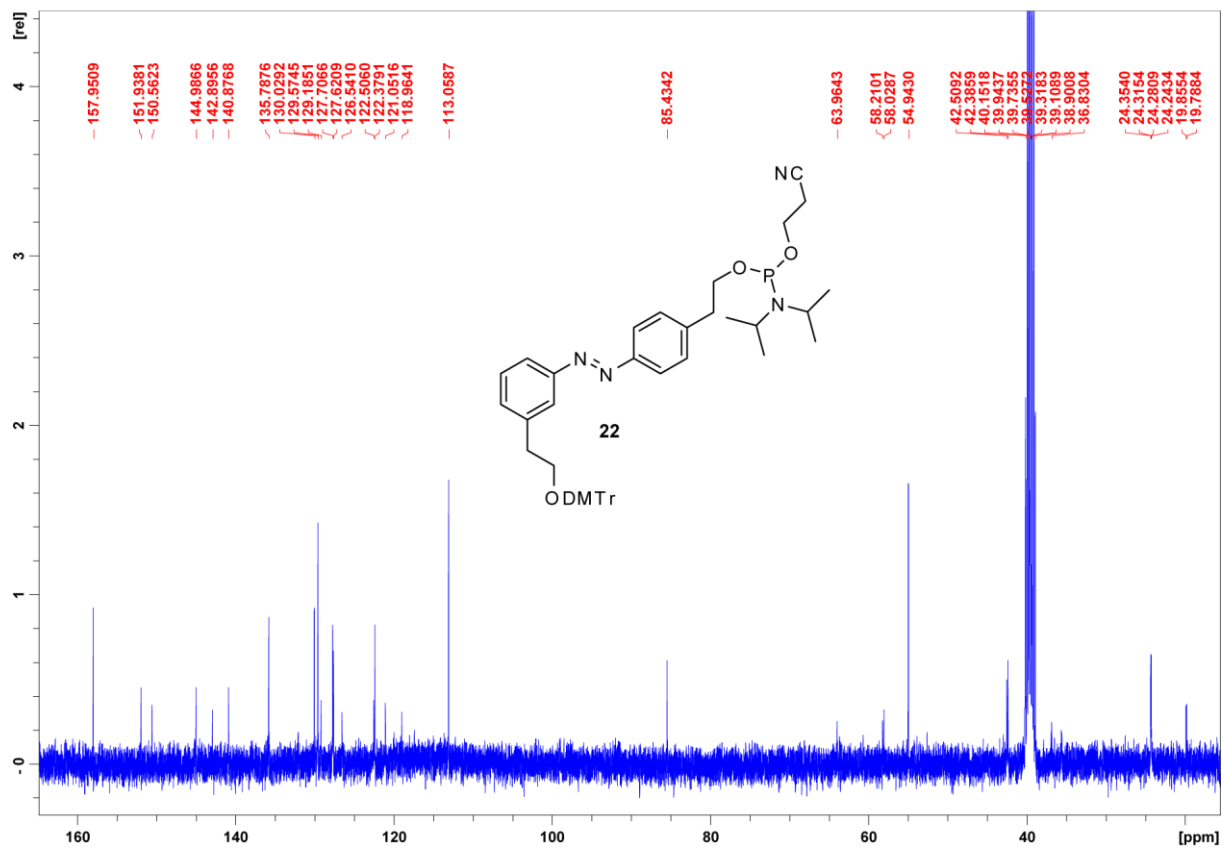


Figure S34. ¹³C-NMR spectrum of **22** in [D₆]DMSO.

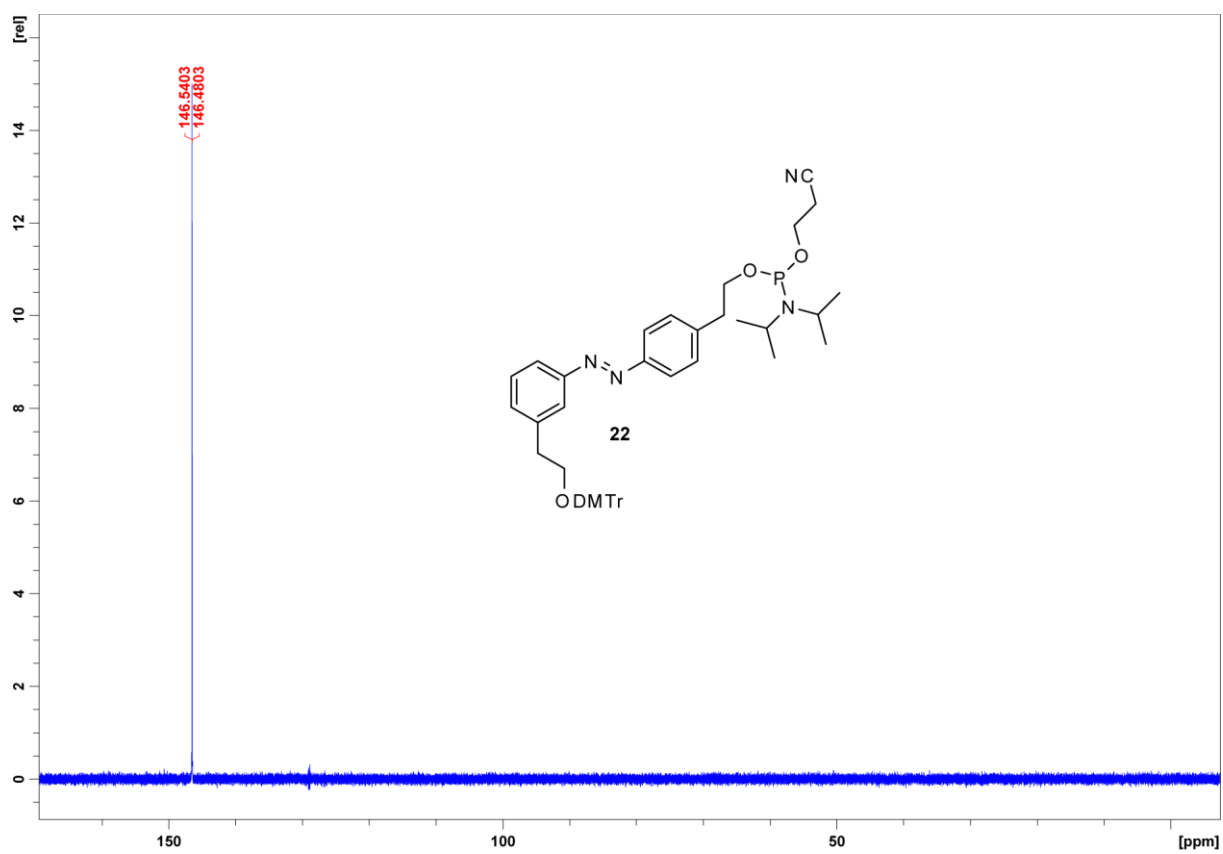


Figure S35. ^{31}P -NMR spectrum of **22** in $[\text{D}_6]\text{DMSO}$.

NMR measurements

NMR sample preparation

The oligonucleotides containing azobenzene linkers, obtained as described on page 5, were precipitated overnight at -20°C with 5 volumes of 2% w/v LiClO_4 in acetone and lyophilized in order to remove residual solvents from HPLC purification.

All NMR spectra were referenced using 3-(Trimethylsilyl)-1-propanesulfonic acid (DSS) and dissolved in either 90% $\text{H}_2\text{O}/10\%$ D_2O or 100% D_2O .

Preliminary screening of GG-**Az1**-GG, GG-**Az2**-GG, GG-**Az3**-GG, GGG-**Az1**-GG, GG-**Az1**-GGG, GG-**Az1**-G and G-**Az1**-GG via 1D ^1H -NMR was performed on samples containing 20-100 μM oligonucleotide in 50 mM Tris-HCl buffer (pH 7.4). Increasing amounts of KCl/NaCl were added directly in the NMR tube in order to obtain the final salt concentration specified in the figures. The sample volume change after each addition was negligible.

NMR samples used for the characterization of GG-**Az1**-GG contained 20-550 μM oligonucleotide in 25 or 50 mM deuterated Tris-HCl buffer (pH 7.4), 50 mM KCl. The sample used for DOSY spectroscopy additionally contained 140 μM 1,4-dioxane as an internal reference.

NMR experiments

All NMR data were collected, processed and analyzed using the software TopSpin 3.2 (Bruker Biospin) and Sparky 3.114.^[2]

^1H -NMR spectra were recorded using either gradient-tailored excitation^[3] or jump-return echo^[4] (excitation maximum at around 11.5 ppm) for water suppression. 2D $^1\text{H},^{13}\text{C}$ -HSQC^[5-7], long-range 2D $^1\text{H},^{13}\text{C}$ -HMBC^[8] and 2D $^1\text{H},^1\text{H}$ -NOESY spectra (50, 100 and 150 ms mixing time) with different water suppression schemes^[3,4] were recorded.

High resolution $^1\text{H},^{13}\text{C}$ -HSQC^[9,10] was recorded on a natural abundance sample to derive $^3J(\text{H}3',\text{P}3')$ and $^3J(\text{H}5'/\text{H}5'',\text{P}5')$ coupling constants (Figure S49 shows the E.COSY pattern) which provided restraints for the backbone torsional angles β and ϵ , respectively^[11]. $^3J(\text{H}2',\text{H}1')$ and $^3J(\text{H}2'',\text{H}1')$ coupling constants were measured in the 2D $^1\text{H},^1\text{H}$ -P.E.COSY^[12] (Figure S48 shows the $\text{H}2',\text{H}1'$ and $\text{H}2'',\text{H}1'$ region) in order to confirm the sugar pucker conformation.

DOSY experiments were performed to determine the hydrodynamic radius of the system. The diffusion time and gradient length was optimized to 125 ms and 4.5 ms respectively. As the hydrodynamic radius of 1,4-dioxane is given as 2.12 \AA ^[13], the hydrodynamic radius of GG-**Az1**-GG could be determined utilizing the formula: $R_{\text{H(DNA)}} = D_{\text{dioxane}}/D_{\text{DNA}} \cdot 2.12 \text{ \AA}$. A hydrodynamic radius of $13.9 \pm 0.5 \text{ \AA}$ could be measured. Calculation of the hydrodynamic radius in Hydropro^[14] yielded a value of 13.9 \AA for GG-**Az1**-GG in the tetrameric model and 11.4 \AA in a dimeric model.

As a proof of principle, a DOSY spectrum of the G-quadruplex forming DNA wtTel26 was recorded and analyzed with the same procedure (data not shown). The hydrodynamic radius was determined as $16.3 \pm 0.5 \text{ \AA}$ by DOSY and 16.1 \AA by calculation.

Screening of the azobenzene derivatives GG-Az1-GG, GG-Az2-GG and GG-Az3-GG using 1D ^1H -NMR spectra

GG-Az1-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl

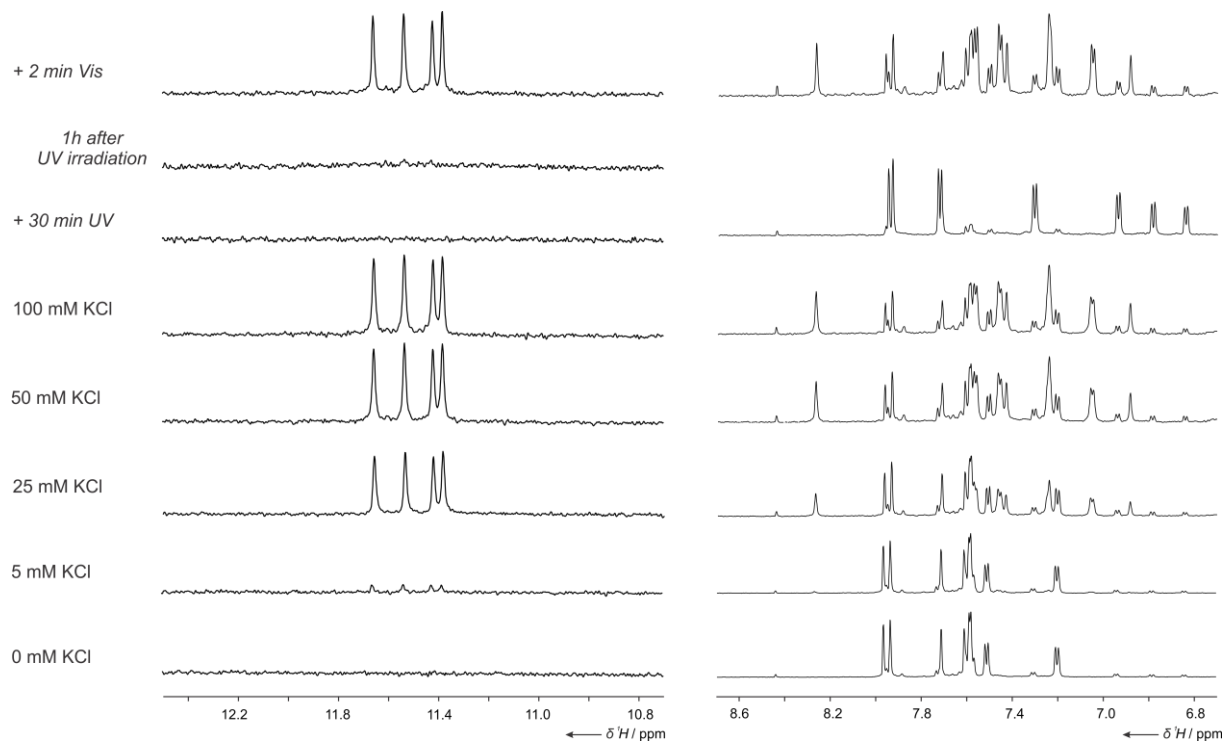


Figure S36. Imino (left) and aromatic (right) region of 1D ^1H -NMR spectrum of 50 μM GG-Az1-GG in presence of increasing amount of KCl. At 100 mM KCl, the sample was irradiated with UV light and then with visible light. Experimental conditions: 50 mM Tris-HCl buffer (pH 7.4), 90% H_2O /10% D_2O , 298K, 600 MHz.

GG-Az2-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl

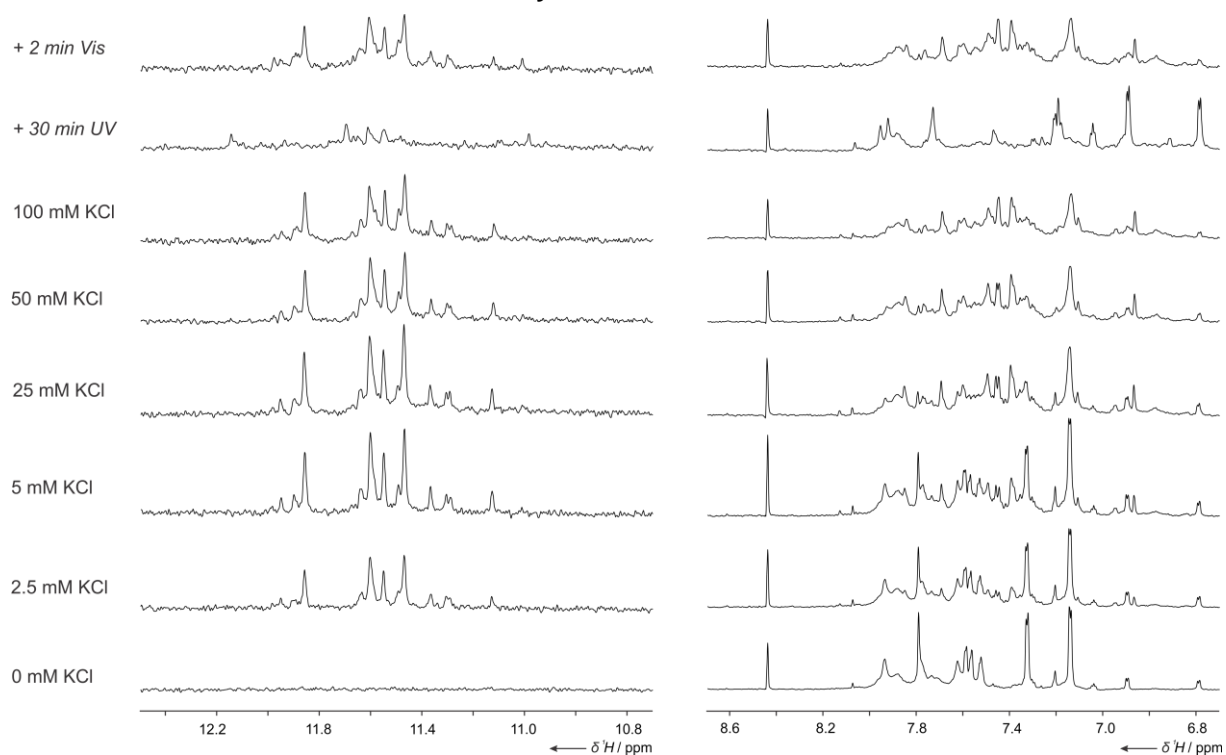


Figure S37. Imino (left) and aromatic (right) region of 1D ¹H-NMR spectrum of 50 μM GG-Az2-GG in presence of increasing amount of KCl. At 100 mM KCl, the sample was irradiated with UV light and then with visible light. Experimental conditions: 50 mM Tris-HCl buffer (pH 7.4), 90% H₂O/10% D₂O, 298K, 950 MHz.

GG-Az3-GG: Titration with KCl followed by UV/vis irradiation at 100 mM KCl

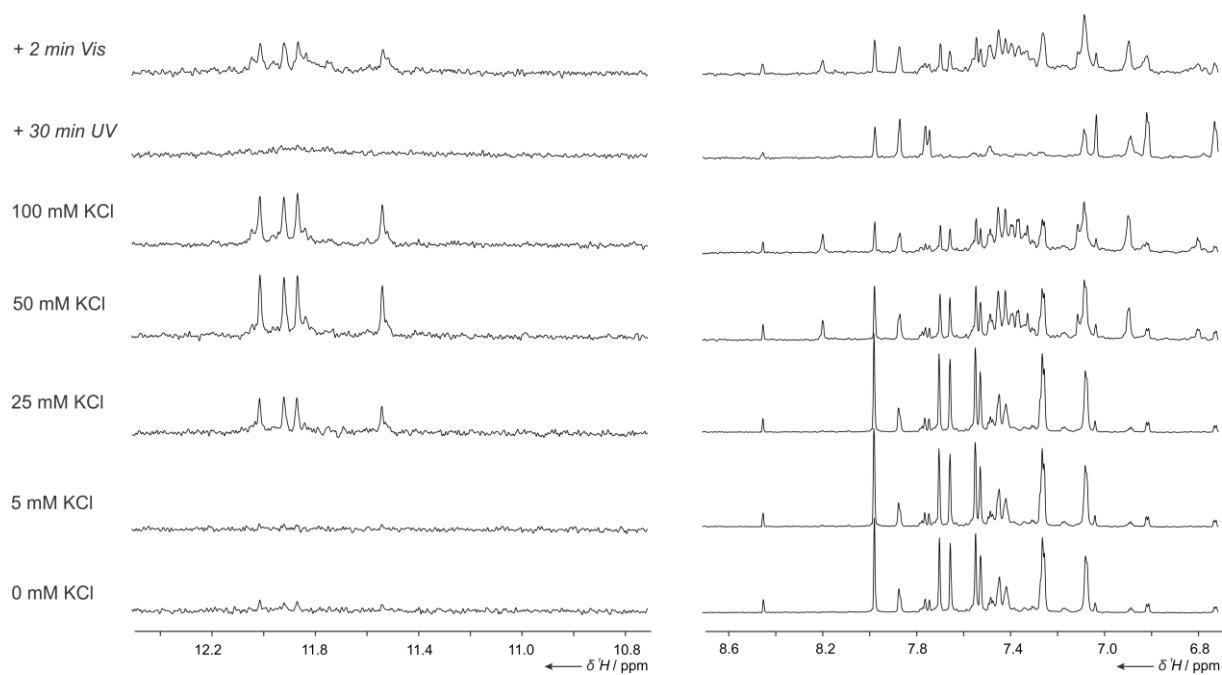


Figure S38. Imino (left) and aromatic (right) region of 1D ¹H-NMR spectrum of 50 μM GG-Az3-GG in presence of increasing amount of KCl. At 100 mM KCl, the sample was irradiated with UV light and then with visible light. Experimental conditions: 50 mM Tris-HCl buffer (pH 7.4), 90% H₂O/10% D₂O, 298K, 950 MHz.

Ion selectivity of GG-Az1-GG: titration with NaCl and KCl followed by UV/vis irradiation at (100 mM NaCl + 100 mM KCl)

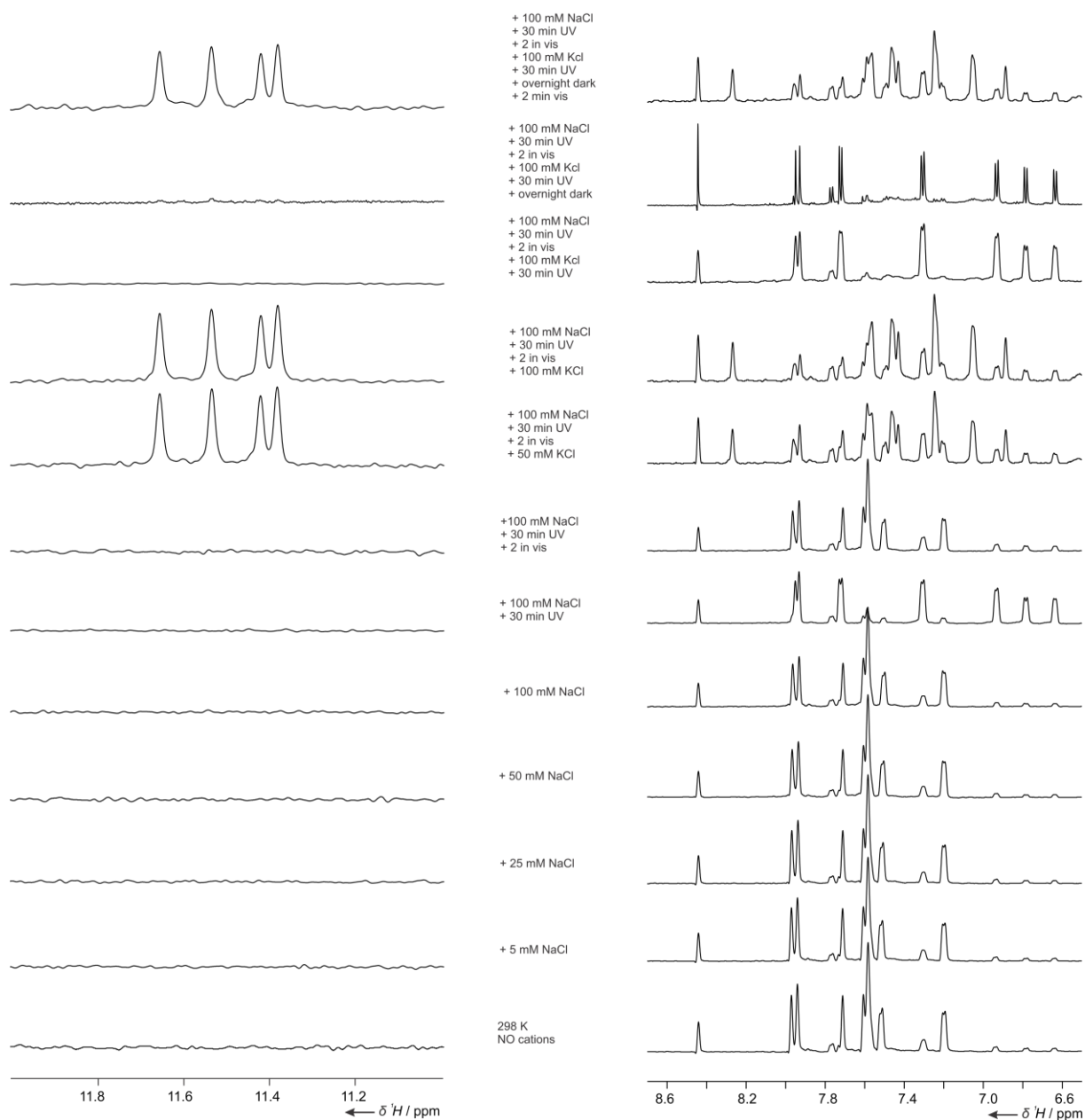


Figure S39. Imino (left) and aromatic (right) region of 1D $^1\text{H-NMR}$ spectrum of $50\ \mu\text{M}$ GG-Az1-GG in presence of increasing amount of NaCl and KCl. At 100 mM NaCl and 100 mM KCl, the sample was irradiated with UV light and then with visible light. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz.

Screening of the azobenzene derivatives GG-Az1-GGG, GGG-Az1-GG, G-Az1-GG and GG-Az1-G using 1D $^1\text{H-NMR}$ spectra

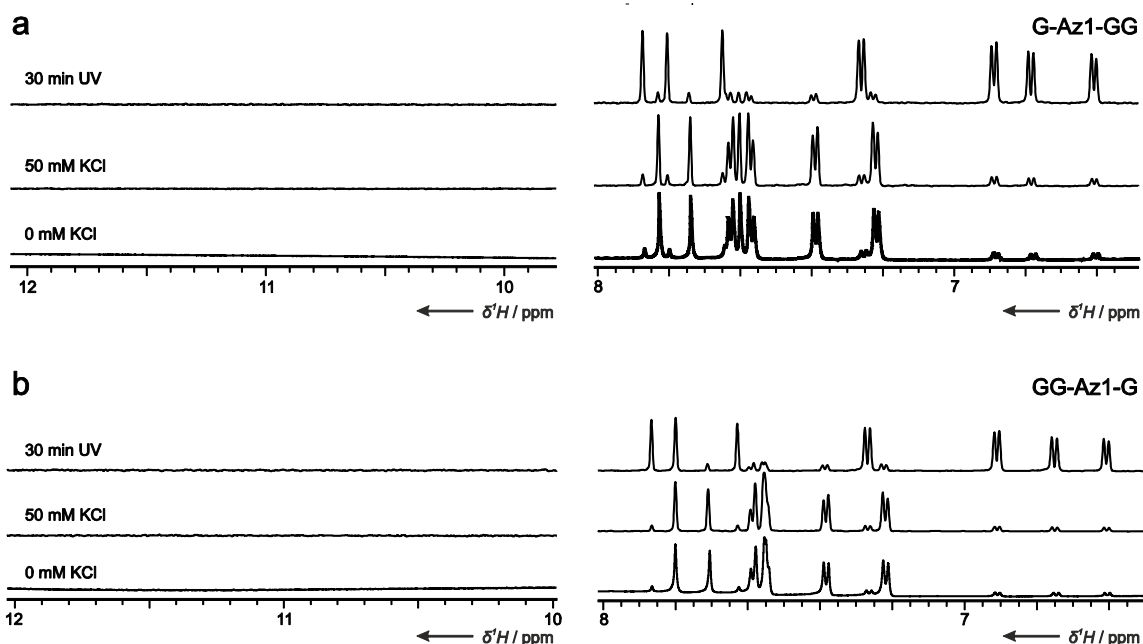


Figure S40. Imino (left) and aromatic (right) region of the 1D $^1\text{H-NMR}$ spectrum of 100 μM GG-Az1-G (a) and G-Az1-GG (b) without KCl and with 50 mM KCl. At 50 mM KCl, the sample was irradiated with UV light. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz.

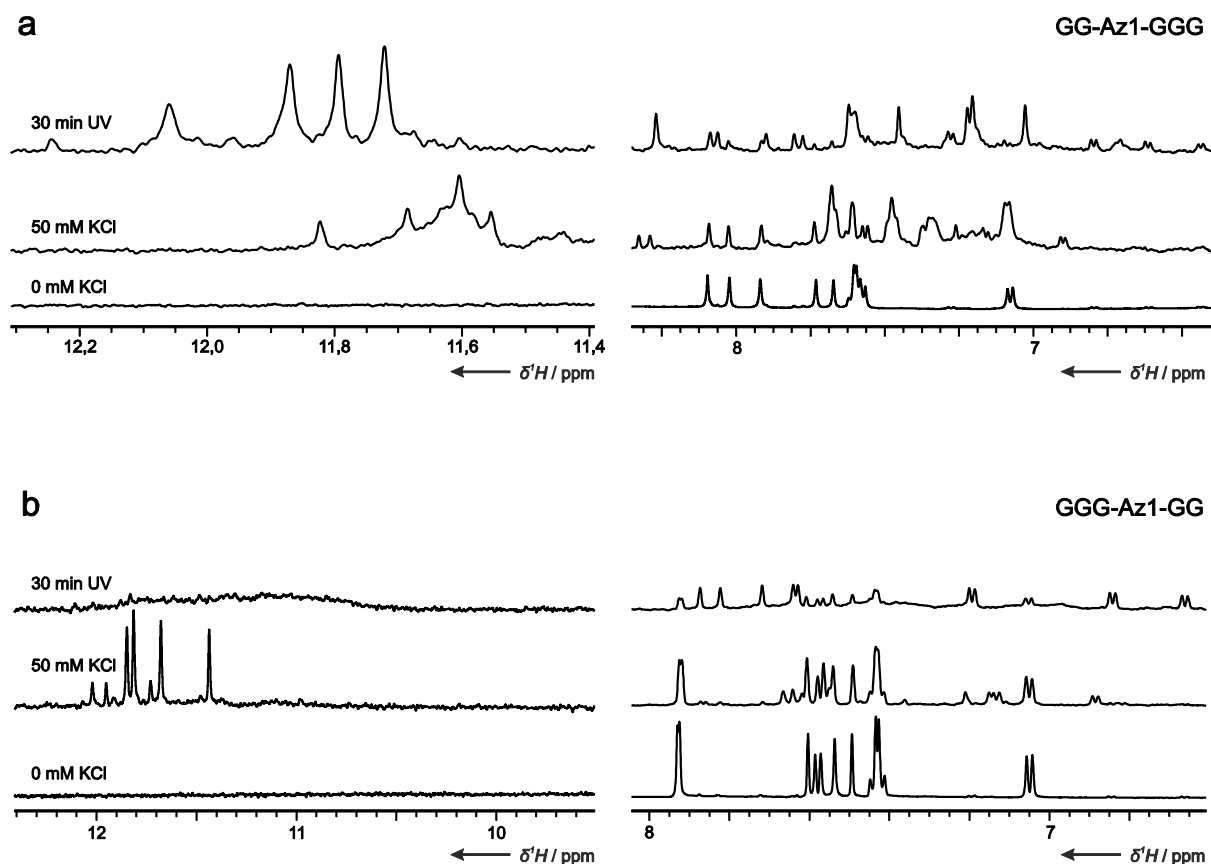


Figure S41. Imino (left) and aromatic (right) region of the 1D $^1\text{H-NMR}$ spectrum of 100 μM GG-Az1-GGG (a) and GGG-Az1-GG (b) without KCl and with 50 mM KCl. At 50 mM KCl, the sample was irradiated with UV light. Thermal recovery of the (*E*)-state monitored by 1D $^1\text{H-NMR}$

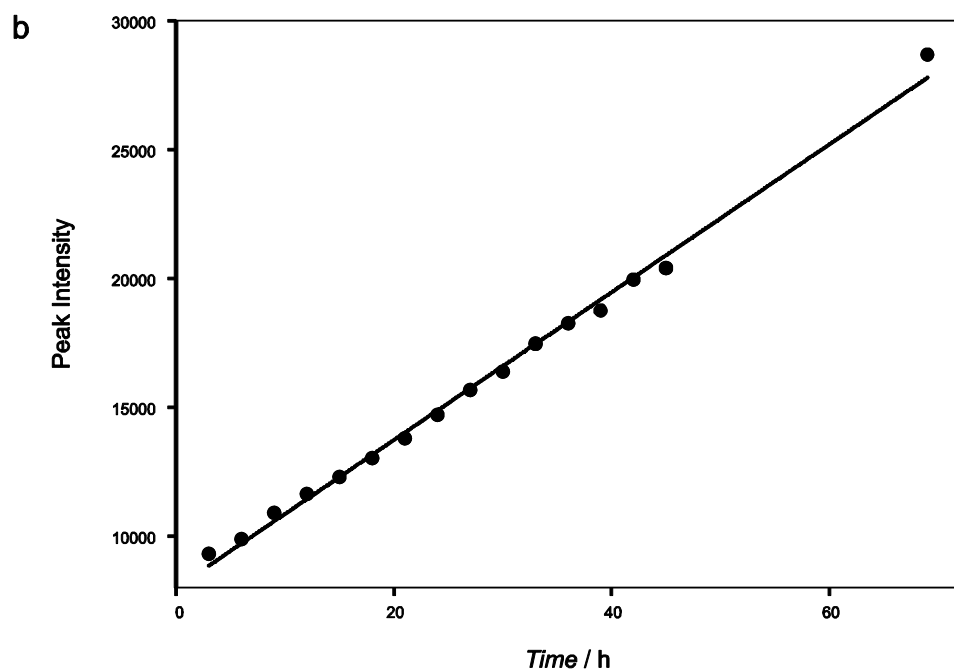
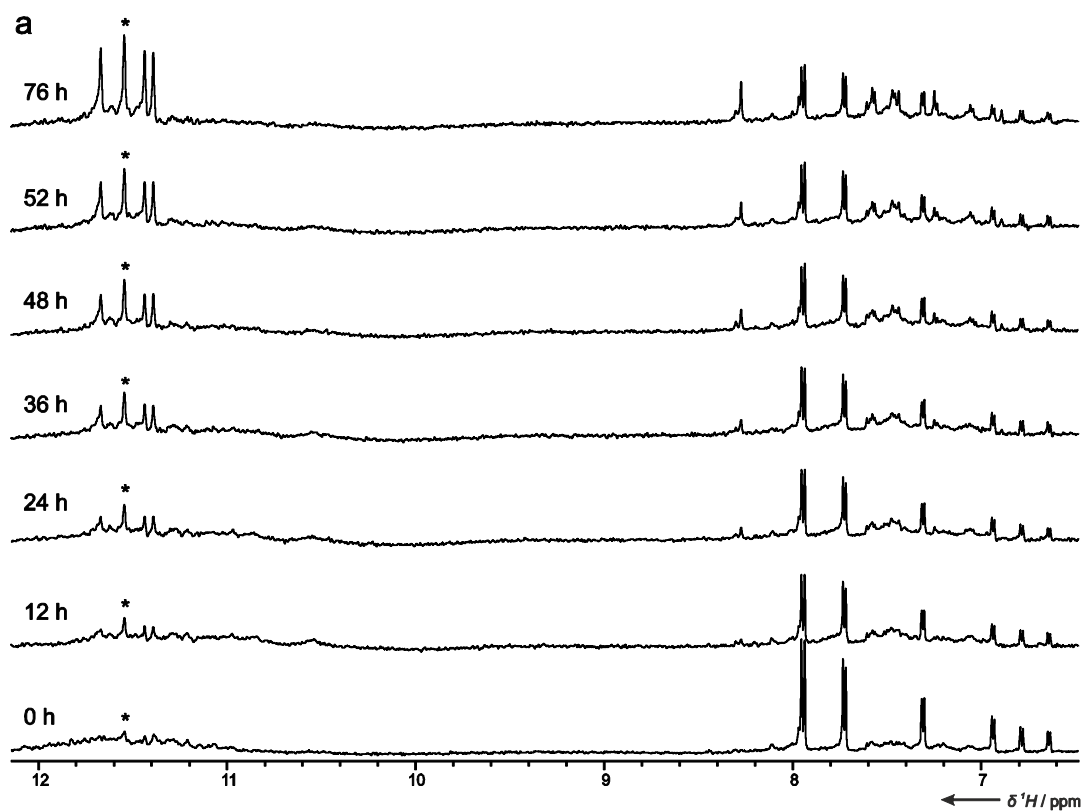


Figure S42. a) Imino (left) and aromatic (right) region of the 1D ¹H-NMR spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl after 30 minutes of UV irradiation. The sample was irradiated with UV light and kept in the dark. The thermal recovery of the *E*-state of the G-quadruplex was monitored by 1D ¹H-NMR. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% H₂O/10% D₂O, 298K, 600 MHz. b) Time dependent intensity-plot of the imino peak originating from G2 (marked with an asterisk in a).

Assignment of the 1D ^1H -NMR spectrum of GG-Az1-GG

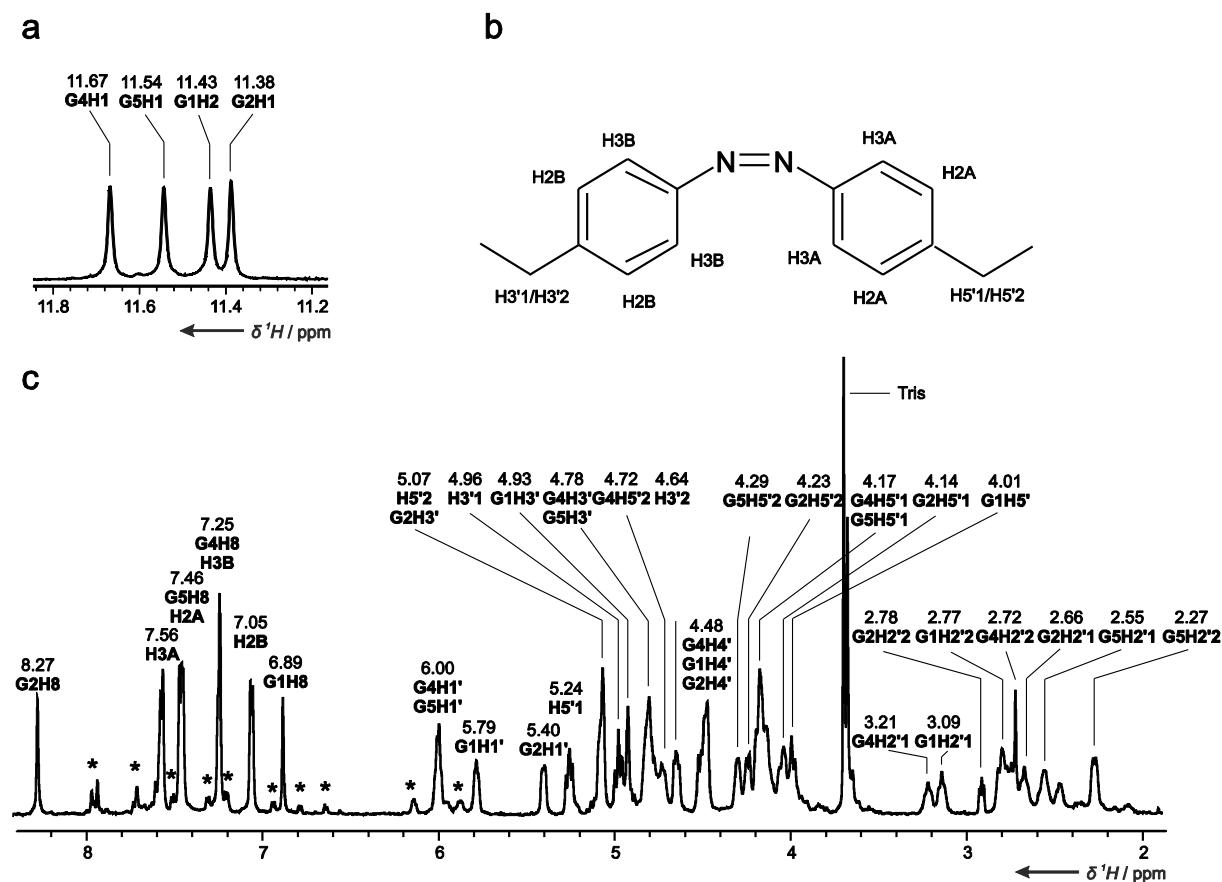


Figure S43. a) Assignment of the imino region of the 1D ^1H -NMR spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz. b) Numbering scheme of the azobenzene moiety, as used in the assignment. c) Assignment of the aromatic and sugar region of the 1D ^1H -NMR spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl. Signals of residual GG-Az1-GG in (Z)-state are marked with an asterisk. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz.

Full assignment of the 2D $^1\text{H},^1\text{H}$ -NOESY spectrum of GG-Az1-GG

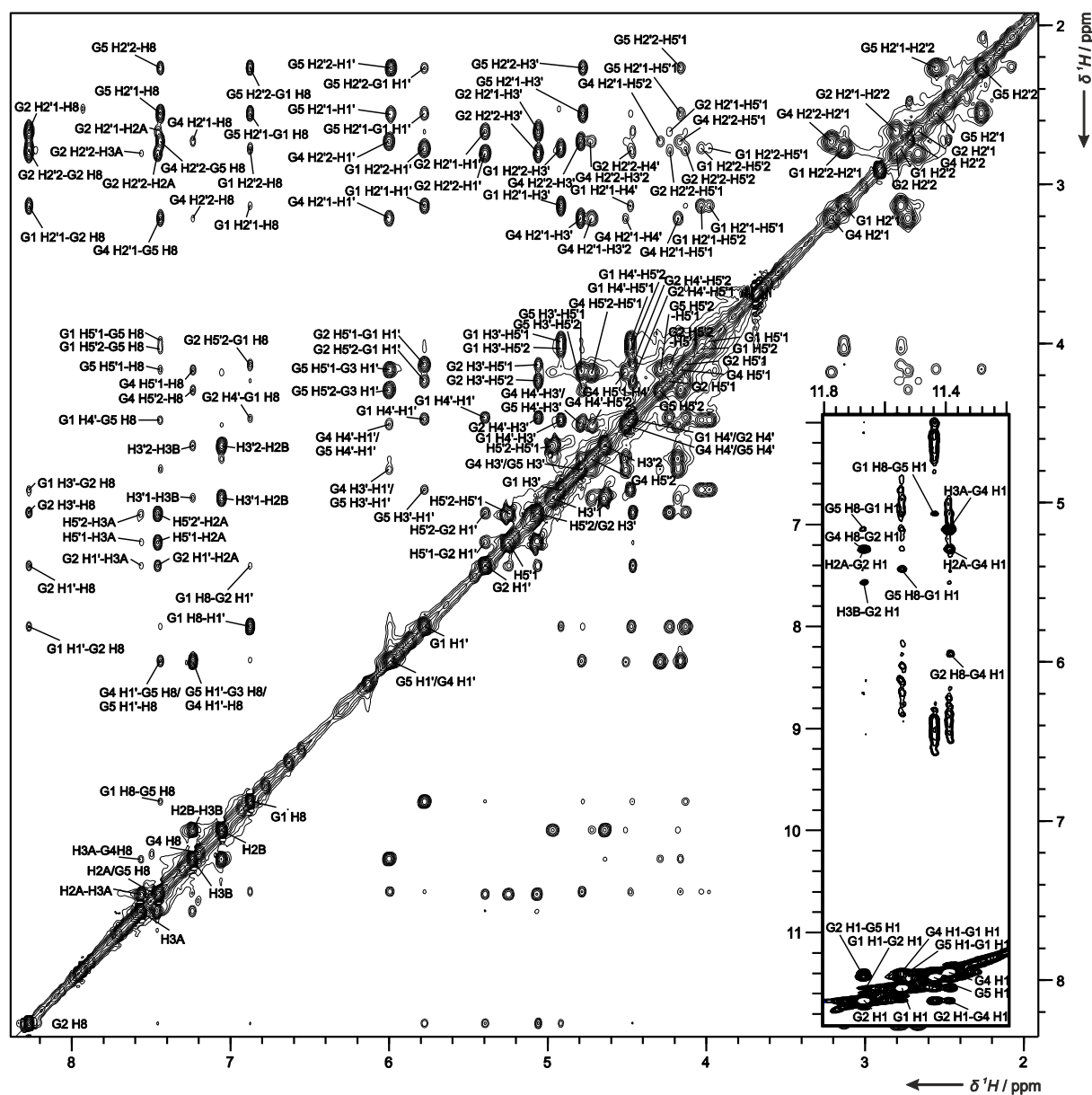


Figure S44. Assignment of the 2D $^1\text{H},^1\text{H}$ -NOESY spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 2048 x 572 points, 144 scans. Insert: Assignment of the imino region including the cross peaks to aromatic protons from the 2D $^1\text{H},^1\text{H}$ -NOESY spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz, 2048 x 572 points, 144 scans.

2D $^1\text{H},^{13}\text{C}$ -HMBC spectrum of GG-Az1-GG: C5 region

2D $^1\text{H},^1\text{H}$ -NOESY spectrum: H1-H8 and H8-H1' region

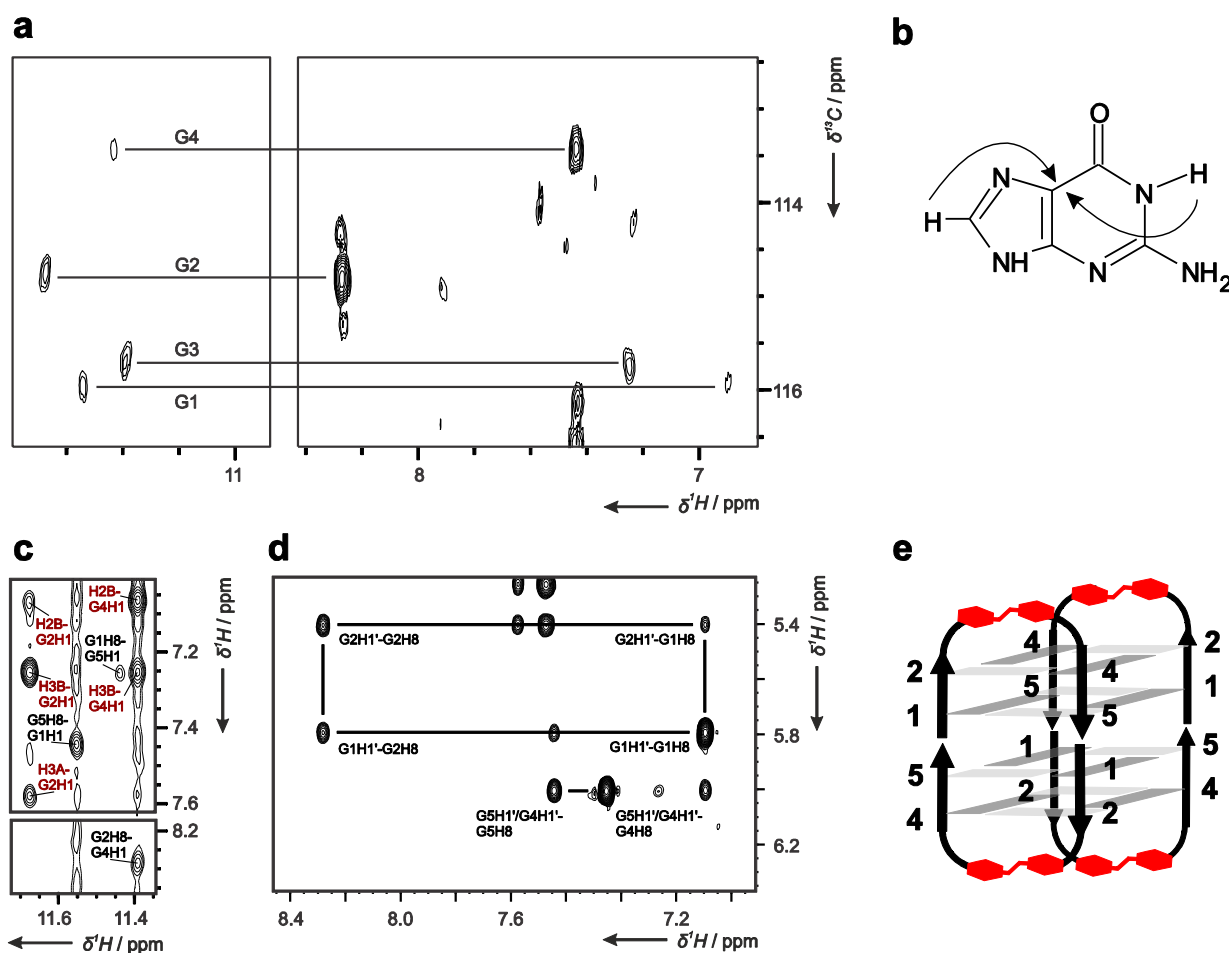


Figure S45. a) 2D $^1\text{H},^{13}\text{C}$ -HMBC natural abundance spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl correlating the imino proton and the H8 proton belonging to the same guanine residue. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz, 2048 x 32 points, 1600 scans. b) Schematic of the magnetization transfer in the long-range $^1\text{H},^{13}\text{C}$ -HMBC experiment. c) Region of the $^1\text{H},^1\text{H}$ -NOESY spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl, showing the imino to aromatic (black annotation) and imino to azobenzene (red annotation) cross peaks. The imino to aromatic interactions allowed defining the intra-tetrad connectivities as shown in panel d). Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz, 2048 x 572 points, 144 scans. d) Region of the $^1\text{H},^1\text{H}$ -NOESY spectrum of 550 μM GG-Az1-GG in presence of 50 mM KCl, showing the H8 to H1' crosspeaks. A rectangular cross-peaks pattern can be observed at the G1-G2 step (black line) and at the G4-G5 step (overlapped), indicating a 5'-*syn-anti*-3' arrangement. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 2048 x 572 points, 144 scans. e) Schematic representation of the folding topology, derived from the NOE assignment. Guanine residues characterized by *anti* glycosidic bond are coloured in light grey, while guanine residues adopting a *syn* glycosidic bond are coloured in dark grey.

H-D exchange experiment: imino region of the 1D ^1H -NMR spectrum of GG-Az1-GG before and after exchange with D_2O

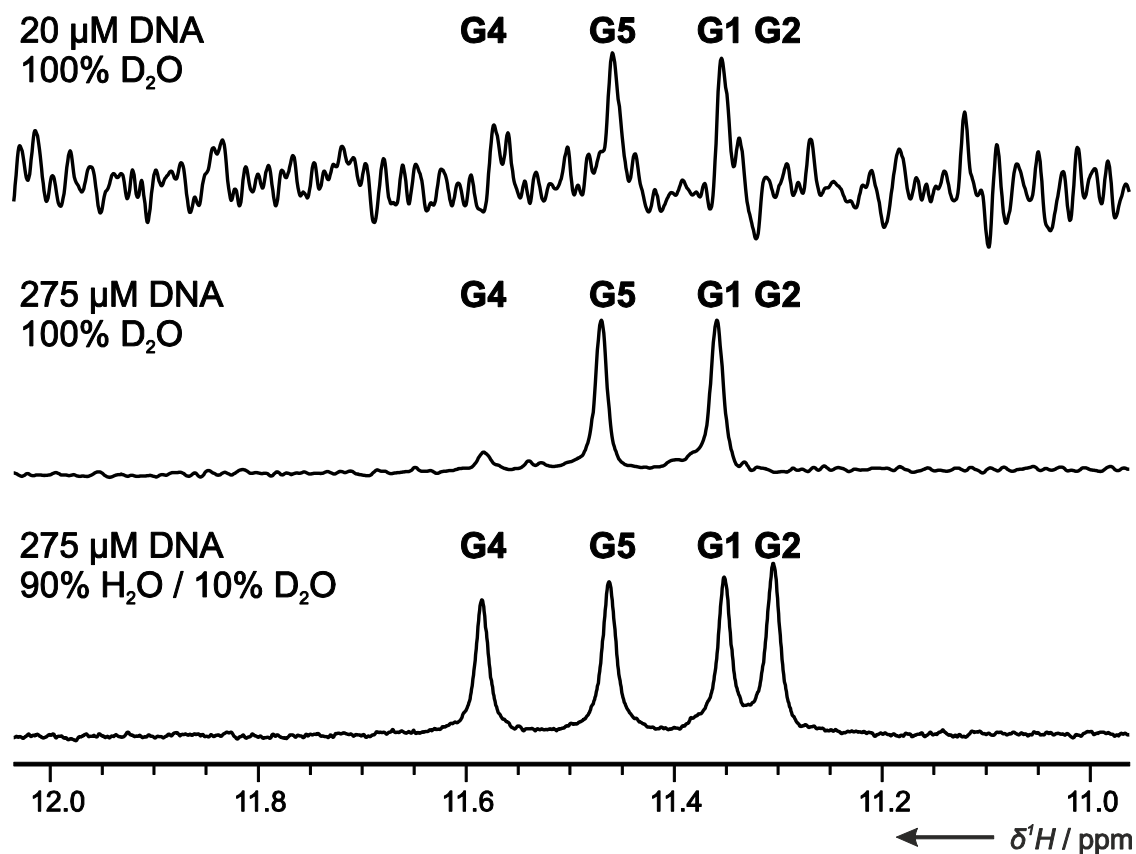


Figure S46. Imino region of the 1D ^1H -NMR spectrum of 275 μM or 20 μM GG-Az1-GG in presence of 50 mM KCl. Experimental conditions (90% H_2O /10% D_2O): 25 mM Tris-HCl buffer (pH 7.4), 90% H_2O /10% D_2O , 298K, 600 MHz. Experimental conditions (100% D_2O): 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz.

DOSY spectrum of GG-Az1-GG

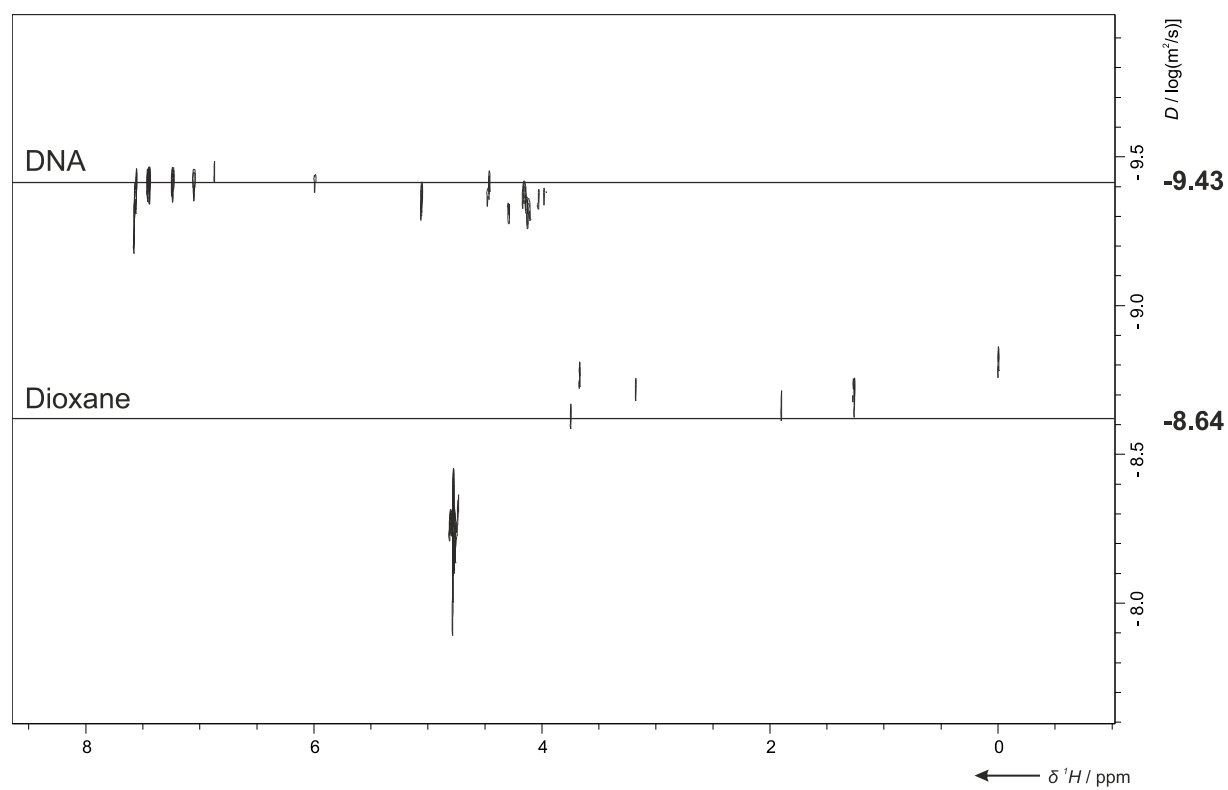


Figure S47. DOSY spectrum of 160 μM GG-Az1-GG in presence of 50 mM KCl and 140 μM 1,4-dioxane. Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 2048 x 64 points, 64 scans.

2D $^1\text{H},^1\text{H}$ -P.E.COSY of GG-Az1-GG: H1'-H2' and H1'-H2'' region

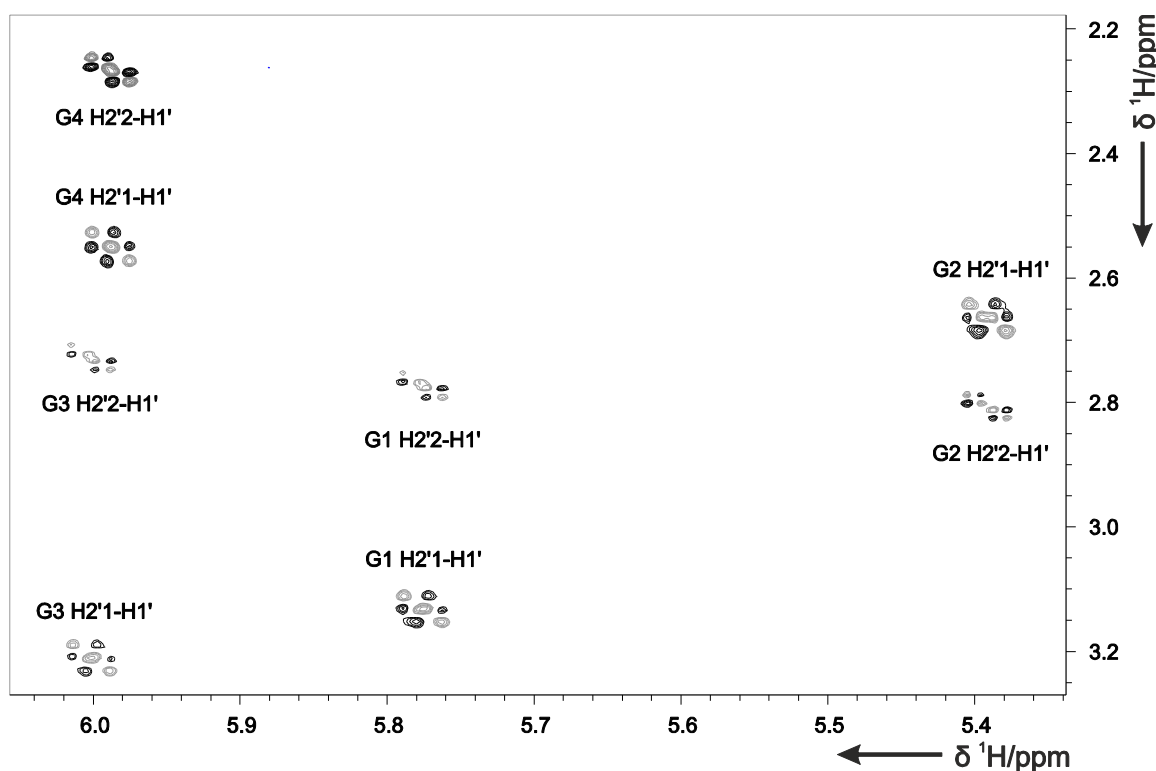


Figure S48. H2'/H2''-H1' region of the 2D $^1\text{H},^1\text{H}$ -P.E.COSY spectrum of 420 μM GG-Az1-GG in presence of 50 mM KCl. Experimental conditions 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 4096 x 2048 points, 32 scans.

2D $^1\text{H},^{13}\text{C}$ -HSQC of GG-Az1-GG: Zoom on a H5'-C5' peak

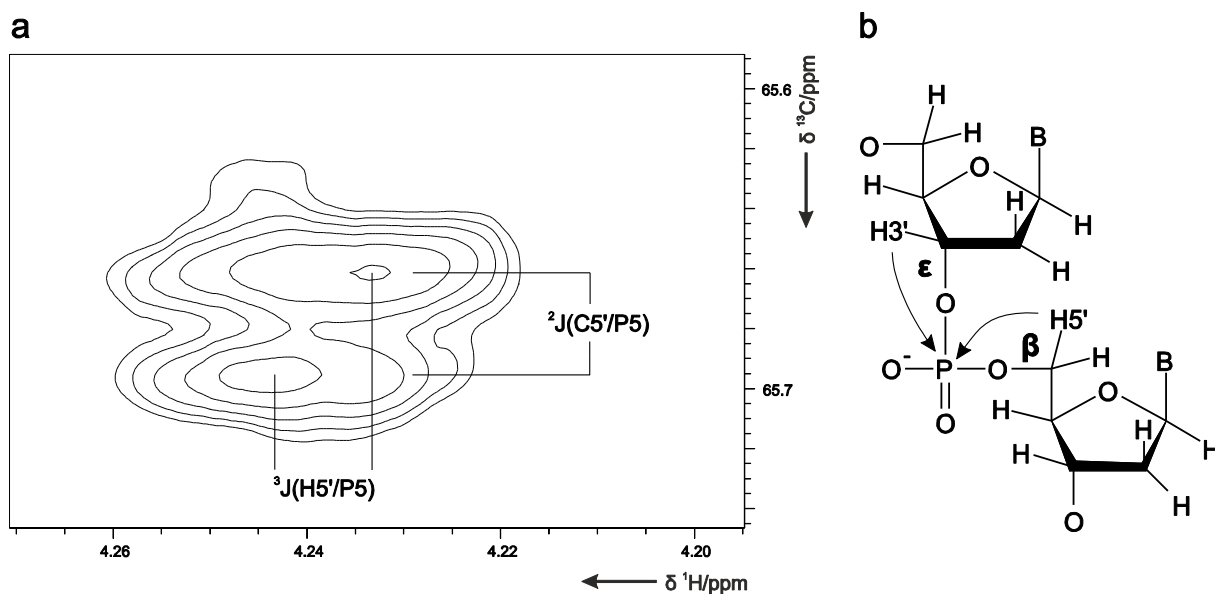


Figure S49. a) Zoom on a H5'C5' peak of the high-resolution $^1\text{H},^{13}\text{C}$ -HSQC spectrum of GG-Az1-GG in presence of 50 mM KCl. Experimental conditions 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 1024 x 3072 points, 128 scans. b) Schematic representation of the magnetization transfer in the high-resolution $^1\text{H},^{13}\text{C}$ -HSQC experiment.

2D $^1\text{H}, ^{31}\text{P}$ -hetero TOCSY-NOESY of GG-Az1-GG:

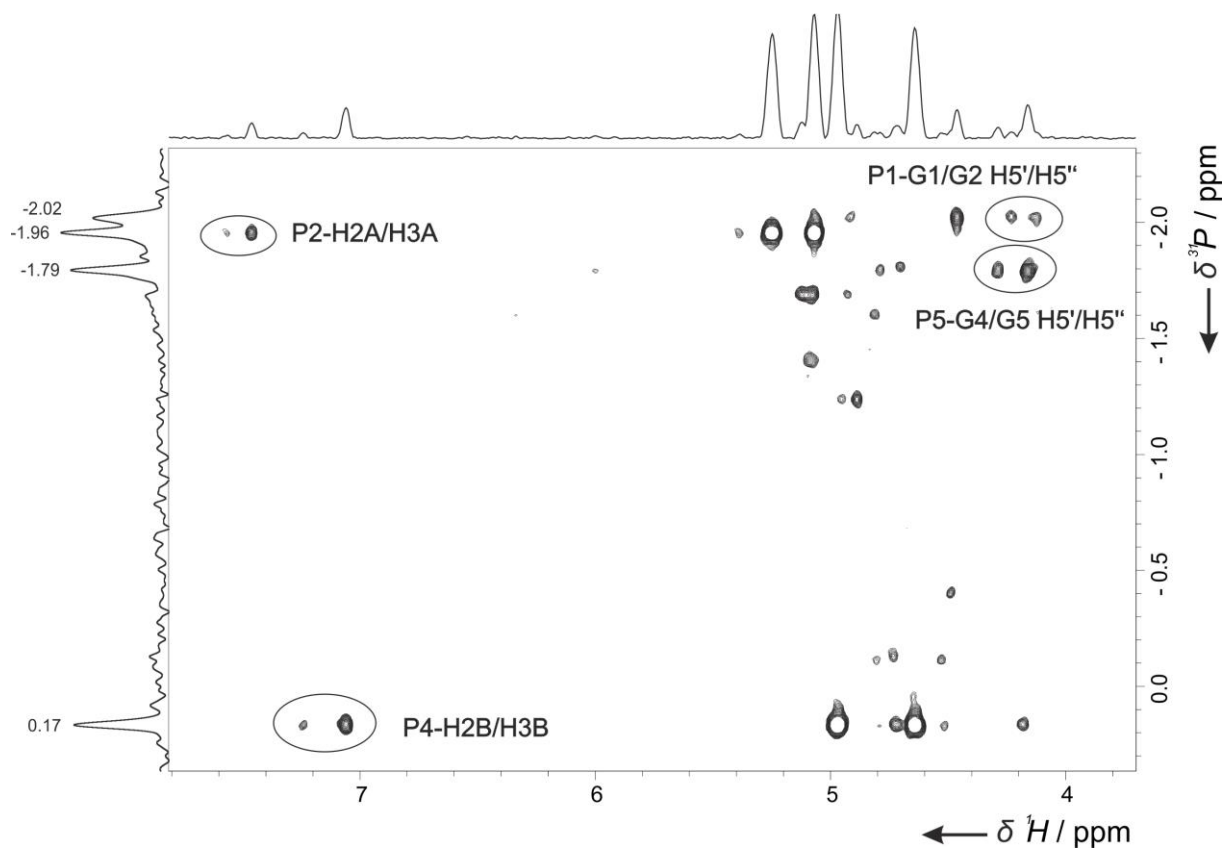


Figure S50. $^1\text{H}, ^{31}\text{P}$ -TOCSY spectrum of GG-Az1-GG in presence of 50 mM KCl. Experimental conditions 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 1024 x 128 points, 512 scans. Peaks that led to the assignment of ^{31}P chemical shifts are marked.

2D $^1\text{H},^{13}\text{C}$ -HSQC of GG-Az1-GG:

C3'-H3', C5'-H5'/H5'', C4'-H4' and C2'-H2'/H2'' region

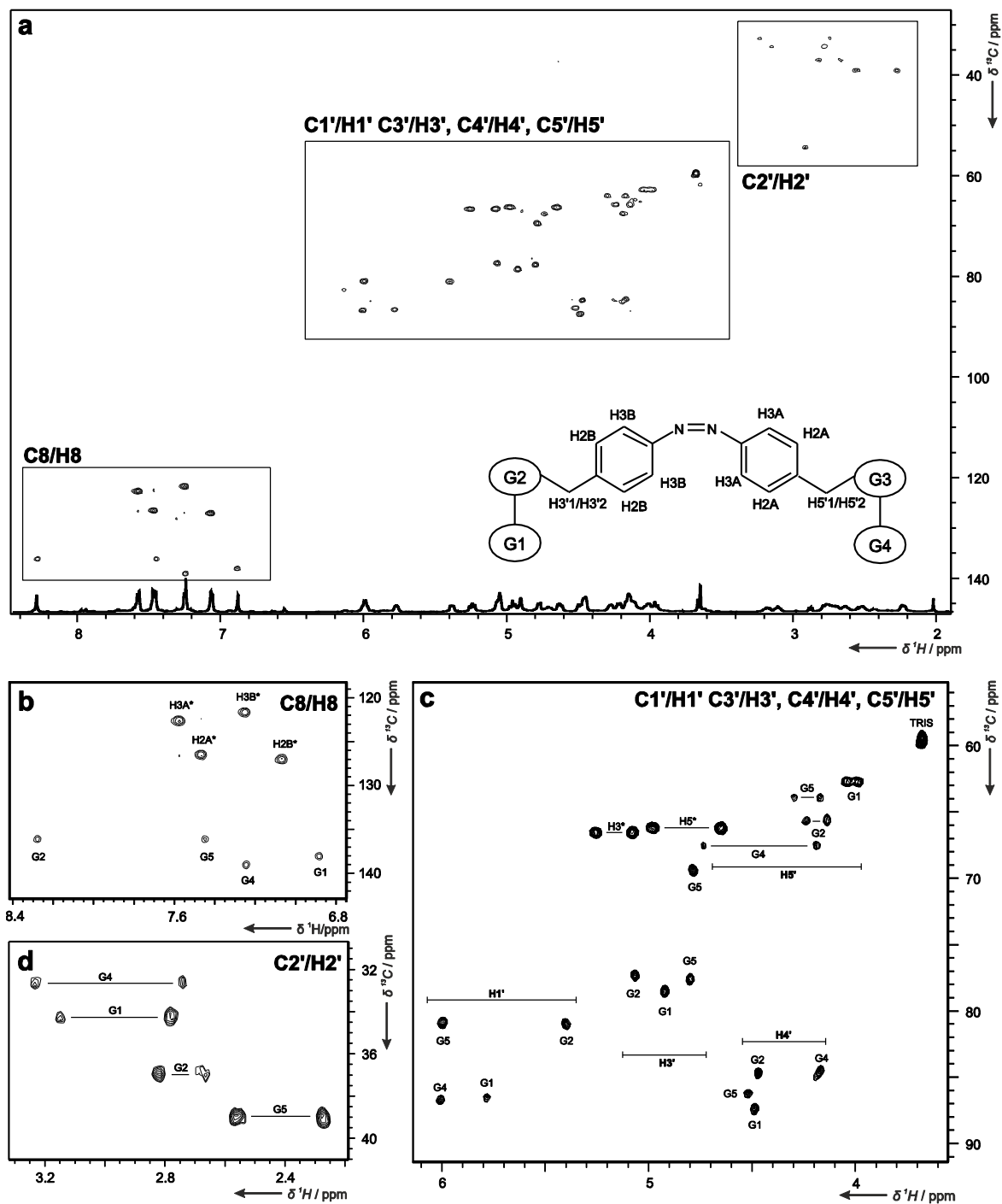


Figure S51. a) $^1\text{H},^{13}\text{C}$ -HSQC spectrum of 420 μM GG-Az1-GG in presence of 50 mM KCl. The insert shows the naming of the protons of the Az1 moiety as used in the assignment. Zooms of the regions marked with a box are shown in panels b), c) and d). Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 100% D_2O , 298K, 600 MHz, 1024 x 256 points, 128 scans. b) Zoom on the C8/H8 region with an assignment of the residues. Peaks of the Az1 moiety are marked with an asterisk and named according to the scheme in a). c) Zoom on the sugar region including an assignment of residues. Peaks of the Az1 moiety are marked with an asterisk and named according to the scheme in a). d) Zoom on the C2'H2' region including an assignment of the residues.

Structure calculations

The NMR structure calculations were performed with CNS 1.1^[15] using the ARIA 1.2^[16] adapted setup^[17] and protocols with the nucleic acid forcefield including OPLS charges and nonbonded parameters.^[18] The topology and parameters for the **Az1** group were generated by hand and are consistent with the nucleic acid and protein forcefields. Three potassium ions were included within the quadruplex tetrads for structure refinement. The standard annealing protocols were used, including NOE distance calibration with spin-diffusion correction. Furthermore, dihedral angle restraints, Watson-Crick basepair hydrogen bonds and weak base-planarity restraints have been added. The proton chemical shift resonances were manually assigned by standard 2D NMR methods using ¹H,¹H-NOESY, ¹H,¹H-COSY, ¹H,¹³C-HSQC, and ¹H,¹³C-HMBC. The 2D ¹H,¹H-NOESY spectra, recorded at 600 MHz, were manually peak picked and assigned to a large extent using Sparky^[2]. The proton resonances and 6 NOE peak lists (with different mixing times: 50, 100 and 100 ms, in D₂O as well as H₂O) were used as input for automated NOESY cross peak assignment and calibration with ARIA. The NOE data was attributed to originate from one monomer in the tetramer environment and subsequently multiplied for the other monomers in accordance with the symmetry. The chemical shift tolerance was set to 0.015 ppm in both proton dimensions. For each iteration (0-7) in which 50 structures were calculated, the NOE distance restraints were recalibrated by ARIA based on the 10 lowest energy structures. The violation tolerance was progressively reduced to 0.1 Å in the last iteration (8) in which 200 structures were calculated. For the structure calculations, a four stage simulated annealing (SA) protocol was used using cartesian angle dynamics. The high temperature stage consisted of 10000 steps at 10000 K. This was followed by refinement and cooling down stages: 8000 steps at 2000 K, 20000 steps to 1000 K and 15000 steps to 50 K. The planarity of the base-pairs were weakly restrained (25 kcal*mol⁻¹Å⁻²) and the potassium ions were coordinated by ample distance restraints. Based on the NOE only structure and confirmed by ³J-couplings and chemical shifts (among others from ¹H,¹H-E.COSY and ¹H,³¹P-TOCSY spectra), ample dihedral angle restraints were included for the sugar pucker of G1, G2 and G4 (2'-*endo*), the glycosidic torsion angle (*syn* for G1 and G4, *anti* for G2 and G5) and the backbone in between the tetrads, only to improve the convergence. During the SA protocol the force constant was progressively increased to 50 kcal*mol⁻¹Å⁻² for the NOE distance restraints and hydrogen bonds and to 200 kcalmol⁻¹rad⁻² for the torsion angles. The final 20 lowest energy structures were further refined in explicit water^[19].

Chemical Shift Resonances

Group	Atom	ppm	Atom	ppm
Gua-1	H1'	5.78	C1'	86.5
	H2' _{1,2}	3.12, 2.78	C2'	34.2
	H3'	4.92	C3'	78.5
	H4'	4.48	C4'	87.4
	H5' _{1,2} *	3.98, 4.03	C5'	62.7
	H8	6.88	C8	138.0
	H1	11.54		
Gua-2	H1'	5.40	C1'	81.0
	H2' _{1,2}	2.67, 2.80	C2'	37.0
	H3'	5.06	C3'	77.4
	H4'	4.47	C4'	84.7
	H5' _{1,2} *	4.14, 4.23	C5'	65.7
	H8	8.27	C8	136.1
	H1	11.67		
Az1-3	H3' _{1,2} *	4.64, 4.97	C3'	66.2
	H5' _{1,2} *	5.07, 5.25	C5'	66.5
	H2A#	7.46	C2A#	126.5
	H2B#	7.06	C2B#	127.0
	H3A#	7.57	C3A#	122.6
	H3B#	7.24	C3B#	121.7
Gua-4	H1'	6.00	C1'	86.7
	H2' _{1,2}	3.22, 2.73	C2'	32.6
	H3'	4.80	C3'	77.6
	H4'	4.51	C4'	86.2
	H5' _{1,2} *	4.18, 4.73	C5'	67.5
	H8	7.24	C8	139.0
	H1	11.39		
Gua-5	H1'	5.99	C1'	80.9
	H2' _{1,2}	2.56, 2.27	C2'	39.0
	H3'	4.78	C3'	69.4
	H4'	4.17	C4'	84.5
	H5' _{1,2} *	4.17, 4.29	C5'	63.9
	H8	7.44	C8	136.1
	H1	11.43		

* not stereospecific assigned

³J coupling constants

Gua-1	H1',H2' ₁	8.8 ± 0.5	H3',P	1.4 ± 0.5
	H1',H2' ₂	6.3 ± 0.2		
Gua-2	H1',H2' ₁	10.9 ± 0.3	H3',P	1.8 ± 0.5
	H1',H2' ₂	5.7 ± 0.5	H5' ₁ ,P	2.8 ± 0.5
				H5' ₂ ,P
Gua-4	H1',H2' ₁	9.7 ± 0.2	H3',P	1.2 ± 0.5
	H1',H2' ₂	6.7 ± 0.4	H5' ₁ ,P	1.2 ± 0.5
				H5' ₂ ,P
Gua-5	H1',H2' ₁	8.9 ± 0.1	H5' ₁ ,P	1.6 ± 0.5
	H1',H2' ₂	6.9 ± 0.4	H5' ₂ ,P	2.3 ± 0.5

FTIR measurements

The mid-IR absorption spectra of GG-Az1-GG were recorded on a Vertex 80 FTIR-Spectrometer (Bruker, Ettlingen) using 50 µm cuvettes built out of two CaF₂ windows. The (*E*) → (*Z*) (PSS^{UV}) and the (*Z*) → (*E*) (PSS^{vis}) conversions were accomplished by illuminating the samples directly in the spectrometer. The PSS^{vis} was reached by illumination with 420 nm (LED, ThorLabs, M420L2), while the PSS^{UV} was reached by illumination with 365 nm (LED, ThorLabs, M420L2). The IR absorption spectra were obtained by subtracting a solvent measurement from the sample measurement, while the difference spectra were obtained by direct subtraction of the corresponding sample measurements.

FTIR absorption spectra of GG-Az1-GG

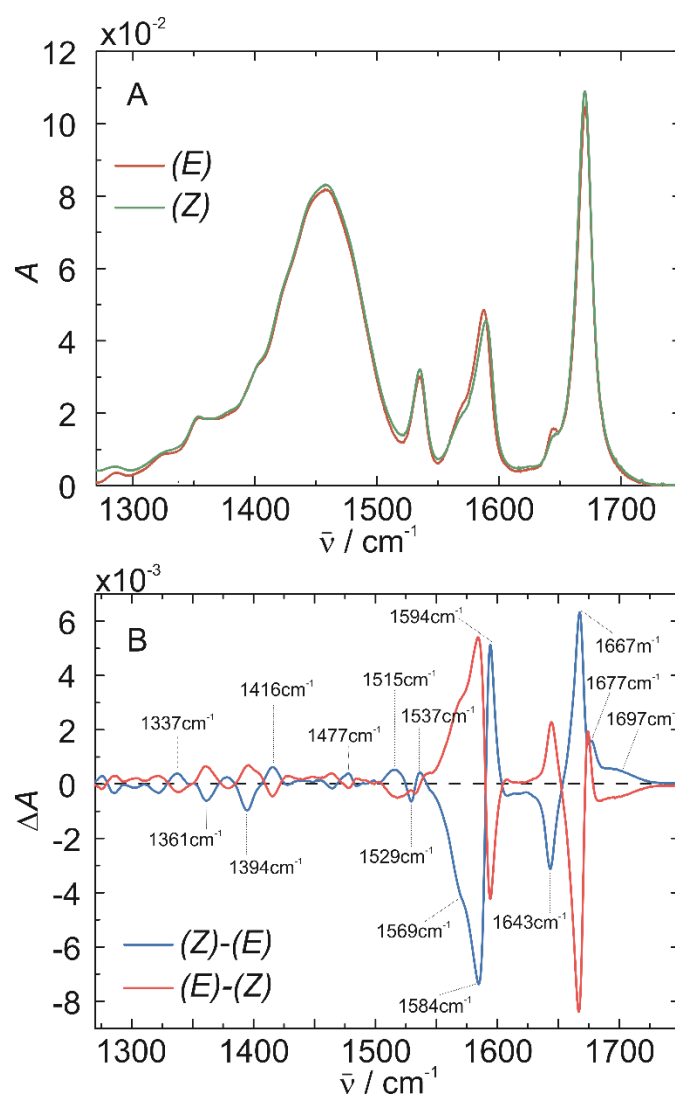


Figure S53. A) FTIR absorption spectra of GG-Az1-GG for both isomerization states of azobenzene obtained by illumination of GG-Az1-GG by 365 nm ((E)→(Z) conversion) and 420 nm ((Z)→(E) conversion); B) IR absorption difference spectra in 1300-1700 cm⁻¹ range indicating the vibrational bands affected by the isomerization of the azobenzene.

Ultrafast UV-pump/mid-IR-probe experiments

The home-built time-resolved mid-IR transient absorption set-up was described previously.^{[20][21]} It is based on a Clark (MXR-CPA-iSeries) oscillator/amplifier system that provides 150 fs pulse of 775 nm wavelength at a repetition rate of 1 kHz. The 335 nm excitation pulses were generated via sum frequency mixing of a non-collinear optical parametric amplified (NOPA)^{[22][23]} output and the laser fundamental. The pump pulses were focused into the sample cell (two CaF₂ windows separated by a 50 μm teflon spacer) to a laser spot of ~ 250 μm diameter. The IR pulses for the probe pulse were generated using a two stage optical parametric amplifier (OPA)^[24] and finally tuned to the mid-IR range (3-10 μm) by difference frequency mixing of the signal and idler in a silver thiogallate crystal (AgGaS₂).^[25] The probe pulses were focused at the sample to a spot size of 200 μm for optimal overlap with the pump pulses. The detection system is composed of 250 mm spectrometer (Sure Spectrum 250is, Chromex) and detected by a liquid nitrogen cooled MCT detector array (32 channels). The final detection resolution of the system is in the range of 3 cm^{-1} . The detector signals are sampled on a single shot basis using a multichannel integrator and analog to digital converter system. (IR-6416 Multi-Channel Laser Pulse Spectroscopy System, Infrared Systems Development). During the experiments the cuvette contain the sample was rotated and shifted sideways to provide a fresh spot for each laser excitation. In addition the sample was continuously illuminated with 420 nm light (LED, ThorLabs, M420L2) to maintain the (*E*) isomer state of the azobenzene. The experimental data was analysed by global lifetime analysis using OPTIMUS.^[26]

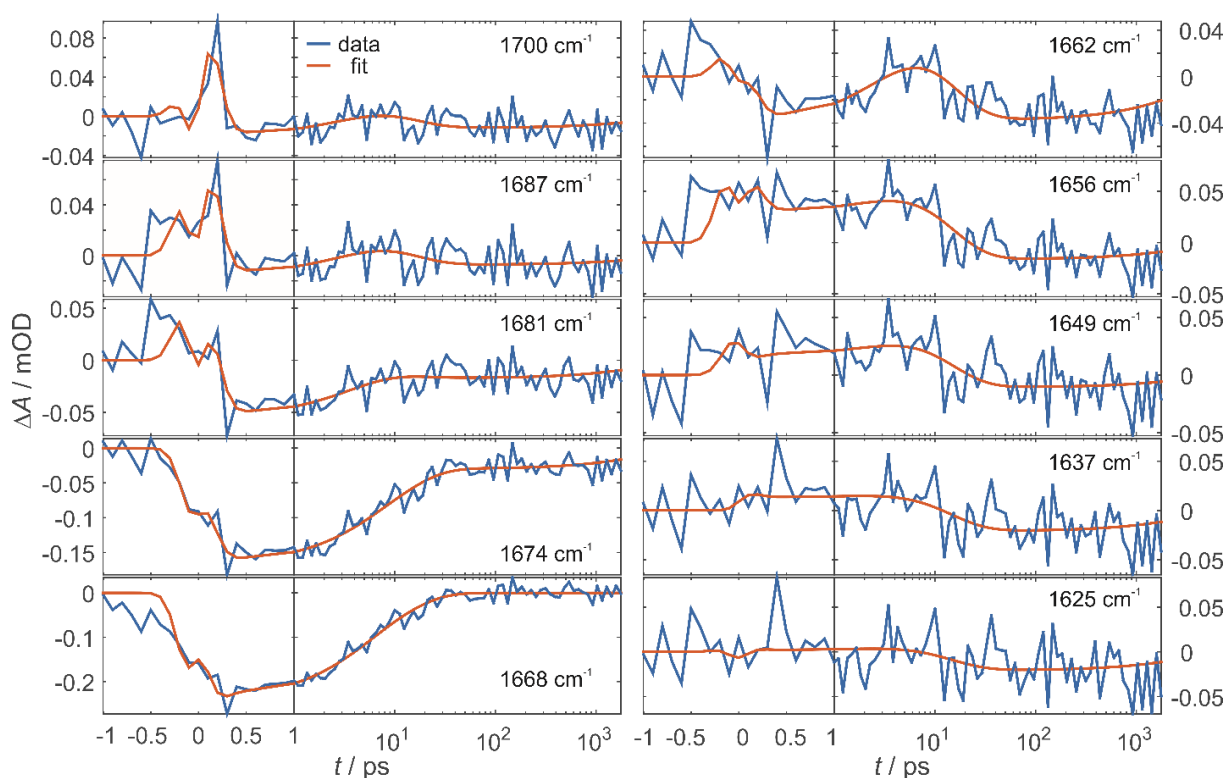


Figure S54. IR transient absorption data at different detection wavelengths demonstrating the ultrafast dynamics in the C=O stretching vibration range induced by the excitation (335 nm) of the azobenzene in GG-Az1-GG.

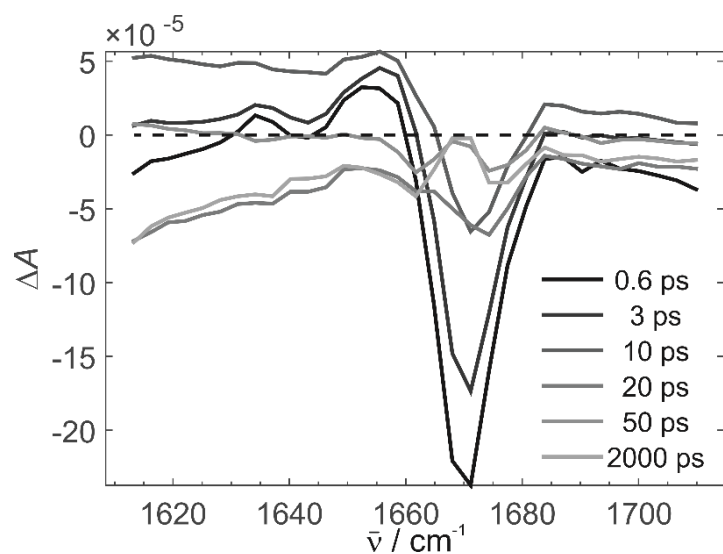


Figure S55. IR transient spectra at different delay times recorded from GG-Az1-GG after excitation of the azobenzene moiety with 335 nm laser pulses.

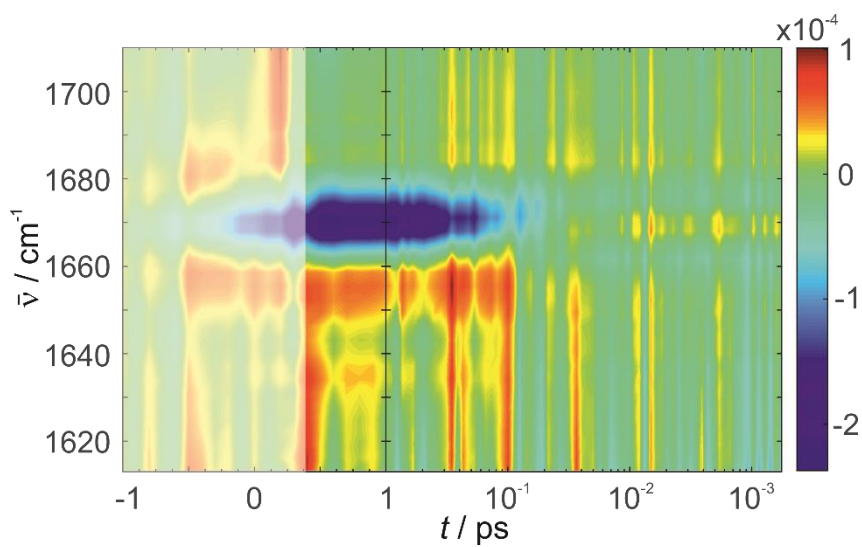


Figure S56. Overview of the IR transient absorption data recorded from GG-Az1-GG after excitation of the azobenzene moiety with 335 nm laser pulses. The transient data shows contribution of perturbed free induction decay at negative delay times and cross phase modulation around time zero, thus the analysis was restricted to delay times $t > 0.3$ ps.

Investigation of the azobenzene derivative GGG-Az1-GGG using 1D ^1H -NMR- and CD-spectroscopy and native gel electrophoresis

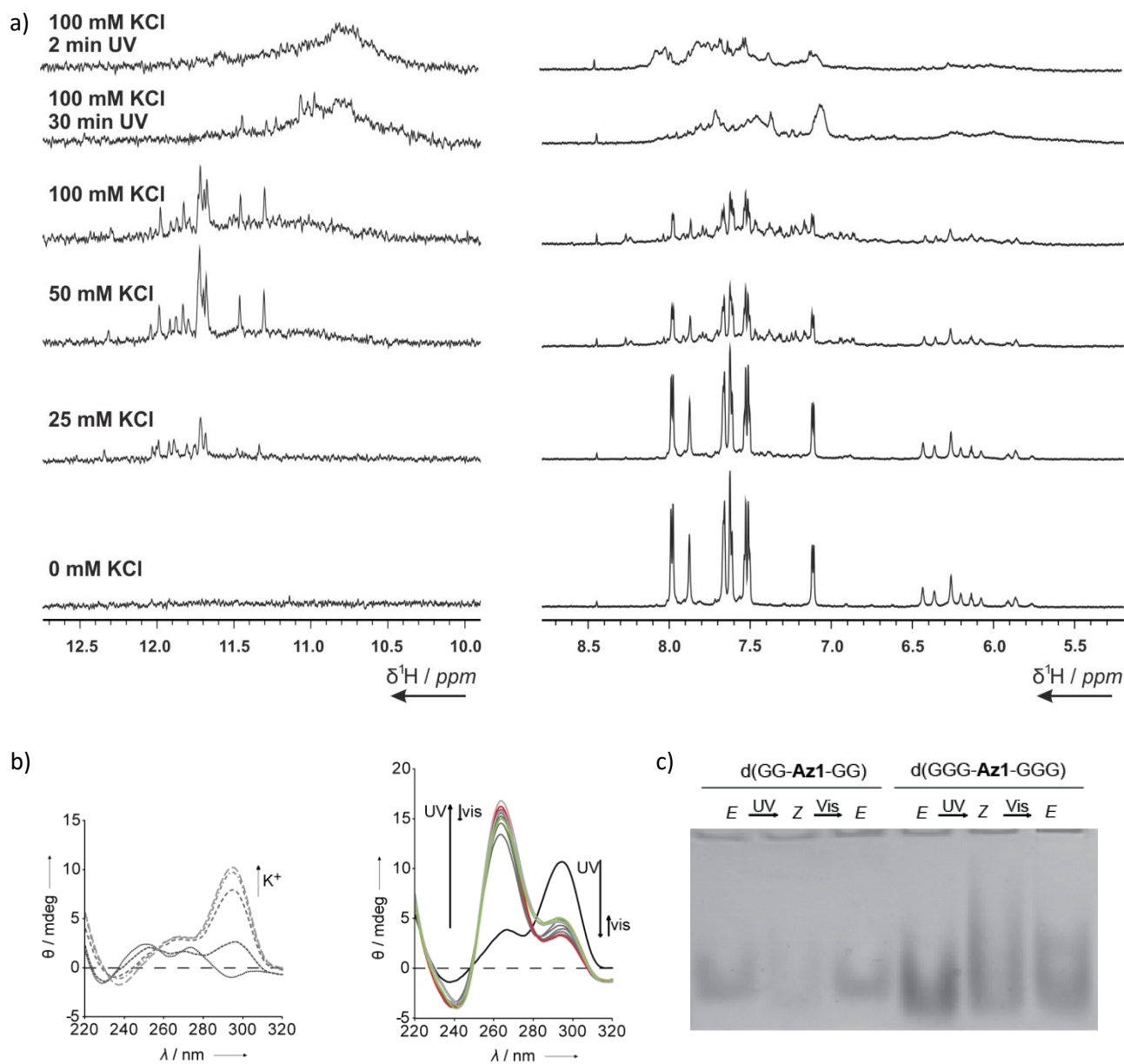


Figure S57. a) Imino (left) and aromatic (right) region of the 1D ^1H -NMR spectrum of 50 μM GGG-Az1-GGG without KCl and with 25, 50 and 100 mM KCl. At 100 mM KCl, the sample was irradiated with UV light (30 min) and afterwards with visible light (2 min). Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4), 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz. b) Series of CD spectra of GGG-Az1-GGG with increasing concentration (0, 5, 25, 50, 100 mM) of K^+ also with photoirradiation (right, 5 min UV, 2 min vis). c) Native gel analysis of GG-Az1-GG and GGG-Az1-GGG (17% polyacrylamide (acrylamide:bis-acrylamide = 29:1) gel in TBE buffer (89 mM Tris buffer, 89 mM boric acid, 2 mM EDTA, pH=8) supplemented with 100 mM KCl; the bands were visualized using StainsAll, Sigma-Aldrich, St. Louis MO, USA)

Native gel electrophoresis of GG-Az1-GG, GG-Az2-GG and GG-Az3-GG

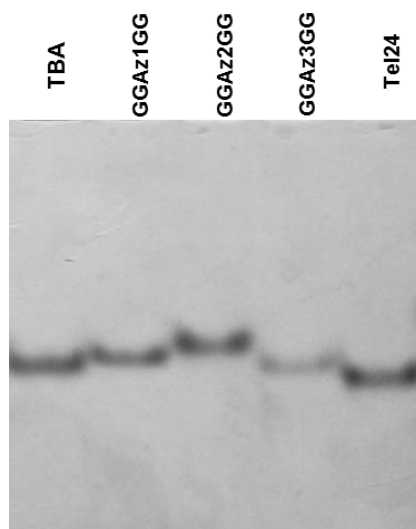


Figure S58. Native gel analysis of the indicated sequences. Molecular sizes of the azobenzene derivatives quadruplexes were characterized on a 17% polyacrylamide (Acrylamide:Bis-acrylamide = 29:1) gel in TBE buffer (89 mM Tris buffer, 89 mM boric acid, 2 mM EDTA, pH=8) supplemented with 50 mM KCl. The NMR samples (0.6 μ g) were loaded on the gel with 50% glycerol containing 1x TBE buffer supplemented with 50 mM KCl. The running buffer contained 1x TBE buffer supplemented with 50 mM KCl. The gel was run at 4°C with water cooling at 40 V for approximately 5 hours. The bands were visualized using StainsAll (Sigma-Aldrich, St. Louis MO, USA). Each lane is labeled with the name of the sequence (TBA: d(GGTTGGTGTGGTTGG), Tel24: d(TTGGGTTAGGGTTAGGGTTAGGGA)).

Structural comparison between the thrombin-binding aptamer and GG-Az1-GG

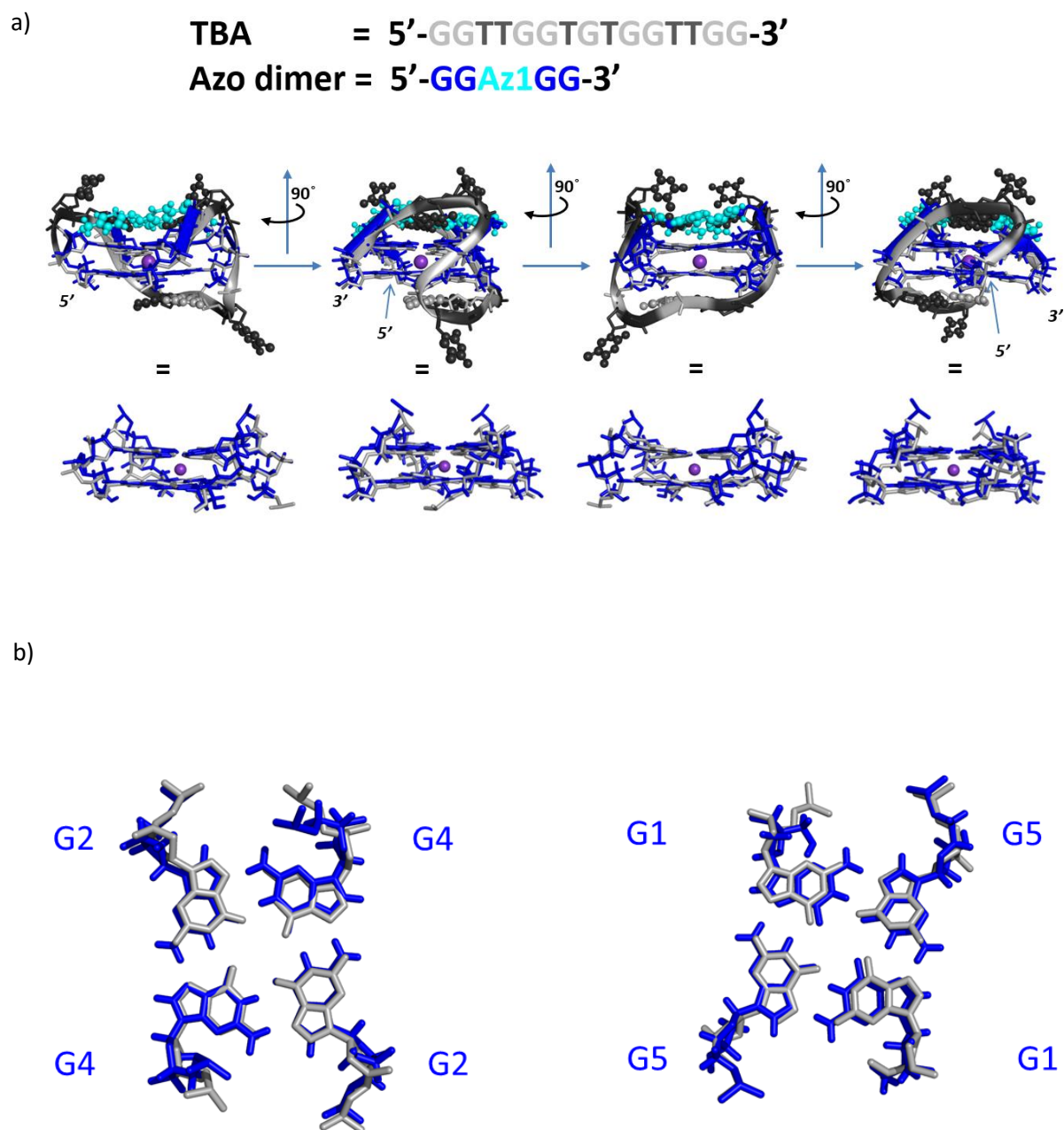


Figure S59. Structural comparison between the thrombin-binding aptamer GGTGGTGTGGTTGG (PDB: 4DII) and GG-Az1-GG.

Investigation of photo-fatigue by ^1H -NMR spectroscopy

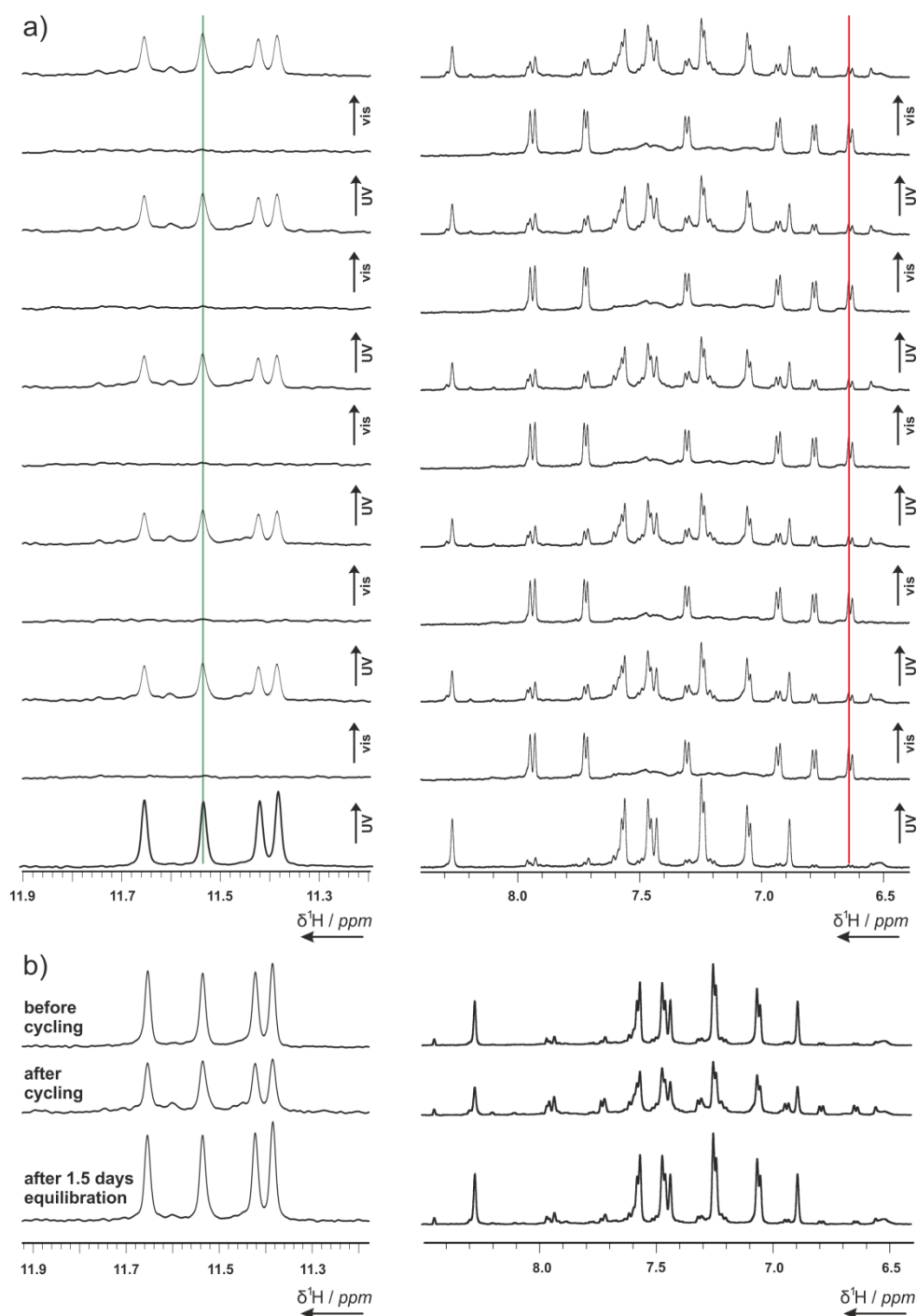


Figure S60. Cycling of photoresponsive structural conversion of GG-Az1-GG by alternate irradiation with UV and visible light monitored by NMR. a) Imino (left) and aromatic (right) region of the 1D ^1H -NMR spectrum of 125 μM GG-Az1-GG in presence of 100 mM KCl. Each cycle was performed in the following way: 1) NMR measurements (15 minutes), 2) irradiation with UV light (15 minutes), 3) NMR measurements (15 minutes), 4) irradiation with visible light (4 minutes). The signal selected for the analysis shown in Figure 3h in the main text are indicated with lines (green, 11.55 ppm, imino proton of residue G5 in *E*-state) and (red, 6.64 ppm, aromatic proton from azobenzene moiety in *Z*-conformation). Experimental conditions: 25 mM Tris-HCl buffer (pH 7.4) supplemented with 100 mM KCl, 90% $\text{H}_2\text{O}/10\%$ D_2O , 298K, 600 MHz. b) Imino (left) and aromatic (right) region of the 1D ^1H -NMR of GG-Az1-GG before the UV-visible irradiation cycles (top), at the end of the UV-visible irradiation cycles (middle) and after thermal recovery (bottom, 1.5 day after the UV-visible cycles).

References

- [1] M. Abe, A. Kubo, S. Yamamoto, Y. Hatoh, M. Murai, Y. Hattori, H. Makabe, T. Nishioka, H. Miyoshi, *Biochemistry* **2008**, *47*, 6260–6266.
- [2] T. D. Goddard and D. G. Kneller, *University of California, San Francisco*
- [3] M. Liu, X. Mao, C. Ye, H. Huang, J. K. Nicholson, J. C. Lindon, *J. Magn. Reson.* **1998**, *132*, 125–129.
- [4] V. Sklenář, A. Bax, *J. Magn. Reson.* **1987**, *74*, 469–479.
- [5] A. G. Palmer, J. Cavanagh, P. E. Wright, M. Rance, *J. Magn. Reson.* **1991**, *93*, 151–170.
- [6] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. J. Glaser, O. W. Sørensen, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 301–6.
- [7] L. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665.
- [8] A. T. Phan, *J. Biomol. NMR* **2000**, *16*, 175–178.
- [9] P. Schmieder, J. H. Ippel, H. van den Elst, G. A. van der Marel, J. H. van Boom, C. Altona, H. Kessler, *Nucleic Acids Res.* **1992**, *20*, 4747–4751.
- [10] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631–644.
- [11] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
- [12] L. Mueller, *J. Magn. Reson.* **1987**, *72*, 191–196.
- [13] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, L. J. Smith, *Biochemistry* **1999**, *38*, 16424–16431.
- [14] A. Ortega, D. Amorós, J. García de la Torre, *Biophys. J.* **2011**, *101*, 892–8.
- [15] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, et al., *Acta Crystallogr. D. Biol. Crystallogr.* **1998**, *54*, 905–21.
- [16] J. P. Linge, S. I. O'Donoghue, M. Nilges, *Methods Enzymol.* **2001**, *339*, 71–90.
- [17] H. R. A. Jonker, *ARIA 1.2 HJ (development version)* **n.d.**
- [18] S. Nozinovic, B. Fürtig, H. R. A. Jonker, C. Richter, H. Schwalbe, *Nucleic Acids Res.* **2010**, *38*, 683–94.
- [19] J. P. Linge, M. A. Williams, C. A. E. M. Spronk, A. M. J. J. Bonvin, M. Nilges, *Proteins* **2003**, *50*, 496–506.

- [20] K. Neumann, M.-K. Verhoefen, I. Weber, C. Glaubitz, J. Wachtveitl, *Biophys. J.* **2008**, *94*, 4796–807.
- [21] K. Neumann, No Title, Goethe University Frankfurt, **2010**.
- [22] T. Wilhelm, J. Piel, E. Riedle, *Opt. Lett.* **1997**, *22*, 1494.
- [23] E. Riedle, M. Beutter, S. Lochbrunner, J. Piel, S. Schenkl, S. Spörlein, W. Zinth, *Appl. Phys. B* **2000**, *71*, 457–465.
- [24] G. Cerullo, S. De Silvestri, *Rev. Sci. Instrum.* **2003**, *74*, 1.
- [25] P. Hamm, R. A. Kaindl, J. Stenger, *Opt. Lett.* **2000**, *25*, 1798–800.
- [26] C. Slavov, H. Hartmann, J. Wachtveitl, *Anal. Chem.* **2015**, *87*, 2328–36.

13. Research article V: ¹⁹F-NMR-based fragment screening for 14 different biologically active RNAs and 10 DNA and protein counter-screens

Oliver Binas*, Vanessa de Jesus*, Tom Landgraf*, Albrecht E. Völklein*, Jason Martins, Daniel Hyman, Jasleen K. Bains, Hannes Berg, Thomas Biedenbänder, Boris Fürtig, Santosh L. Gande, Anna Niesteruk, Andreas Oxenfarth, Nusrat S. Qureshi, Tatjana Schamber, Robbin Schnieders, Alix Tröster, Anna Wacker, Julia Wirmer-Bartoschek, Maria A. Wirtz Martin, Elke Stinal, Kamal Azzaoui, Christian Richter, Sridhar Sreeramulu, Marcel J. J. Blommers and Harald Schwalbe

*authors contributed equally to the work

ChemBioChem, 2020, *accepted*

Small molecule drug screening attempts are performed predominantly on proteins. However, RNA is currently emerging as an additional target, since many pathologically relevant proteins are deemed not druggable and RNA is becoming increasingly researched as a regulatory factor, especially in bacteria (see **Chapter 7.1**). In order to investigate on the druggability of 24 biological targets, including 14 RNAs, five DNAs and five proteins, a fragment-based screening relying on ¹⁹F-NMR was performed for this article. T₂-modulated ¹⁹F-1D spectroscopy was applied to fragment mixtures, monitoring signal intensities in dependence of biomolecule abundance. Targets were chosen from biomolecules currently investigated in the Schwalbe group and cover a broad range of biological functions as well as molecular sizes. For selected targets and observed binders, follow-up experiments including 2D-TOCSY, 2D-¹⁵N-correlation and competition screens were performed particularly focusing riboswitch targets. In addition, hits were analyzed bioinformatically determining hit enrichment in dependence of target and molecular scaffold. As a proof of principle, follow-up chemistry was demonstrated on a fragment which was determined to bind weakly to a 39 nt terminator RNA. Upon linking to the known RNA binder acridine, low micromolar affinity was reached.

The author of the thesis contributed to the experiments with production of one screened RNA and subsequent fragment-based screening, establishing the approach in the group under supervision of Dr. C. Richter and Dr. S. Sreeramulu. All follow-up experiments on the ZMP-sensing riboswitch were performed and analyzed by the author. Additionally, the author contributed largely to development and assembly of the necessary toolset for screening data analyzation and distribution of the results to the large group of investigators. The author was responsible for data consolidation and presentation and contributed significant parts of the manuscript as well as most graphic material. Cheminformatics were carried out by Dr. M.J.J. Blommers and Dr. K. Azzaoui. Follow-up chemistry was performed by J. Martins, A. Tröster and Dr. N. S. Qureshi.

¹⁹F-NMR-based fragment screening for 14 different biological-ly active RNAs and 10 DNA and protein counter-screens

Oliver Binas,^{‡[a]} Vanessa de Jesus,^{‡[a]} Tom Landgraf,^{‡[a]} Albrecht Eduard Völklein,^{‡[a]} Jason Martins,^[a] Daniel Hymon,^[a] Jasleen Kaur Bains,^[a] Hannes Berg,^[a] Thomas Biedenbänder,^[a] Boris Fürtig,^[a] Santosh Lakshmi Gande,^[a] Anna Niesteruk,^[a] Andreas Oxenfarth,^[a] Nusrat Shahin Qureshi,^[a] Tatjana Schamber,^[a] Robbin Schnieders,^[a] Alix Tröster,^[a] Anna Wacker,^[a] Julia Wirmmer-Bartoschek,^[a] Maria Alexandra Wirtz Martin,^[a] Elke Stinal,^[a] Kamal Azzaoui,^[b] Christian Richter,^{*[a]} Sridhar Sreeramulu,^{*[a]} Marcel Jules José Blommers,^{[b]*} Harald Schwalbe,^{*[a]}

‡ Oliver Binas, Vanessa de Jesus, Tom Landgraf, Albrecht Eduard Völklein contributed equally to the work.

[a] Oliver Binas, Vanessa de Jesus, Tom Landgraf, Albrecht Eduard Völklein, Jason Martins, Daniel Hymon, Jasleen Kaur Bains, Hannes Berg, Thomas Biedenbänder, Dr. Boris Fürtig, Dr. Santosh Lakshmi Gande, Dr. Anna Niesteruk, Andreas Oxenfarth, Dr. Nusrat Shahin Qureshi, Tatjana Schamber, Robbin Schnieders, Alix Tröster, Dr. Anna Wacker, Dr. Julia Wirmmer-Bartoschek, Maria Alexandra Wirtz Martin, Elke Stinal, Dr. Christian Richter, Dr. Sridhar Sreeramulu, Prof. Dr. Harald Schwalbe
Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ),
Johann Wolfgang Goethe-University Frankfurt
Max-von-Laue Straße 7, 60438 Frankfurt am Main, Germany
E-Mail: schwalbe@nmr.uni-frankfurt.de, ric@nmr.uni-frankfurt.de, sridhar@nmr.uni-frankfurt.de

[b] Dr. Marcel Jules José Blommers, Dr. Kamal Azzaoui
Saverna Therapeutics, Gewerbestrasse 24, 4123 Allschwil, Switzerland
E-Mail: mjj.blommers@saverna.com

Abstract: We report here on the nuclear magnetic resonance (NMR) ¹⁹F screening of 14 RNA targets with different secondary and tertiary structure to systematically assess druggability of RNAs. Our RNA targets include representative bacterial riboswitches that naturally bind with nanomolar affinity and high specificity to cellular metabolites of low molecular weight. Based on counter-screens against five DNAs and five proteins, we can show that RNA can be specifically targeted. To demonstrate the quality of the initial fragment library that has been designed for easy follow-up chemistry, we further show how to increase binding affinity from an initial fragment hit by chemistry that links the identified fragment to the intercalator acridine. Thus, we achieve low micromolar binding affinity without losing binding specificity between two different terminator structures.

Introduction

Proteins constitute the vast majority of validated drug targets. The ribosome, a large RNA-protein-complex, is the most prominent RNA drug target. Most antibiotics inhibit protein synthesis by targeting the interface of RNA and proteins in the ribosome^[1]. Beyond being target for antibiotics, RNA has for long been considered undruggable. Recently, however, this view has changed and RNA emerged as a potential target for drug discovery as well^[2-6]. Clinical success of compounds initially identified as RNA binders^[7] inspires thorough exploration of the non-coding RNA target space. Here, potential targets range from RNA involved in oncology and inflammation to RNA involved in bacterial and viral infections, to mention a few areas with unmet medical need. Concurrently, the continuous identification of

new regulatory RNAs, including riboswitches or sRNAs further increases potential applications^[8]. Riboswitches in particular have come into focus to combat multi drug resistant (MDR) bacteria that pose a major health threat for modern human society^[9], which can only be tackled by new antibiotics. The development of drugs targeting riboswitches is therefore an important research focus^[10].

Structure-based drug discovery is a key methodology for rational drug discovery. Here, X-ray crystallography^[11] and nuclear magnetic resonance (NMR) spectroscopy^[12] provide the essential structural information. Insight from target structures is often supported by computational methods to aid in library design. Virtual screening by *in silico* docking of a compound library against available target model structures guides medicinal chemistry in the development from initial hits towards the generation of lead compounds^[13].

High-throughput screening in drug discovery requires robust detection of binding of a large number of test compounds, a library, to a biological target. Such screening can involve up to 1-2 million compounds from which routinely a very small number of potential lead compounds are identified. Screening large libraries thus requires high preparative and infrastructural effort.

An alternative to this classic approach is fragment-based screening. Fragments are often weak binders and their binding specificity can be lower than expected from a lead compound^[14]. Thus, fragment-based drug discovery requires that the initial hits are further processed into lead compounds by chemical modification such as growing the compound to fit the desired number of binding interactions with the target or linkage of two fragments which bind to binding sites in spatial proximity. Fragment-based drug discovery is nowadays the basis of many hit-to-lead re-

search programs^[15]. In fact, it has been shown that FDA-approved drugs targeting proteins can be developed starting from fragment screens^[15–21].

Methods used to screen RNA include fluorescent-based assays^[22,23], mass spectrometry^[24,25], small molecule microarrays (SMM)^[26,27], microscale thermophoresis (MST)^[28,29] and NMR spectroscopy^[30]. However, most of these studies focus only on a single target RNA^[31,32]. By contrast, we here report on the screening of a library of 102 fragments against 14 different RNAs of different sizes and different architectures, using ¹⁹F-NMR as the main method for hit identification. The targeted RNAs include small stem loop structures, aptamer domains of riboswitches, full-length riboswitches, terminators and antiterminators of riboswitches, ribozymes as well as tRNAs, traditionally serving as control RNAs in screening (Figure 1). Thus, our results allow us to delineate specificities of fragments towards different RNAs. We further counter-screened against five DNAs and five proteins to test whether the fragment library can target for these different classes of biomacromolecules or whether the library is biased towards binding a subset of biomacromolecules. The DNA targets include regular double-stranded DNA as well as G-quadruplex structures of different morphology, and the proteins include RNA-binding proteins as well as the important enzyme classes kinases and phosphatases that bind to phosphorylated moieties as part of the enzymatic function. With our multi-target approach, we show that selective fragments for an RNA target can be found but often a good hit is broadly binding to RNAs of the same size and structural complexity. Also we show that the se-

lective targeting of RNA over other classes of biomacromolecules is possible with this library.

By way of example, we further show that by fast follow-up chemistry that involves the linkage of an RNA-binding fragment with the intercalator acridine, we obtain low micromolar RNA binders with more than 10-fold specificity towards different RNAs.

Results and Discussion

Target choice is an important step for any screening especially in a multi-target approach in which a broad spectrum of biologically relevant target molecules is crucial to success. Therefore, we showcase our measures of target choice in the following section. All RNA constructs screened are summarized in Table 1.

RNA hairpin structures

Stem-loop/hairpin structures represent the most common small secondary structure motifs in RNA^[33]. Common loop lengths range from three to seven nucleotides, but more than 50% of all loops are tetraloops^[34]. Tetraloops are not only very abundant, they also exhibit a high thermodynamic stability as they are usually stabilized by hydrogen bonding and stacking interactions. When drug-screening approaches are employed on biologically relevant stem-loops, further characterization of the binding mode is needed to distinguish stem-binding^[35] from loop-binding ligands^[27] to understand the structural basis of the ligand-induced change in biological function^[36].

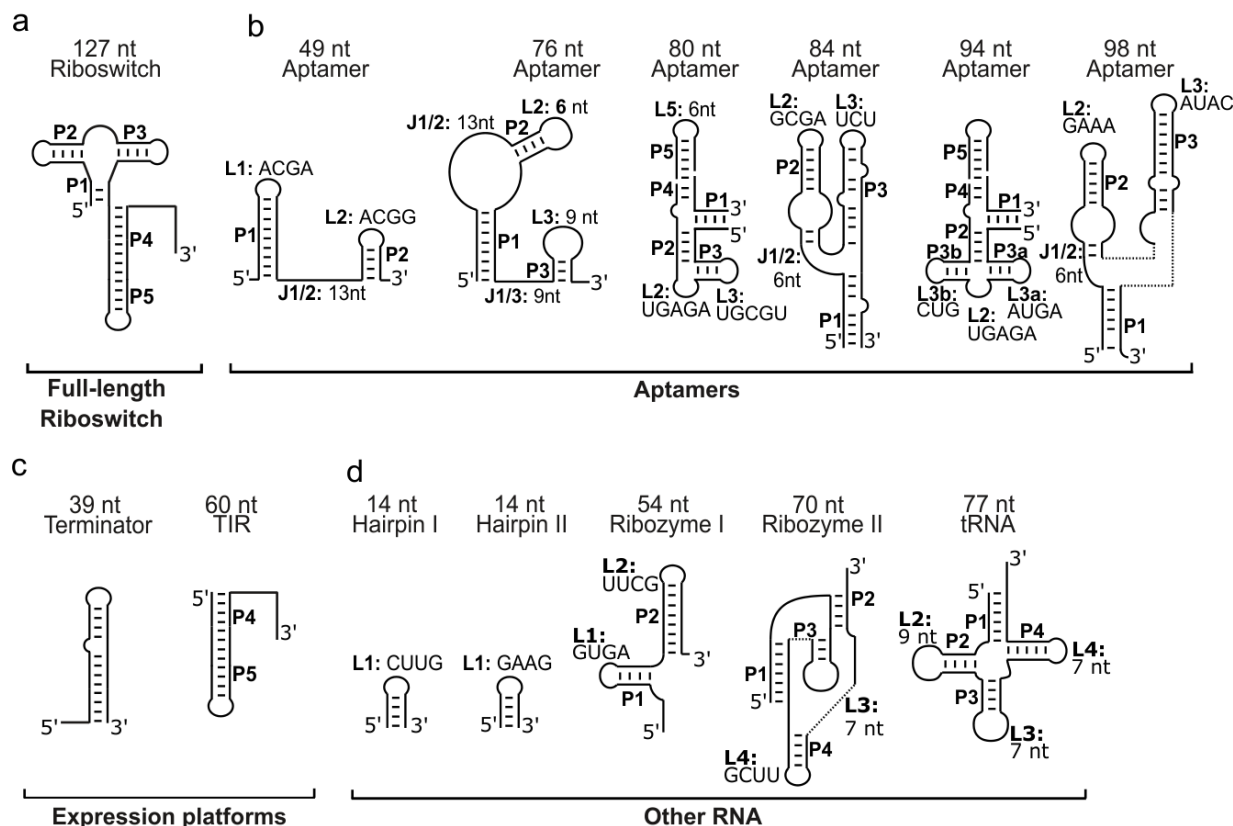


Figure 1. Overview of RNA targets. Schematic secondary structures of the RNA targets investigated by ^{19}F -FBS. Stems (P), loops (L) and junctions (J) are annotated, respectively. Tri-, tetra- and pentaloop sequences are listed explicitly.

Rational ligand design led to the development of compounds that specifically bind to an RNA loop region^[37]. In our study, we included two 14 nucleotides (nt)-containing stem-loop structures exhibiting a GAAG and a CUUG tetraloop, respectively in order to detect fragments binding to this abundant secondary structure motif. Also, we included the guanidine-sensing riboswitch as an example of a functional RNA with hairpin structures. Loop-loop-interactions^[38] are part of the stabilizing function with purine-sensing riboswitches that are part of the RNA targets in this study (Supplementary Figure 3).

RNA bulges, internal loops and pseudoknots

Helix-junction-helix (HJH) structure elements occur between two helices or three- and four-way junctions. They can be divided into bulges and internal loops, where the first is characterized by short single-stranded intersections on one side of an RNA stem and the latter features unpaired regions on both sides of the stem^[39,40]. Bulges and internal loops constitute conformational hinges, allowing helices to adopt different conformations with respect to each other. Formation of these structures allows for inter helix motions such as dynamic nucleobase stacking or rotation in and out of a junction.^[39] They can often be targeted by low molecular weight ligands as previously shown for the Tat-TAR interaction, which was mimicked by arginamide^[41,42]. In drug screening approaches, internal

loops and bulges can be valuable targets for ligand design since their potential for ligand recognition is relatively high. Examples of virtual screenings directed on HIV-1 TAR show that sampling of the entirety of the allowed topological space leads to an ensemble of discrete conformations that can bind different ligands^[36,43,44].

In our screening pool, several examples of small and large bulge regions, internal loops and pseudoknots are present in RNAs (Figure 1).

Riboswitches

Riboswitches are structured RNA elements which regulate gene expression by allosteric structural re-arrangements of an expression platform element in response to sensing environmental changes by an aptamer element. Most riboswitches respond to changes in concentration of small molecules, mostly metabolites, which they bind with remarkable specificity^[45,46]. Examples showing the observation of binding via homonuclear and heteronuclear 2D-NMR spectroscopy are displayed in Supplementary Figures 2 and 3. Intricate tertiary structures are formed by most aptamers to achieve high-affinity binding required for optimal sensitivity, alongside with sufficient discrimination against non-cognate ligands. These binding pockets feature a closely defined chemical space, which is usually thoroughly described by structural data (see Supplementary Figure 1). The complex and specific chemical envi-

ronments aid development of specific ligands and therefore especially fragment-based drug discovery approaches, which rely on chemical adaption in particular. Thus, it is not surprising that riboswitch aptamer domains have been identified as excellent drug targets very early on^[47,48]. Of the 14 RNAs screened, eight are derived from riboswitches including natural ligand binding aptamer domains (*vide infra*).

In this study, we screened the aptamer domains of riboswitches from the second-messenger-sensing class, the guanidinium-sensing class, the purine-sensing class and the thiamin-pyrophosphate-(TPP)-sensing riboswitch. The TPP-sensing riboswitch represents the most abundant riboswitch found in different prokaryotes and even eukary-

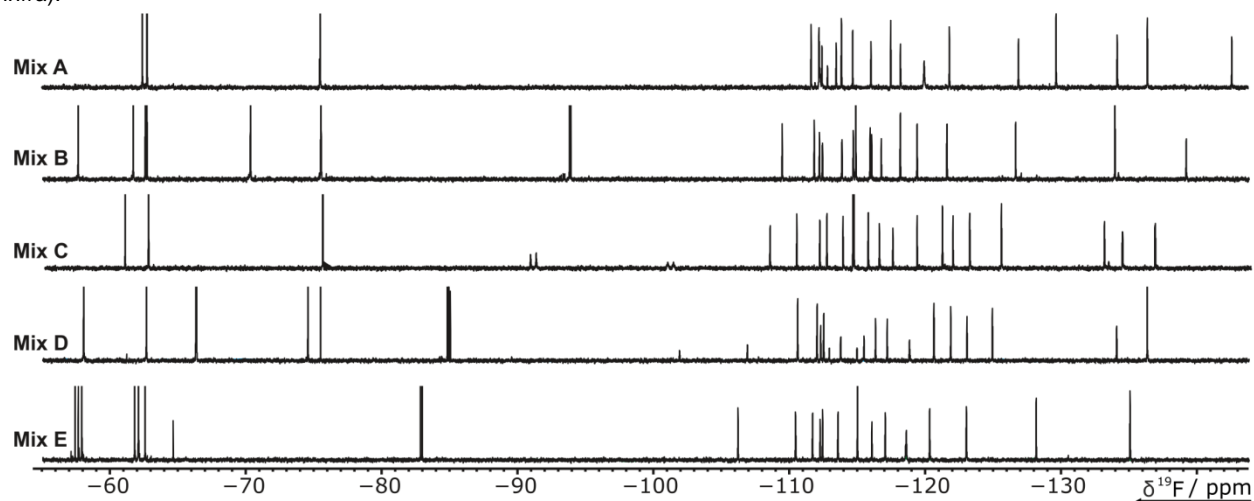


Figure 2. ¹⁹F-1D NMR-spectra of the ¹⁹F-library fragment mixtures. The ¹⁹F-library contains 101 compounds (Supplementary Table 1). Five mixtures of either 20 or 21 ligands were generated to avoid signal overlap. The spectra of the mixtures (A, B, C, D, E) in the screening buffer are displayed.

otes (Table 1). For purine-sensing riboswitches, operating either on the transcriptional or the translational level, we have previously reported on the mechanism of full-length riboswitch function^[49,50].

Counter screens

To maximize the coverage of conformational space and to rule out unspecific binding we added five other RNAs ranging from 14 nt to 77 nt in length to the pool of target RNAs. We screened tRNA^{fMet} produced in house. tRNAs, being ubiquitously present in all kingdoms of life, are used as counter screen RNA in many applications and especially in high-throughput screens of RNA molecules, such as an RNA G-quadruplex^[51] and the transactivation response element (TAR) RNA^[52]. To rule out binders not specific to RNA, we additionally screened five DNAs (Supplementary Figure 24 – 31), including four G-quadruplexes and five proteins with molecular weights ranging from 18 to 100 kDa (Supplementary Figure 32 – 39).

¹⁹F-CPMG based screening by NMR spectroscopy

NMR spectroscopy is well suited for the identification of initial fragment hits as it is fast and reliable and offers the possibility to detect weak binding in solution. There are a large number of NMR experiments to detect binding, including NOEs, chemical shift perturbation, saturation transfer difference^[53], WaterLOGSY^[54] or T₂-relaxation spectroscopy^[55]. By these methods, interactions characterized by dissociation constants in the range of 10 mM down to

low nM can be detected, depending on experimental setup. Further improvements are currently developed with more sophisticated methods of dynamic nuclear polarization (DNP) or hyperpolarization^[56], decreasing the lower detection limit. Additionally, NMR offers the possibility to observe the interaction of the target with fragments in a mix of several fragments at the same time, greatly reducing operational effort. Often, NMR screenings utilize ¹H-detection. The high number of hydrogen atoms in fragment compounds, however, leads to severe overlap of NMR signals, reducing the number of fragments within one mixture that can be screened in a single experiment.

¹⁹F-detection^[57] is an attractive alternative to ¹H detection. NMR signals in ¹⁹F-spectra show a much higher chemical shift dispersion covering a range of around 50 kHz (83 ppm) compared to around 6 kHz (10 ppm) for protons. Furthermore, if ¹H-decoupling can be performed, which depends on the spectrometer configurations and NMR probehead used, each ¹⁹F resonates at a single resonance frequency, allowing the observation of several fragments in a single mixture. Figure 2 shows the design of the fragment mixture containing 20 or 21 different ligands per mixture. An overview of all 101 fragments screened is available in Supplementary Table 1. In this study, we measured ¹⁹F transverse relaxation experiments which apply CPMG pulse trains^[55,58] for varying relaxation delays. CPMG T₂ measurements exploit the different relaxation properties of unbound fragments in comparison to (transiently) bound fragments to biological targets. Low molecular

weight fragments with short rotational correlation times (τ_c) in solution will show changes in CPMG T_2 values upon (transient) binding to a high-molecular weight macromolecule (4 kDa - 100 kDa), which exhibits much slower τ_c and consequently faster T_2 -relaxation. This relaxation effect is observed even if the population of bound fragment is 1% or lower. We chose this method since it is very sensitive to even low affinity interactions and is therefore beneficial to fragment-based approaches.

Validation of hits from ^{19}F screens

After this initial broad screening follow-up screens integrate cheminformatic-based searches for similar ligands that are commercially available but also cross-validation of binding to other targets. In fact some of the fluorine containing ligands show binding to almost all RNAs (#s of the unspecific binder/s). Screening of fragments against RNA targets yielded several hits with a hit rate of up to 26%. These hits were roughly estimated to bind with an at least low millimolar K_D or tighter to the target RNA. This estimation stems from the observation that most RNA signals did not show large chemical shifts changes in the follow up 2D ^1H - ^{15}N -correlation experiments, upon interaction addition of fragments. This assumption is made, although it is not clear whether changes in chemical shifts are strictly correlated with changes in the chemical environment. After the initial ^{19}F screening the hits can be confirmed, K_D values determined and information on the fragment binding site obtained. To confirm hits, we prepared samples containing only the fragment that showed binding in initial screenings mixtures. We were able to observe altered T_2 -relaxation behavior for all hits investigated, confirming their RNA binding capability and justifying the pooled-mix screening approach (Supplementary Figure 5 – 39).

Mapping binding to a specific site in RNAs is usually performed by analysis of chemical shift perturbations (CSPs) that fragment binding causes on the RNA resonances in proximity to the binding epitope. In this approach, care has to be taken to distinguish direct binding induced CSPs from remote effects whose origins are sometimes difficult to assess. Most of the RNAs studied are riboswitches, which bind to metabolites of low molecular weight with an affinity several orders of magnitude higher than the expected affinity of the fluorinated fragments. Thus, orthosterically binding fragments can be detected in a competition experiment. By adding the natural ligand to an RNA sample containing the hit fragment, the natural ligand will compete for the RNA binding site and eventually displace the lower-affinity binding fragment. Accordingly, if a CPMG-experiment with long mixing time is recorded, fragment signals will be recovered upon addition of the natural ligand. Although these experiments can provide evidence for ligands that interact with the natural ligand binding site, this might not always be the case since structural rearrangement upon binding of the natural ligand can also obscure other binding sites, which were formerly accessible to fragments. To evaluate the T_2 -modulated 1D-NMR spectra,



Figure 3. Interaction table of all fragments and biological targets screened. Hits were classified into no binding ($Q_{bind} > 0.67$, alternating grey and white), weak ($0.66 < Q_{bind} < 0.33$, yellow) or strong binding ($Q_{bind} < 0.32$, green) in 19F-CPMG experiments. For protein screens, hits for ~5% of the ligands could not unambiguously be assigned (light blue).

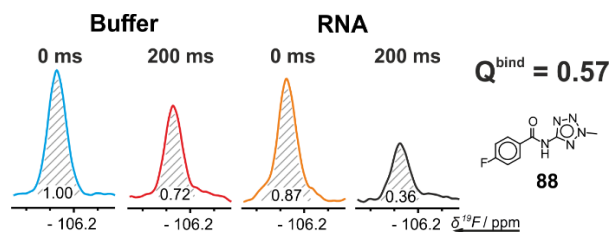


Figure 4. Determination of Q^{bind} . Four ^{19}F -CPMG experiments are recorded to determine the binding factor Q^{bind} from peak integrals as discussed in the main text. The relaxation loss at 200 ms relaxation dephasing time relative to 0 ms dephasing for the ^{19}F signal of the ligand is recorded in the presence and absence of biomolecular target.

we measured the fragment signal integrals and calculated the ratios between 200 ms CPMG and 0 ms CPMG applied to ^{19}F -NMR spectra.

The quotient Q^{bind} of the integral ratios

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

with

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

defines a quantifiable factor to classify the ligand-target interaction into no binding, strong or weak binding (Figure 4). The exact effect, however, depends on the overall rotational tumbling time τ_c that increases with increasing molecular weight. Q^{bind} is a rough criterion, since we did not differentiate between aromatic and aliphatic bound fluorines, which we however deem sufficient for the task. The quotient ($\text{Peak Integral}^{+\text{Target}} / \text{Peak Integral}^{-\text{Target}}$) was automatically calculated; all potential hits were then manually checked. The results are summarized in Figure 3 and spectra regions of all hits are displayed in Supplementary Figures 5 – 39. We obtained hits for all targets. The results show a clear trend that riboswitch RNAs show a high hit rate, ranging from 7 to 26 hits per riboswitch. Riboswitches contain aptamer domains that bind metabolites in the same size range as the fragments or even lower (e.g. F^- -sensing^[59] or Mg^{2+} -sensing^[60] riboswitches). Only one to five hits could be determined for all other RNAs. Although CPMG measurements as relaxation-based experiments are biased towards increased sensitivity for larger constructs, increased affinities of riboswitch targets are still striking in comparison to the 77 nt tRNA.

DNA structures which were included to sample other nucleic acid structures, namely duplex and G-quadruplex, showed very different behavior. For the duplex, only a single fragment showed binding. For G-quadruplexes, we observed between 12 and 20 hits with some overlap to hits that also bind to riboswitch RNAs. Of the five proteins investigated,

four showed a large number of hits, ranging from 16 up to 55 (Table 1).

The 18 kDa phosphatase MtpA showed only four hits, in line with the difficult druggability of phosphatases. In general, for 101 fragments screened across 24 different biomolecular targets involving either DNA/RNA/proteins, approximately 5% of the data were not analyzable. This is a result of the necessity to optimize buffer conditions in particular for proteins. The different buffer conditions can lead to solubility issues and chemical stability issues for this subset of ligands. The remaining fragments show a broad variety of target selectivity from fragments binding exclusively a single target (fragment 100) to highly promiscuous binding behavior (fragment 57).

From the pool of the screened biological targets, we chose the aptamer domains containing 76 nt, 84 nt and 98 nt of the secondary-messenger-sensing riboswitches for follow-up investigation. Weak hits were omitted. To verify hits and rule out effects of fragment mixing we confirmed binding of hits for single fragments. For all investigated fragment and RNA combinations, we were able to observe the same strong effects as in the mixtures. To validate the observed effects and to characterize the binding epitope, we analyzed the effect of fragment 75 addition to the 76 nt riboswitch (Figure 5). We used ^{15}N -isotopically labeled RNA to conduct ^{15}N -correlated 2D spectroscopy on imino hydrogens (Figure 5a) to detect possible chemical shift perturbation of RNA signals introduced by the addition of the screening hit. In these spectra, only helical imino hydrogen signals are visible, which shift distinctly in case of helix groove binding fragments. In our case, we could only observe very small shifts and sometimes additional small signals in the spectra. More pronounced effects could be observed on the signals of aromatic hydrogens in $^1\text{H},^1\text{H}$ -TOCSY spectra (Figure 5b). Here, clear signal shifts over 10 Hz on H5-H6 cross peaks of pyrimidine residues were detected.

At concentrations appropriate for screening (in this case 50 μM), and for large RNAs, such as the riboswitches investigated, only the strongest H5-H6 peaks are visible. In the example shown for compound 75 (Figure 5b), dose-dependent chemical shift perturbation was detected for three signals. Combined with the data obtained from ^{15}N -correlation spectroscopy, which showed only minor alteration, we can conclude that the respective binding site is located in a flexible region of the RNA. A NMR-titration experiment analyzing the ^1H -1D-chemical shift perturbations showed a 400 μM K_D for a mixture of fragment 75 and the 76 nt secondary-messenger-sensing riboswitch (Figure 6b)^[61].

Additional information on the binding site was obtained from a competitive binding assay. We added the fragment under investigation to the RNA target and characterized the effect of addition of native ligand on the T_2 -modulated signal. As observed earlier, the signal is completely suppressed upon addition of the RNA, but addition of native ligand leads to signal recovery by approximately 15%. The incomplete recovery of fragment signals points to the possibility that orthosteric binding of the fragment hit to the binding site of the

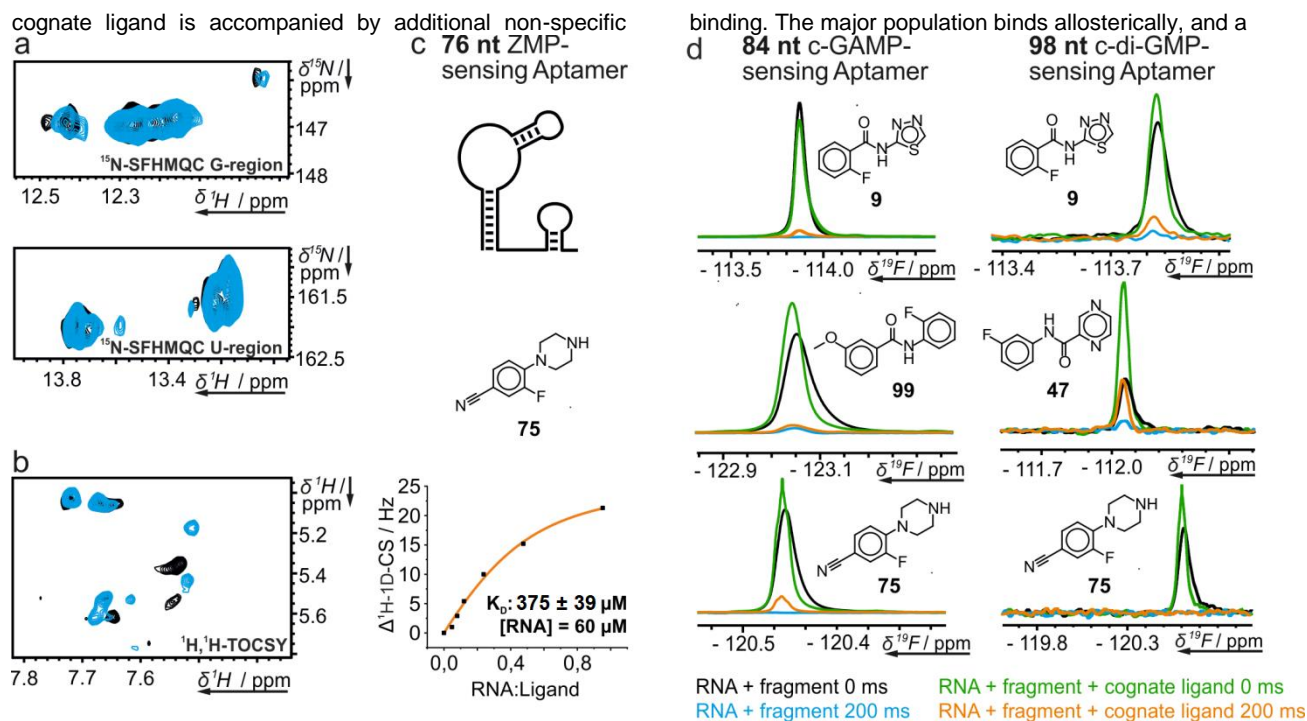


Figure 5. Hit validation and competition experiments. Validation of ^{19}F -CPMG screening hits for the aptamer domains of the three secondary messenger-sensing riboswitches. a) Spectral regions with signals from guanosine (top) and uridine (bottom) residues of the ^1H , ^{15}N -correlation experiment of the 76 nt riboswitch with (blue) and without 75 (black) shown under c. b) ^1H , ^1H -TOCSY spectrum with (blue) and without 75 (black). c) ^{19}F -1D-NMR-titration of 75 with the RNA. K_D was determined according to Williamson.^[61] d) (Partially) competitive binding of fragments to the 84 nt and 98 nt riboswitch observed in T_2 -modulated 1D- ^1H experiments.

smaller population binds orthosterically to the same binding site as cognate ligand. We found that signals of other compounds could be recovered by 83% (47 Figure 5d) after addition of native ligand, pointing to a larger influence on the binding site. The most convenient way to obtain affinity data by NMR is the observation of the ^{19}F fragment signal, since it does not require isotopic enrichment and there is only one signal. ^{19}F signals in general are very sensitive to changes in their chemical environment and thus, presumably, show also a substantial CSP for the ^{19}F upon addition of ligand. In contrast, the chemical shift dispersion of the aromatic hydrogens is smaller and the largest CSP in ^1H RNA signals was only around 5 - 8 Hz. From ^{19}F CSP data we could obtain affinity constants in the high μM range, such as 400 μM for fragment 75^[62].

The screening of a large number of biomolecular targets demonstrates that the fragment library exhibits the highest hit rate to proteins, followed by RNA and then DNA. RNA hits are observed predominantly for RNAs containing loop regions, bulges, and internal loops. Some of the fragments in the library are promiscuously binding to all three different classes of biomolecular targets. The overlap for hits binding to proteins and RNA is also around 20%. The most promising result, however, is the observation that each biomolecular target class can be specifically targeted (See Supplementary Figure 4).

Cheminformatic analysis of hit data

The 69 hit compounds were chemically clustered using Hierarchical Clustering (DistMatrix, Morgan fingerprint, distance threshold 0.6, using Knime software 4.0.2) resulting in 38 singletons and 4 chemical families sharing a closely related scaffold. The largest cluster contains 5 members that were binders for proteins, DNA/RNA, and DNA/RNA/proteins targets. No cluster seems to be specific to any of the target families screened.

In order to check if there are any correlations between chemical structures and the number of targets that bind to it, we generated a number of molecular descriptors linked to the shape, electrostatic and hydrophobic interactions. The correlation matrix of all data shows no significant correlations between the number of target hits and molecular descriptors. The lowest and the highest correlations found are respectively with the number of aromatic atoms ($R = +0.27$) and SP3 descriptor (sp3 carbon atom count/total carbon atom count reflecting the flatness of the molecules) ($R = -0.23$).

The statistical analysis of the number of aromatic atoms for each category of binders shows higher average value of this descriptor for compounds hitting DNA/RNA/proteins (9.2) after the DNA/RNA binders (10.4) but in this last case the number of hits is too small to draw a conclusion. The SP3,

like the other molecular descriptors, shows no significant difference between the category for hits (Figure 6). The

substructure counting of popular and frequent motifs in

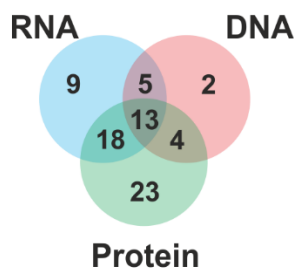
a Gaussian distribution for aromatic atoms count per class value

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

Gaussian distribution for SP3 descriptor per class value

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

b



c ■ Riboswitch ■ Aptamer ■ Expression platform ■ Other RNA ■ DNA ■ Protein

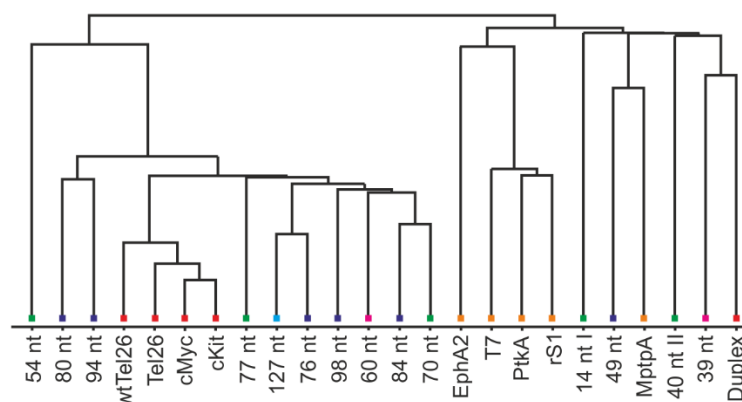


Figure 6. Cheminformatic analysis of hit data for all RNA, DNA and Protein biomolecules. a) Gaussian distributions for aromatic atoms and SP3 descriptor over categories of biomolecules. SP3 descriptor (sp3 carbon atom count/total carbon atom count) reflects the flatness of the fragment molecules. b) Visualization of categories in a Venn diagram. c) Euclidian distribution of hits to the target biomolecules.

organic molecules reported in the Table 6a, also shows no significant enrichment in different categories of binders. Because of the small size of the fragments in the current ^{19}F -library, there is no privileged class of compounds or relevant physicochemical properties that can be specific to a family of biological targets (RNA/DNA/proteins). The ^{19}F -fragment library is therefore suitable for RNA, DNA, and proteins and can be used to generate starting points for follow-up screens and to examine the presence of potential binding pockets for ligands in the individual targets. Moreover, correlation analysis (Figure 6c and Supplementary Figure 4) shows striking clustering of hits between Riboswitches and Aptamers, DNA and Proteins respectively. Currently research focuses on the development of libraries solely suited for RNA, taking the repetitive nature of the RNA backbone, the higher charges and RNA dynamics into account.^[63,64] While the here shown library of ^{19}F -fragments was shown to be suited for proteins and RNA likewise, these recent developments could be

taken into account in order to enrich current fragment libraries into a more RNA-focused library.

Follow-up chemistry

After hit identification, the next major aspect in fragment-based drug discovery is to link initial hits to increase binding affinity. Such an improvement can be achieved by linkage of the fragment to an RNA binder, as shown for a Neomycin-Acridine conjugate.^[65] To generate affinity we used an acridine moiety, which is a well-known intercalator that also allows fast follow-up chemistry.^[66,67] Herein we describe a three step synthesis to generate an acridine moiety followed by linking it to a fragment. We chose a fragment from an additional 1H-Screening for the 39 nt terminator stem (P2D11) with a similar structure to the Hits from ^{19}F -Screening because of better availability and facilitated synthesis. By Linkage of this fragment with the acridine a bind-

ing in low μM range was observed as described in the Supplementary information.

Conclusion

In summary, we outline the screening of ^{19}F -containing libraries to 14 different RNA targets. Commercially available fragment library can be used to identify low molecular weight fragments. We also show the general versatility of the used poised library, allowing straight-forward follow-up chemistry to increase binding affinity to as low as $1\ \mu\text{M}$, while observing a 15-fold selectivity between different RNA targets. Identified hits report on the general druggability of RNA targets^[68]. Cheminformatics allow delineation of features within the fragments that are specific for each class of biomolecular target. Our study will aid our current effort to identify new ligands with antiviral activity in the context of combatting the COVID-19-pandemic (covid19-nmr.de).

Experimental Section

Initial checks

All fragments in the library utilized for screening were checked for inconsistencies by running 1D ^1H - and ^{19}F -spectra of each compound. Spectra were analyzed by hand and compounds showing wrong or additional signals were ruled out.

RNA preparation

All RNAs were prepared in house through in vitro transcription with T7 RNAP^[69]. DNA templates included the necessary T7 promotor and were either obtained from linearizing plasmid containing the sequence of interest or PCR run-off. Transcription conditions were optimized for yield and sample purity and in vitro transcription was performed in 10 to 20 ml scale dependent on the expected yield. Purification was performed either by HPLC, preparative PAGE or buffer exchange to NMR buffer if necessary^[70]. Concentration and purity of the samples were analyzed by UV-Vis spectroscopy and analytical PAGE respectively. For some of the follow-up experiments ^{15}N -uniformly labeled RNA was prepared with the same procedure using ^{15}N -isotopically labeled rNTPs.

Sample generation

Each fragment mixture contained 20 or 21 fragments at 2.5 mM concentration in 90%- d_6 -DMSO with 10% D_2O . Mixtures were designed (minimize signal overlap) in an excel sheet, using the chemical shift obtained from individual compound measurement. Sample-Pro Tube robot was used for automated pipetting of the samples into the NMR tubes. The final sample volume was 170 μL with 5% D_2O as locking solvent. For each target, two samples (with and without target) were prepared per mixture. ^{19}F -screening was performed at a ratio of 1:1 with respect to RNA and fragments. The final [RNA]([protein])-ligand concentration was around 50 μM . The screening buffer was 25 mM KPi, pH 6.2, 50 mM KCl, 5 mM MgCl_2 in 94% H_2O / 5% D_2O / 1% d_6 -DMSO in a 3 mm tube. For DNA, buffer conditions were 25mM KPi, pH 7.0, 70mM KCl, in 94% H_2O / 5% D_2O /

1% d_6 -DMSO. For proteins MtpA and PtkA buffer conditions were 25 mM HEPES/NaOH, pH 7.0, 150 mM NaCl, 10 mM DTT and 50 mM HEPES/NaOH, pH 7.5, 300 mM NaCl, 10 mM DTT, 10 mM MgCl_2 respectively. For EphA2 buffer conditions were 20 mM Tris pH 8, 200 mM NaCl, 5 mM MgCl_2 , 3 mM TCEP. For T7 and rS1 buffer conditions were 25 mM KPi (pH 7.2), 150 mM KCl, 5 mM DTT. NMR screening data of ^{19}F -1D and ^{19}F -CPMG- T_2 measurements were recorded with mixing times of 0 ms, 200 ms, 400 ms for the CPMG experiments. Strong hits from screening of three investigated RNAs were chosen for follow-up experiments. Samples of these fragments with the respective RNA were prepared in the same way omitting mixture generation, to confirm binding of the single fragment. In the same way, samples for the competition experiments were generated. Native ligand was added at an equimolar concentration to the RNA. Additionally, samples of ^{15}N -labeled RNA were prepared with an RNA concentration between 50 μM and 100 μM and fragment concentration of 1.25 mM and were used in follow-up experiments. Reference samples contained no fragment.

NMR spectroscopy

Spectra acquisition was carried out on a Bruker AVIIIHD-600 NMR spectrometer equipped with a five mm $^1\text{H}/^{19}\text{F}$ [^{13}C , ^{15}N]-TCI prodigy cryo-probe and high throughput sample changer for 579 samples with temperature option for sample storage. For the screening process the following spectra were acquired: ^{19}F -1D, water-suppressed proton 1D and ^{19}F -1D with CPMG spinlock (0, 200 and 400 ms). All ^{19}F spectra were recorded at room temperature, without ^1H -decoupling and processed with line broadening function of 10 Hz. The CPMG spin lock was applied using an adiabatic WURST pulse with a bandwidth of 120 ppm and a length of 2 ms. The pulse is calculated on-the-fly by wavemaker software in Topspin. The inter-pulse delay of the CPMG spin lock was set to 9 ms. Screening data were analyzed by integration with Topspin 4.0 (Bruker Biospin) and manually checked using the integrated fragment-based screening software tool. Strong hits for three RNAs investigated were chosen for follow-up experiments. ^{19}F -1D CPMG spectra were run on single fragment samples with the same parameters as in screening. The following spectra were acquired: ^{15}N -SFHMQC, proton 1Ds with excitation sculpting^[71] or jump-and-return echo^[72] scheme for water suppression, and ^1H , ^1H -TOCSY with excitation sculpting.

Synthesis

All experimental details for the synthesis and characterization of compound 1 are described in the Supporting Information.

Acknowledgements

The work was supported by funds from DFG in collaborative research center 902: "Molecular principles of RNA-based regulation", by EU-funded project iNEXT-discovery. Work at BMRZ is supported by the state of Hesse. VDJ was supported by Boehringer Ingelheim Fonds. RS was supported by Fonds der Chemischen Industrie.

Key words: FBS • fragment-based screening • fluorine • ^{19}F • RNA • DNA • proteins

References

- [1] Q. Vicens, E. Westhof, *ChemBioChem* **2003**, *4*, 1018–1023.
- [2] K. D. Warner, C. E. Hajdin, K. M. Weeks, *Nat. Rev. Drug Discov.* **2018**, *17*, 547–558.
- [3] A. Garcia-Lopez, F. Tessaro, H. R. A. Jonker, A. Wacker, C. Richter, A. Comte, N. Berntenis, R. Schmucki, K. Hatje, O. Petermann, G. Chiriano, R. Perozzo, D. Sciarra, P. Konieczny, I. Faustino, G. Fournet, M. Orozco, R. Artero, F. Metzger, M. Ebeling, P. Goekjian, B. Joseph, H. Schwalbe, L. Scapozza, *Nat. Commun.* **2018**, *9*, 1–12.
- [4] M. Sivaramakrishnan, K. D. McCarthy, S. Campagne, S. Huber, S. Meier, A. Augustin, T. Heckel, H. Meistermann, M. N. Hug, P. Birrer, A. Moursy, S. Khawaja, R. Schmucki, N. Berntenis, N. Giroud, S. Golling, M. Tzouros, B. Banfai, G. Duran-Pacheco, J. Lamerz, Y. Hsiu Liu, T. Luebbbers, H. Ratni, M. Ebeling, A. Cléry, S. Paushkin, A. R. Krainer, F. H. T. Allain, F. Metzger, *Nat. Commun.* **2017**, *8*, 1–13.
- [5] A. Donlic, A. E. Hargrove, *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1477.
- [6] M. Matsui, D. R. Corey, *Nat. Rev. Drug Discov.* **2017**, *16*, 167–179.
- [7] J. Palacino, S. E. Swalley, C. Song, A. K. Cheung, L. Shu, X. Zhang, M. Van Hoosear, Y. Shin, D. N. Chin, C. G. Keller, M. Beibel, N. A. Renaud, T. M. Smith, M. Salcius, X. Shi, M. Hild, R. Servais, M. Jain, L. Deng, C. Bullock, M. McLellan, S. Schuierer, L. Murphy, M. J. J. Blommers, C. Blaustein, F. Berenshteyn, A. Lacoste, J. R. Thomas, G. Roma, G. A. Michaud, B. S. Tseng, J. A. Porter, V. E. Myer, J. A. Tallarico, L. G. Hamann, D. Curtis, M. C. Fishman, W. F. Dietrich, N. A. Dales, R. Sivasankaran, *Nat. Chem. Biol.* **2015**, *11*, 511–517.
- [8] L. S. Waters, G. Storz, *Cell* **2009**, *136*, 615–628.
- [9] D. van Duin, D. L. Paterson, *Infect. Dis. Clin. North Am.* **2016**, *30*, 377–390.
- [10] M. S. Butler, M. A. Cooper, *Curr. Drug Targets* **2012**, *13*, 373–387.
- [11] D. R. Davies, Humana Press, New York, NY, **2014**, pp. 315–323.
- [12] T. Sugiki, K. Furuita, T. Fujiwara, C. Kojima, *Molecules* **2018**, *23*, 148.
- [13] X. Liu, D. Shi, S. Zhou, H. Liu, H. Liu, X. Yao, *Expert Opin. Drug Discov.* **2018**, *13*, 23–37.
- [14] D. A. Erlanson, in *Top. Curr. Chem.*, **2011**, pp. 1–32.
- [15] D. A. Erlanson, I. J. P. de Esch, W. Jahnke, C. N. Johnson, P. N. Mortenson, *J. Med. Chem.* **2020**, acs.jmedchem.9b01581.
- [16] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science (80-)*. **1996**, *274*, 1531–1534.
- [17] J. Tsai, J. T. Lee, W. Wang, J. Zhang, H. Cho, S. Mamo, R. Bremer, S. Gillette, J. Kong, N. K. Haass, K. Sproesser, L. Li, K. S. M. Smalley, D. Fong, Y.-L. Zhu, A. Marimuthu, H. Nguyen, B. Lam, J. Liu, I. Cheung, J. Rice, Y. Suzuki, C. Luu, C. Settachatgul, R. Shellooe, J. Cantwell, S.-H. Kim, J. Schlessinger, K. Y. J. Zhang, B. L. West, B. Powell, G. Habets, C. Zhang, P. N. Ibrahim, P. Hirth, D. R. Artis, M. Herlyn, G. Bollag, *Proc. Natl. Acad. Sci.* **2008**, *105*, 3041–3046.
- [18] P. A. Brough, W. Aherne, X. Barril, J. Borgognoni, K. Boxall, J. E. Cansfield, K.-M. J. Cheung, I. Collins, N. G. M. Davies, M. J. Drysdale, B. Dymock, S. A. Eccles, H. Finch, A. Fink, A. Hayes, R. Howes, R. E. Hubbard, K. James, A. M. Jordan, A. Lockie, V. Martins, A. Massey, T. P. Matthews, E. McDonald, C. J. Northfield, L. H. Pearl, C. Prodromou, S. Ray, F. I. Raynaud, S. D. Roughley, S. Y. Sharp, A. Surgenor, D. L. Walmsley, P. Webb, M. Wood, P. Workman, L. Wright, *J. Med. Chem.* **2008**, *51*, 196–218.
- [19] S. Howard, V. Berdini, J. A. Boulstridge, M. G. Carr, D. M. Cross, J. Curry, L. A. Devine, T. R. Early, L. Fazal, A. L. Gill, M. Heathcote, S. Maman, J. E. Matthews, R. L. McMenamin, E. F. Navarro, M. A. O'Brien, M. O'Reilly, D. C. Rees, M. Reule, D. Tisi, G. Williams, M. Vinković, P. G. Wyatt, *J. Med. Chem.* **2009**, *52*, 379–388.
- [20] C.-M. Park, M. Bruncko, J. Adickes, J. Bauch, H. Ding, A. Kunzer, K. C. Marsh, P. Nimmer, A. R. Shoemaker, X. Song, S. K. Tahir, C. Tse, X. Wang, M. D. Wendt, X. Yang, H. Zhang, S. W. Fesik, S. H. Rosenberg, S. W. Elmore, *J. Med. Chem.* **2008**, *51*, 6902–6915.
- [21] Y.-S. Wang, C. Strickland, J. H. Voigt, M. E. Kennedy, B. M. Beyer, M. M. Senior, E. M. Smith, T. L. Nechuta, V. S. Madison, M. Czarniecki, B. A. McKittrick, A. W. Stamford, E. M. Parker, J. C. Hunter, W. J. Greenlee, D. F. Wyss, *J. Med. Chem.* **2010**, *53*, 942–950.
- [22] E. Largy, F. Hamon, M.-P. Teulade-Fichou, *Anal. Bioanal. Chem.* **2011**, *400*, 3419–3427.
- [23] P. N. Asare-Okai, C. S. Chow, *Anal. Biochem.* **2011**, *408*, 269–276.
- [24] S. Laughlin, W. Wilson, *Int. J. Mol. Sci.* **2015**, *16*, 24506–24531.
- [25] H. Shimizu, F. Jinno, A. Morohashi, Y. Yamazaki, M. Yamada, T. Kondo, S. Asahi, *J. Mass Spectrom.* **2012**, *47*, 1015–1022.
- [26] J. Sztuba-Solinska, S. R. Shenoy, P. Gareiss, L. R. H. Krumpke, S. F. J. Le Grice, B. R. O'Keefe, J. S. Schneekloth, *J. Am. Chem. Soc.* **2014**, *136*, 8402–8410.
- [27] C. M. Connelly, R. E. Boer, M. H. Moon, P. Gareiss, J. S. Schneekloth, *ACS Chem. Biol.* **2017**, *12*, 435–443.
- [28] M. H. Moon, T. A. Hillimire, A. M. Sanders, J. S. Schneekloth, *Biochemistry* **2018**, *57*, 4638–4643.
- [29] M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, *J. Mol. Struct.* **2014**, *1077*, 101–113.
- [30] M. Pellicchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, G. Siegal, *Nat. Rev. Drug Discov.* **2008**, *7*, 738–745.
- [31] M. Garavís, B. López-Méndez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas, C. González, *ACS Chem. Biol.* **2014**, *9*, 1559–1566.
- [32] M. Zeiger, S. Stark, E. Kalden, B. Ackermann, J. Ferner, U. Scheffer, F. Shoja-Bazargani, V. Erdel, H. Schwalbe, M. W. Göbel, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 5576–5580.
- [33] D. R. Groebe, O. C. Uhlenbeck, *Nucleic Acids Res.* **1988**, *16*,

- 11725–11735.
- [34] J. P. Sheehy, A. R. Davis, B. M. Znosko, *RNA* **2010**, *16*, 417–429.
- [35] J. L. Childs-Disney, J. Hoskins, S. G. Rzuczek, C. A. Thornton, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 856–862.
- [36] A. C. Stelzer, A. T. Frank, J. D. Kratz, M. D. Swanson, M. J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D. M. Markovitz, H. M. Al-Hashimi, *Nat. Chem. Biol.* **2011**, *7*, 553–559.
- [37] C. Hong, M. Hagihara, K. Nakatani, *Angew. Chemie Int. Ed.* **2011**, *50*, 4390–4393.
- [38] J. P. Marino, R. S. Gregorian, G. Csankovszki, D. M. Crothers, *Science (80-.)*. **1995**, *268*, 1448–1454.
- [39] M. H. Bailor, A. M. Mustoe, C. L. Brooks, H. M. Al-Hashimi, *Curr. Opin. Struct. Biol.* **2011**, *21*, 296–305.
- [40] M. H. Bailor, C. Musselman, A. L. Hansen, K. Gulati, D. J. Patel, H. M. Al-Hashimi, *Nat. Protoc.* **2007**, *2*, 1536–1546.
- [41] F. Hamy, V. Brondani, A. Flörsheimer, W. Stark, M. J. J. Blommers, T. Klimkait, *Biochemistry* **1998**, *37*, 5086–5095.
- [42] A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **1997**, *267*, 624–639.
- [43] A. T. Frank, A. C. Stelzer, H. M. Al-Hashimi, I. Andricioaei, *Nucleic Acids Res.* **2009**, *37*, 3670–3679.
- [44] N. N. Patwardhan, L. R. Ganser, G. J. Kapral, C. S. Eubanks, J. Lee, B. Sathyamoorthy, H. M. Al-Hashimi, A. E. Hargrove, *Medchemcomm* **2017**, *8*, 1022–1036.
- [45] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577–586.
- [46] H. Schwalbe, J. Buck, B. Fürtig, J. Noeske, J. Wöhnert, *Angew. Chemie Int. Ed.* **2007**, *46*, 1212–1219.
- [47] S. D. Gilbert, S. J. Mediatore, R. T. Batey, *J. Am. Chem. Soc.* **2006**, *128*, 14214–14215.
- [48] E. R. Lee, K. F. Blount, R. R. Breaker, *RNA Biol.* **2009**, *6*, 187–194.
- [49] A. Reining, S. Nozinovic, K. Schlepckow, F. Buhr, B. Fürtig, H. Schwalbe, *Nature* **2013**, *499*, 355–9.
- [50] C. Helmling, D. Klötzner, F. Sochor, R. A. Mooney, A. Wacker, R. Landick, B. Fürtig, A. Heckel, H. Schwalbe, **n.d.**
- [51] M. Garavís, B. López-Méndez, A. Somoza, J. Oyarzabal, C. Dalvit, A. Villasante, R. Campos-Olivas, C. González, *ACS Chem. Biol.* **2014**, *9*, 1559–1566.
- [52] L. R. Ganser, J. Lee, A. Rangadurai, D. K. Merriman, M. L. Kelly, A. D. Kansal, B. Sathyamoorthy, H. M. Al-Hashimi, *Nat. Struct. Mol. Biol.* **2018**, *25*, 425–434.
- [53] D. W. Begley, S. O. Moen, P. G. Pierce, E. R. Zartler, in *Curr. Protoc. Chem. Biol.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2013**, pp. 251–268.
- [54] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349–359.
- [55] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **1954**, *94*, 630–638.
- [56] Y. Kim, C. Hilty, **2019**, pp. 501–526.
- [57] C. Dalvit, P. E. Fagermess, D. T. A. Hadden, R. W. Sarver, B. J. Stockman, *J. Am. Chem. Soc.* **2003**, *125*, 7696–7703.
- [58] S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **1958**, *29*, 688–691.
- [59] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, R. R. Breaker, *Science* **2012**, *335*, 233–235.
- [60] C. E. Dann, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, W. C. Winkler, *Cell* **2007**, *130*, 878–892.
- [61] M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
- [62] S. H. Rüdissler, N. Goldberg, M.-O. Ebert, H. Kovacs, A. D. Gossert, *J. Biomol. NMR* **2020**, *7*, DOI 10.1007/s10858-020-00325-x.
- [63] S. G. Rzuczek, M. R. Southern, M. D. Disney, *ACS Chem. Biol.* **2015**, *10*, 2706–15.
- [64] B. S. Morgan, J. E. Forte, R. N. Culver, Y. Zhang, A. E. Hargrove, *Angew. Chemie Int. Ed.* **2017**, *56*, 13498–13502.
- [65] S. R. Kirk, N. W. Luedtke, Y. Tor, *J. Am. Chem. Soc.* **2000**, *122*, 980–981.
- [66] P. Prasher, M. Sharma, *Medchemcomm* **2018**, *9*, 1589–1618.
- [67] J. F. Arambula, S. R. Ramisetty, A. M. Baranger, S. C. Zimmerman, *Proc. Natl. Acad. Sci.* **2009**, *106*, 16068–16073.
- [68] P. J. Hajduk, J. R. Huth, C. Tse, *Drug Discov. Today* **2005**, *10*, 1675–1682.
- [69] J. K. Bains, J. Blechar, V. de Jesus, N. Meiser, H. Zetzsche, B. Fürtig, H. Schwalbe, M. Hengesbach, *Methods* **2019**, *153*, DOI 10.1016/j.ymeth.2018.10.004.
- [70] C. Helmling, S. Keyhani, F. Sochor, B. Fürtig, M. Hengesbach, H. Schwalbe, *J. Biomol. NMR* **2015**, *63*, 67–76.
- [71] T. L. Hwang, A. J. Shaka, *J. Magn. Reson. Ser. A* **1995**, *112*, 275–279.
- [72] G. Marius Clore, B. J. Kimber, A. M. Gronenborn, *J. Magn. Reson.* **1983**, *54*, 170–173.
- [73] L. Huang, J. Wang, D. M. J. Lilley, *Cell Chem. Biol.* **2017**, *24*, 695–702.e2.
- [74] A. Ren, K. R. Rajashankar, D. J. Patel, *Structure* **2015**, *23*, 1375–1381.
- [75] A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **2006**, *441*, 1167–1171.
- [76] 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.
- [77] K. D. Smith, S. V. Lipchock, A. L. Livingston, C. A. Shanahan, S. A. Strobel, *Biochemistry* **2010**, *49*, 7351–7359.
- [78] K. D. Smith, S. V. Lipchock, T. D. Ames, J. Wang, R. R. Breaker, S. A. Strobel, *Nat. Struct. Mol. Biol.* **2009**, *16*, 1218–1223.
- [79] N. Kulshina, N. J. Baird, A. R. Ferré-D'amaré, *Nat. Struct. Mol. Biol.* **2009**, *16*, 1212–1217.
- [80] A. Serganov, Y.-R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, D. J. Patel, *Chem. Biol.* **2004**, *11*, 1729–1741.
- [81] J. R. Stagno, Y. Liu, Y. R. Bhandari, C. E. Conrad, S. Panja, M. Swain, L. Fan, G. Nelson, C. Li, D. R. Wendel, T. A. White, J. D. Coe, M. O. Wiedorn, J. Knoska, D. Oberthuer, R. A. Tuckey, P. Yu, M. Dyba, S. G. Tarasov, U. Weierstall, T. D. Grant, C. D. Schwieters, J. Zhang, A. R. Ferré-D'amaré, P. Fromme, D. E. Draper, M. Liang, M. S. Hunter, S. Boutet, K. Tan, X. Zuo, X. Ji, A. Barty, N. A. Zatsepin, H. N. Chapman, J. C. H. Spence, S. A.

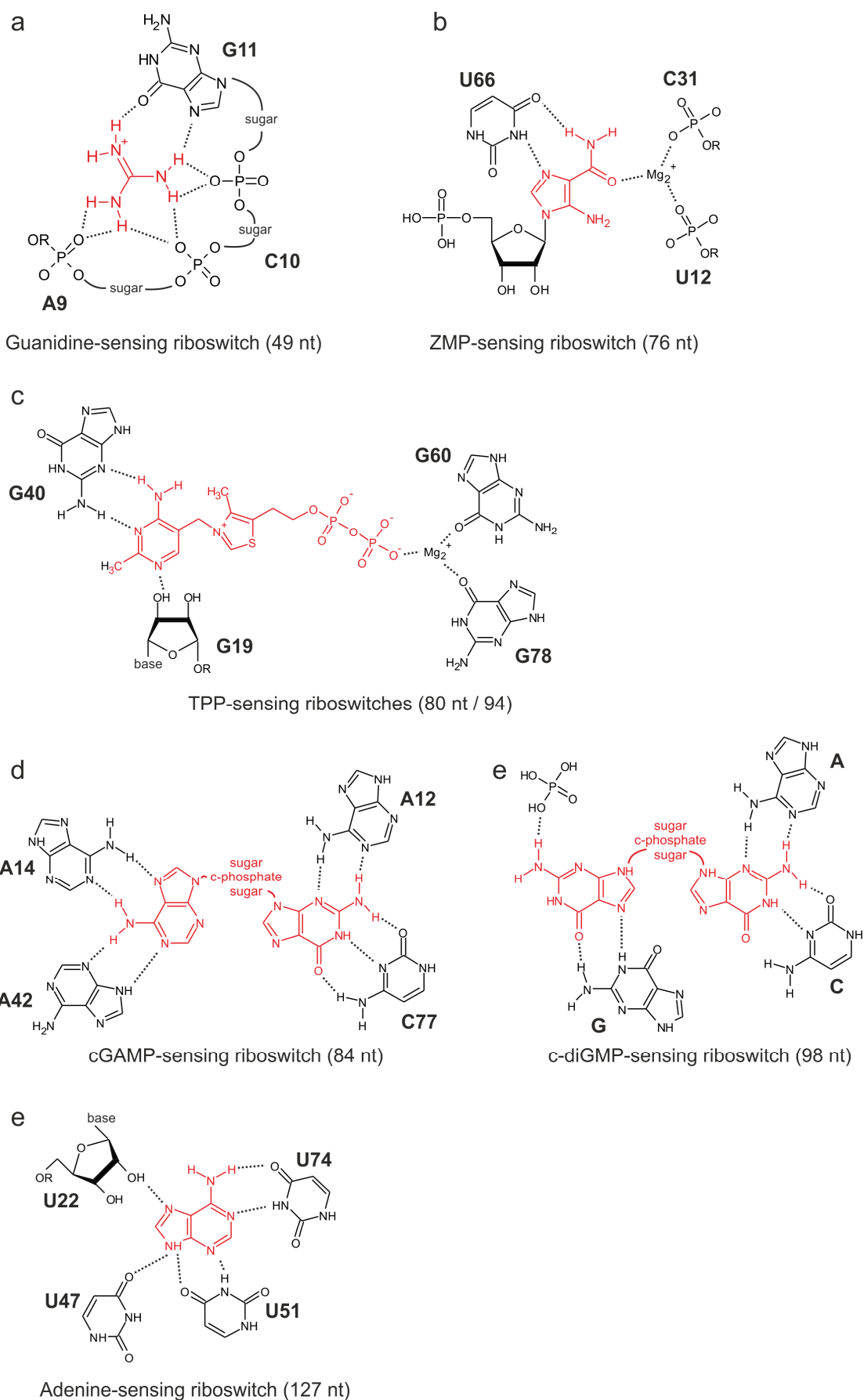
- Woodson, Y. X. Wang, *Nature* **2017**, *541*, 242–246.
- [82] J. Zhang, A. R. Ferré-D'Amaré, *Structure* **2014**, *22*, 1363–1371.
- [83] C. Helmling, A. Wacker, M. T. Wolfinger, I. L. Hofacker, M. Hengesbach, B. Fürtig, H. Schwalbe, *J. Am. Chem. Soc.* **2017**, *139*, 2647–2656.
- [84] R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBioChem* **2007**, *8*, 896–902.
- [85] R. Schnieders, C. Richter, S. Warhaut, V. de Jesus, S. Keyhani, E. Duchardt-Ferner, H. Keller, J. Wöhnert, L. T. Kuhn, A. L. Breeze, W. Bermel, H. Schwalbe, B. Fürtig, *J. Biomol. NMR* **2017**, *69*, 31–44.
- [86] K. Okada, M. Takahashi, T. Sakamoto, G. Kawai, K. Nakamura, A. Kanai, *Nucleosides, Nucleotides and Nucleic Acids* **2006**, *25*, 383–395.
- [87] F. M. Jucker, A. Pardi, *Biochemistry* **1995**, *34*, 14416–14427.
- [88] W. G. Scott, J. T. Finch, A. Klug, *Cell* **1995**, *81*, 991–1002.
- [89] A. R. Ferré-D'Amaré, K. Zhou, J. A. Doudna, *Nature* **1998**, *395*, 567–574.
- [90] N. H. Woo, B. A. Roe, A. Rich, *Nature* **1980**, *286*, 346–351.
- [91] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [92] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [93] D. Wei, G. N. Parkinson, A. P. Reszka, S. Neidle, *Nucleic Acids Res.* **2012**, *40*, 4691–4700.
- [94] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386–4392.
- [95] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 2179–2183.
- [96] J. Dai, C. PUNCHIHEWA, A. Ambrus, D. Chen, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
- [97] W. Liu, Y. F. Zhong, L. Y. Liu, C. T. Shen, W. Zeng, F. Wang, D. Yang, Z. W. Mao, *Nat. Commun.* **2018**, *9*, DOI 10.1038/s41467-018-05810-4.
- [98] J. Wirmer-Bartoschek, L. E. Bendel, H. R. A. Jonker, J. T. Grün, F. Papi, C. Bazzicalupi, L. Messori, P. Gratteri, H. Schwalbe, *Angew. Chemie - Int. Ed.* **2017**, *56*, 7102–7106.
- [99] J. Dai, M. Carver, C. PUNCHIHEWA, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
- [100] C. Madhurantakam, E. Rajakumara, P. A. Mazumdar, B. Saha, D. Mitra, H. G. Wiker, R. Sankaranarayanan, A. K. Das, *J. Bacteriol.* **2005**, *187*, 2175–2181.
- [101] T. Stehle, S. Sreeramulu, F. Löhr, C. Richter, K. Saxena, H. R. A. Jonker, H. Schwalbe, *J. Biol. Chem.* **2012**, *287*, 34569–34582.
- [102] A. Niesteruk, H. R. A. Jonker, C. Richter, V. Linhard, S. Sreeramulu, H. Schwalbe, *J. Biol. Chem.* **2018**, *293*, 11823–11836.
- [103] S. Heinzlmeir, D. Kudlinzki, S. Sreeramulu, S. Klaeger, S. L. Gande, V. Linhard, M. Wilhelm, H. Qiao, D. Helm, B. Ruprecht, K. Saxena, G. Médard, H. Schwalbe, B. Kuster, *ACS Chem. Biol.* **2016**, *11*, 3400–3411.
- [104] P. Giraud, J. B. Créchet, M. Uzan, F. Bontems, C. Sizun, *Biomol. NMR Assign.* **2015**, *9*, 107–111.
- [105] P. Salah, M. Bisaglia, P. Aliprandi, M. Uzan, C. Sizun, F. Bontems, *Nucleic Acids Res.* **2009**, *37*, 5578–5588.
- [106] Y. W. Yin, T. A. Steitz, *Science (80-)*. **2002**, *298*, 1387–1395.

Table 1. List of all biomolecules used in the study listed with their biological host organism (if applicable), PDB accession codes of X-ray structures and primary publication.

	Organism	X-Ray	NMR
Riboswitches and Aptamers			
Guanidine (Gdn-II)-sensing riboswitch (49 nt)	<i>Escherichia Coli</i>	5NDI ^[73]	
ZMP-sensing riboswitch (76 nt)	<i>Thermosinus carboxydivorans</i>	4ZNP ^[74]	
thiM TPP-sensing riboswitch (80 nt)	<i>E. Coli</i>	2GDI ^[75]	
pilM 3', 3'-cGAMP-sensing riboswitch (84 nt)	<i>Geobacter metallireducens</i>	4YAZ ^[76]	
TenA TTP- sensing riboswitch (94 nt)	<i>Staphylococcus aureus</i>		

cyclic di-GMP-1 riboswitch (98 nt)	<i>Clostridium difficile</i>	3MXH ^{#[77]} 3IRW ^{#[78]} 3IWN ^{#[79]}	
Adenine-sensing riboswitch (127 nt)	<i>Vibrio vulnificus</i>	1y26 ^{#[80]} 5E54 ^{#[81]} 4TZX ^{#[82]}	[49] [69]
Riboswitch Elements			
2'-deoxyguanosine-sensing-riboswitch terminator (39 nt)	<i>Mesoplasma florum</i>		[83]
SAM-sensing riboswitch anti-terminator (38 nt)	<i>Bacillus subtilis</i>		
Adenine-sensing riboswitch terminator (51 nt)	<i>B. subtilis</i>		[84]
Adenine-sensing riboswitch expression platform (60 nt)	<i>V. vulnificus</i>		[69] [85]
Other RNAs			
RNA with GAAG tetraloop (14 nt)	<i>artificial</i>		2F87 ^{#[86]}
RNA CUUG tetraloop (14 nt)	<i>artificial</i>		1RNG ^{#[87]}
Hammerhead ribozyme (54 nt)		1MME ^{#[88]}	
Hepatitis delta virus ribozyme (70 nt)	<i>Hepatitis delta virus</i>	1DRZ ^{#[89]}	
tRNA ^{Met} (77 nt)	<i>E. coli</i>		[90]
DNA			
cMyc G-Quadruplex (22 nt)	<i>Homo sapiens</i>		1XAV ^{#[91]} 2L7V ^{#[92]}
cKit G-Quadruplex (24 nt)	<i>H. sapiens</i>	3QXR ^{#[93]} 2WO2	2O3M ^{#[94]}
DNA Duplex (24 nt)	<i>artificial</i>	1BNA ^{#[95]}	
Tel26 G-Quadruplex (26 nt)	<i>H. sapiens</i>		2HY9 ^{#[96]} 5Z80 ^{#[97]}
wtTel26 G-Quadruplex (26 nt)	<i>H. sapiens</i>		5MVB ^{#[98]} 2JPZ ^{#[99]}
Proteins			
Mycobacterium tuberculosis Protein Tyrosine Phosphatase A (MptpA, 18 kDa)	<i>Mycobacterium tuberculosis</i>	1U2P ^{#[100]}	2LUO ^{#[101]}
Protein tyrosine Kinase A (PtkA, 30 kDa)	<i>M. tuberculosis</i>		6F2X ^{#[102]}
Receptor tyrosine kinase EphA2 (34 kDa)	<i>H. sapiens</i>	5I9U ^{#[103]}	
Ribosomal protein S1 (61 kDa)	<i>V. vulnificus</i>		2MFI ^{#[104]} 2KHI ^{#[105]}
T7 RNA polymerase (100 kDa)	<i>Escherichia phage T7</i>	1MSW ^{#[106]}	

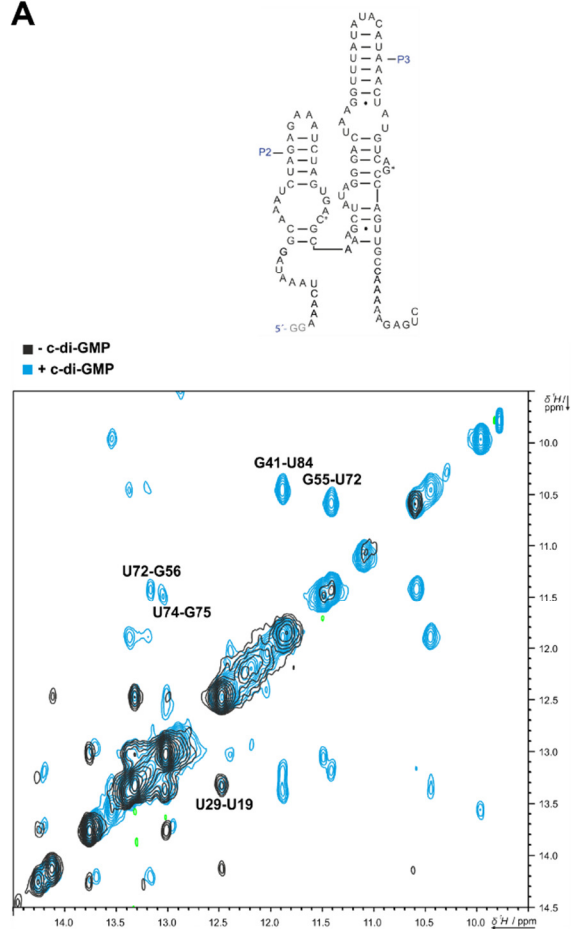
Accepted Manuscript



Supplementary Figure 1: Molecular recognition of ligands (red) by riboswitches used in ^{19}F -screening. (A) Guanidine-sensing riboswitch, (B) ZMP-sensing riboswitch, (C) TPP-sensing riboswitches, (D) cGAMP-sensing riboswitch, (E) c-diGMP-sensing riboswitch, (F) Adenine-sensing riboswitch.

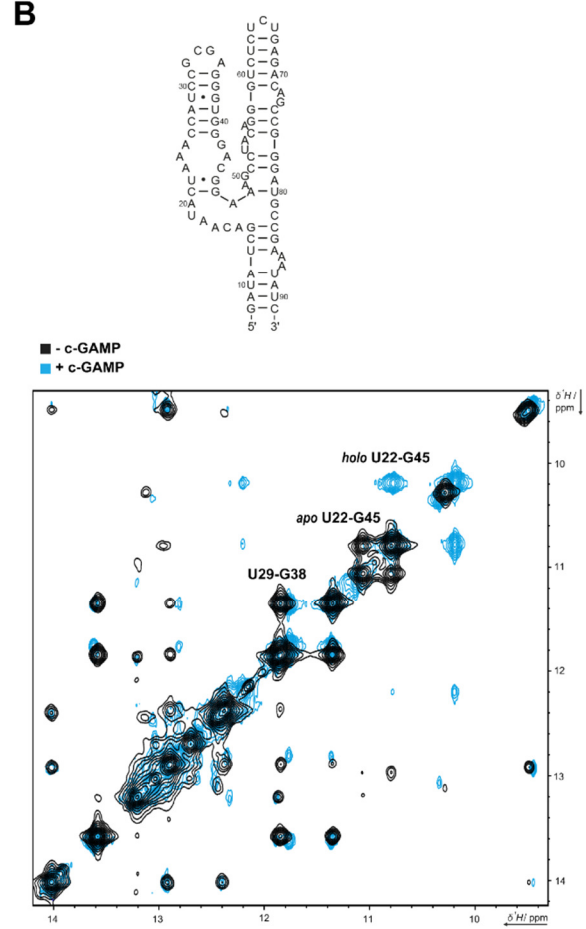
c-di-GMP-sensing riboswitch

A



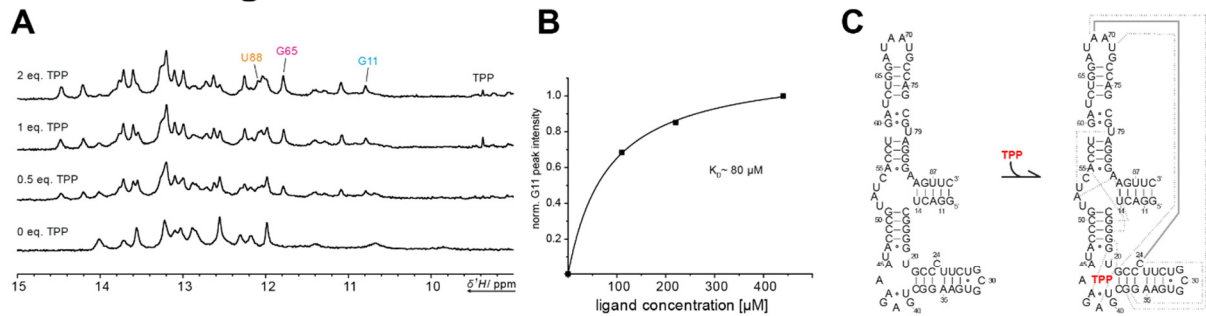
GAMP-sensing riboswitch

B

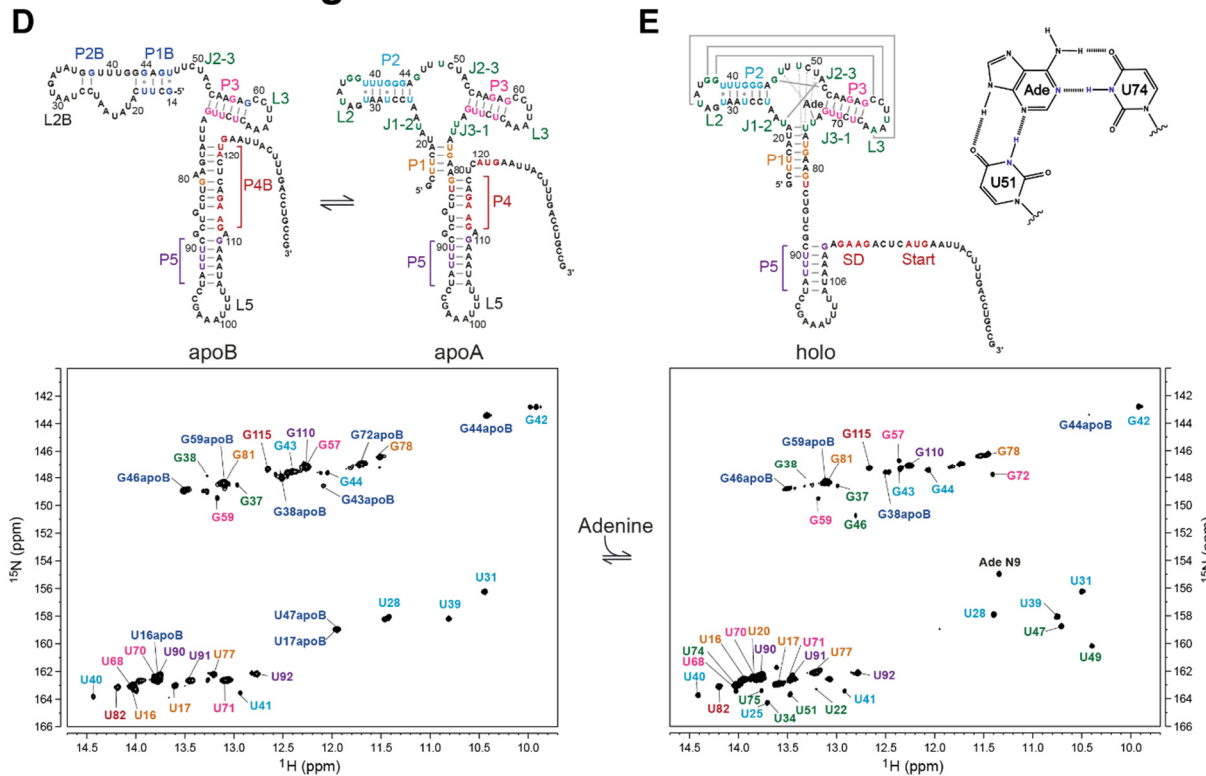


Supplementary Figure 2: (A) $[^1\text{H}, ^1\text{H}]$ -NOESY spectra of the 98 nt c-di-GMP-sensing Cd1 riboswitch from *Clostridium difficile* in the apo and holo conformation. The secondary structure of the apo conformation is shown. Spectra were recorded in 50 mM Bis-Tris buffer with 120 mM NaCl and 20 mM MgCl_2 at 298 K, 600 MHz and pH 6.1. Spectra were recorded with a spectral width of 24 ppm and 16 ppm in the indirect dimension. Spectra were recorded with 2048 points in the direct and 380 points in the indirect dimension and an acquisition time of 71 ms. The spectrum without c-di-GMP was recorded with 384 scans and the spectrum with c-di-GMP with 364. The sample concentration was 500 μM . (B) $[^1\text{H}, ^1\text{H}]$ -NOESY spectra of the pilM 3',3'-cGAMP-sensing riboswitch (84 nt) from *Geobacter metallireducens* in the apo and holo conformation. The secondary structure of the apo conformation is shown. The spectra were recorded in 25 mM potassium phosphate buffer (pH 6.2), 50 mM KCl, 5 mM MgCl_2 and 10% D_2O . DSS was used as reference. The mixing time was 100 ms. The spectra were recorded at 308 K. The apo spectrum was recorded at an RNA concentration of 1.6 mM at 700 MHz with 256 scans. The spectrum was recorded with a spectral width of 24 ppm in the direct and 15 ppm in the indirect dimension and an acquisition time of 63 ms. The apo spectrum was recorded with 3172 points in the direct and 488 points in the indirect dimension. The holo spectrum was recorded at an RNA concentration of 1.1 mM with 1.6 mM cGAMP at 950 MHz with 312 scans. The spectrum was recorded with a spectral width of 24 ppm in the direct and 15.5 ppm in the indirect dimension and an acquisition time of 64 ms. Spectrum was recorded with 2914 points in the direct and 384 points in the indirect dimension.

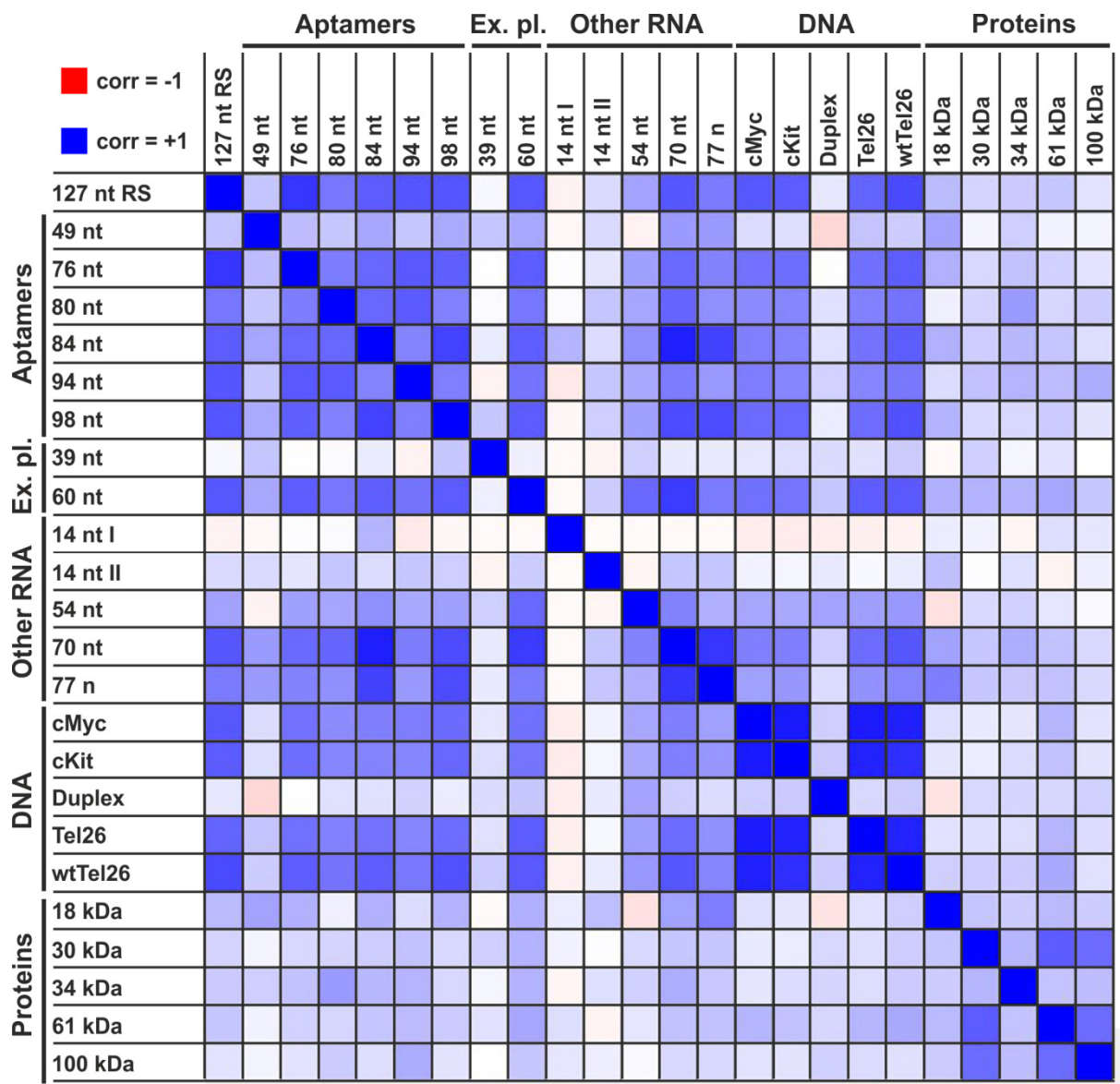
TPP-sensing riboswitch thiM



Adenine-sensing riboswitch



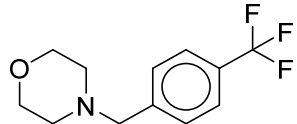
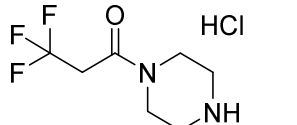
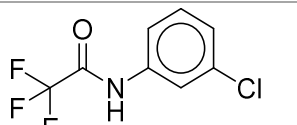
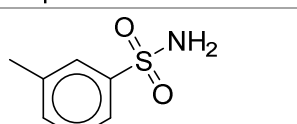
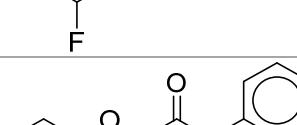
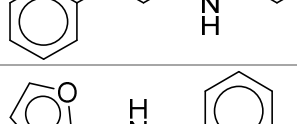
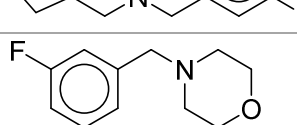
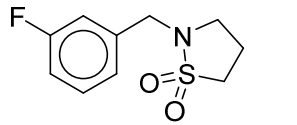
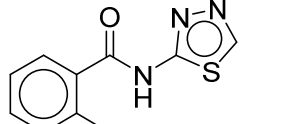
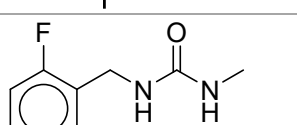
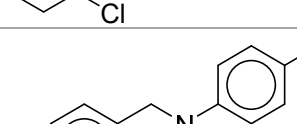
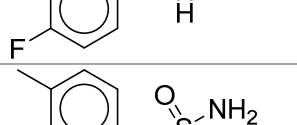
Supplementary Figure 3: (A) Imino proton region of 1D ^1H spectra of $u\text{-}^{13}\text{C}$, ^{15}N labeled thiM RNA (80 nt) with 5 mM Mg^{2+} at increasing TPP concentrations. Spectra were recorded at a 700 HD MHz Bruker spectrometer at 298 K. All spectra were recorded with 8 scans. The spectral width was 24 ppm. Sample had an RNA concentration of 220 μM in NMR buffer (25 mM potassium phosphate buffer, pH 6.2, 10% D_2O , 100 μM DSS) (B) K_D value determination of TPP ligand binding to thiM RNA. The intensities of the G11 peak in the titration shown in (C) were normalized to the highest value and fitted using a hyperbolic one-site binding curve. (C) Secondary structure of thiM RNA (80 nt). (D) ^1H - ^{15}N -BEST TROSY spectra of 127 nt adenine-sensing riboswitch from *Vibrio vulnificus* in apo and holo conformation. Spectra were recorded in potassium phosphate buffer (25 mM KPi, 150 mM KCl and 5 mM MgCl_2 , pH 7.2) at 800 MHz. The proton carrier frequency was set to the resonance frequency of the solvent (4.7 ppm). The interscan delay was set to 0.4 seconds. The spectra were recorded with a spectral width of 24 ppm in the direct dimension and 30 ppm in the indirect dimension. The spectra were recorded with an acquisition time of 103 ms representing 3998 points in the ^1H dimension. In the indirect dimension, 256 points were recorded with an acquisition time of 52 ms. (A) The apo spectrum was recorded with 4 scans at an RNA concentration of 465 μM . (E) The holo spectrum was recorded with 8 scans at an RNA concentration of 229 μM .

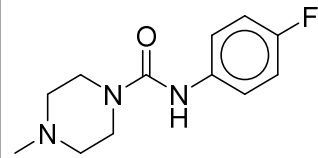
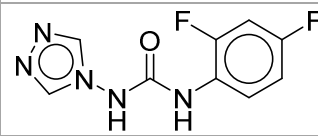
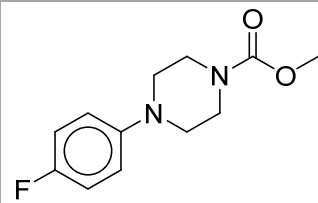
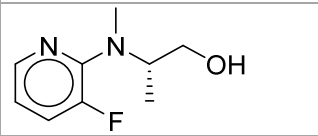
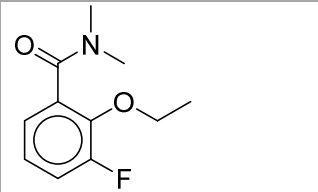
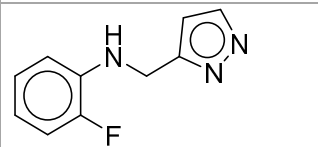
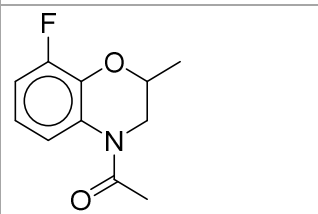
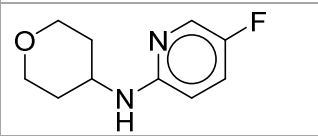
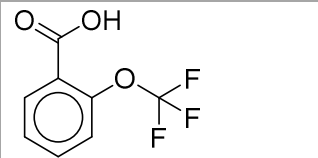
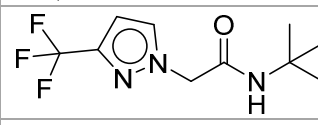
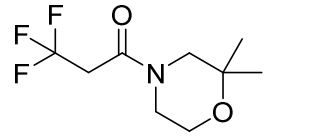


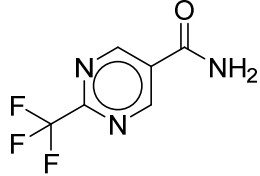
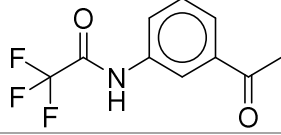
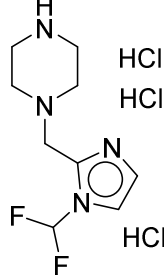
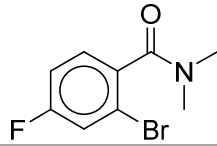
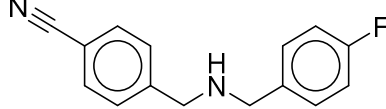
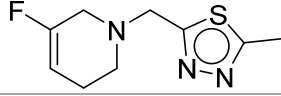
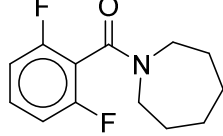
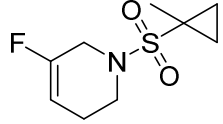
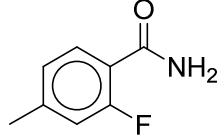
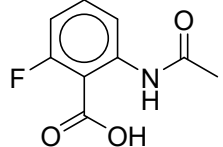
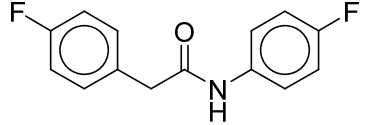
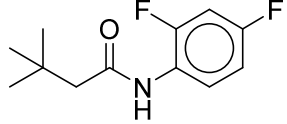
Supplementary Figure 4: Correlation matrix of hit clusters, displaying hit correlation between different targets screened by ¹⁹F-FBS.

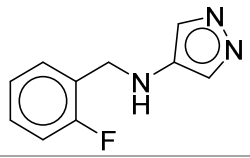
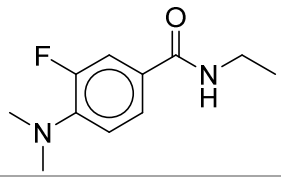
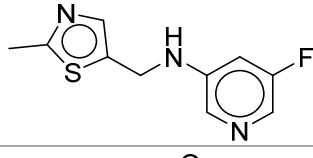
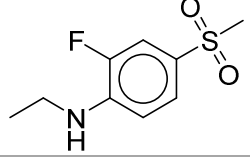
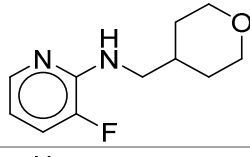
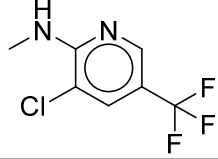
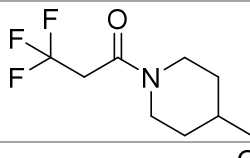
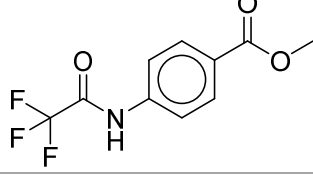
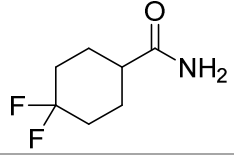
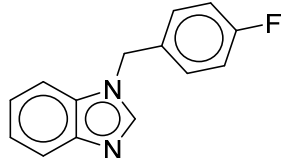
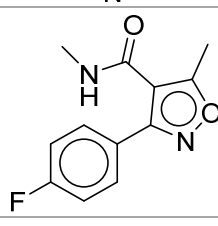
¹⁹F Library

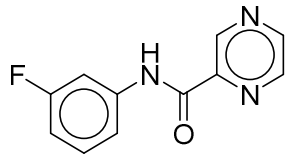
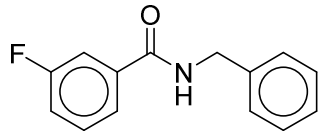
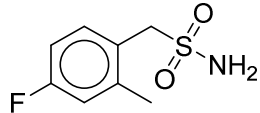
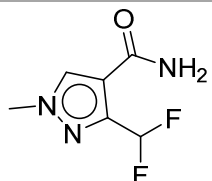
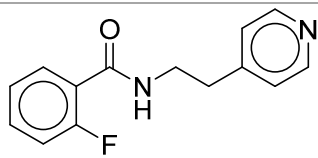
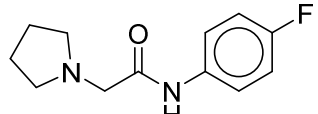
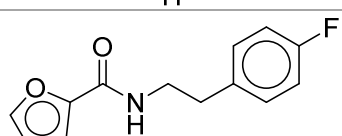
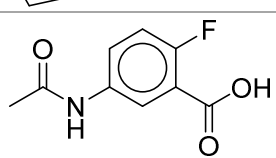
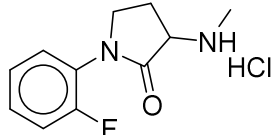
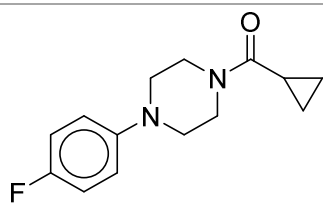
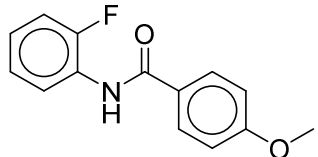
Supplementary Table 1: ¹⁹F library for the ¹⁹F-NMR-based fragment screening.

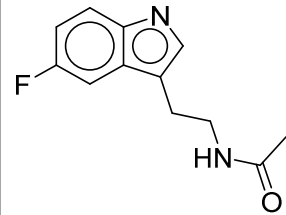
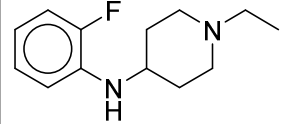
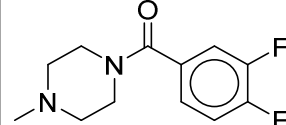
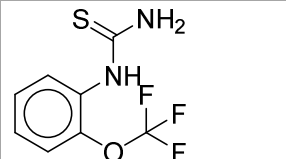
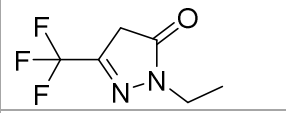
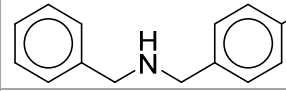
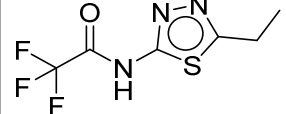
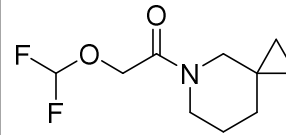
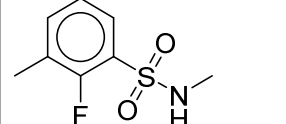
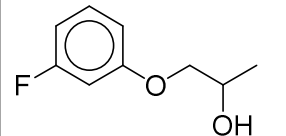
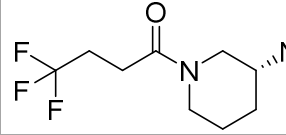
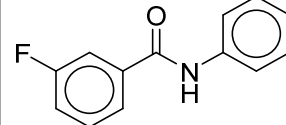
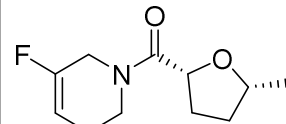
	SMILES code	Formula	Chemical structure
1.	<chem>FC(F)(F)c1ccc(cc1)CN2CCOCC2</chem>	C ₁₂ H ₁₄ F ₃ NO	
2.	<chem>Cl.O=C(CC(F)(F)F)N1CCNCC1</chem>	C ₇ H ₁₂ ClF ₃ N ₂ O	
3.	<chem>Clc1cc(NC(=O)C(F)(F)F)ccc1</chem>	C ₈ H ₅ ClF ₃ NO	
4.	<chem>Fc1cc(C)cc(c1)S(=O)(N)=O</chem>	C ₇ H ₈ FNO ₂ S	
5.	<chem>Fc1cccc(c1)NC(=O)COc2ccccc2</chem>	C ₁₄ H ₁₂ FNO ₂	
6.	<chem>Fc2cccc(CNCc1ccco1)c2</chem>	C ₁₂ H ₁₂ FNO	
7.	<chem>Fc2cc(CN1CCOCC1)ccc2</chem>	C ₁₁ H ₁₄ FNO	
8.	<chem>O=S2(=O)CCCN2Cc1cccc(F)c1</chem>	C ₁₀ H ₁₂ FNO ₂ S	
9.	<chem>O=C(Nc1nncs1)c2ccccc2F</chem>	C ₉ H ₆ FN ₃ OS	
10.	<chem>Clc1cccc(F)c1CNC(=O)NC</chem>	C ₉ H ₁₀ ClFN ₂ O	
11.	<chem>COc2ccc(NCc1ccc(F)cc1)cc2</chem>	C ₁₄ H ₁₄ FNO	
12.	<chem>Cc1ccc(CS(N)(=O)=O)cc1F</chem>	C ₈ H ₁₀ FNO ₂ S	

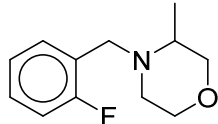
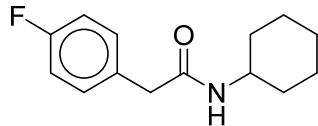
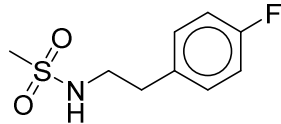
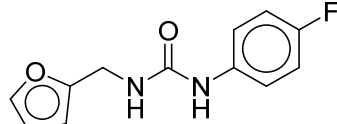
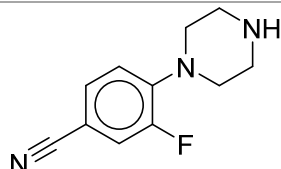
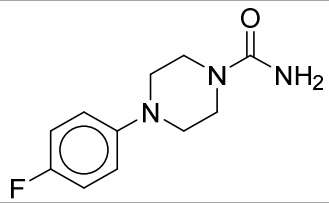
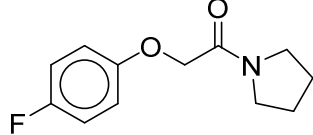
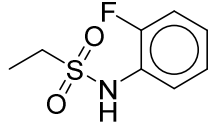
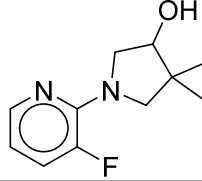
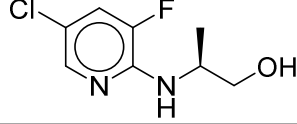
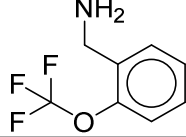
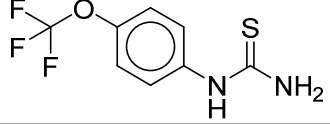
13.	<chem>O=C(Nc1ccc(F)cc1)N2CCN(C)CC2</chem>	C12H16FN3O	
14.	<chem>O=C(Nc1ccc(F)cc1F)Nn2cnnc2</chem>	C9H7F2N5O	
15.	<chem>COC(=O)N1CCN(CC1)c2ccc(F)cc2</chem>	C12H15FN2O2	
16.	<chem>Fc1cccnc1N(C)[C@@H](C)CO</chem>	C9H13FN2O	
17.	<chem>CCOc1c(cccc1F)C(=O)N(C)C</chem>	C11H14FNO2	
18.	<chem>Fc2ccccc2NCc1ccnn1</chem>	C10H10FN3	
19.	<chem>CC(=O)N2CC(C)Oc1c(F)cccc12</chem>	C11H12FNO2	
20.	<chem>Fc1ccc(nc1)NC2CCOCC2</chem>	C10H13FN2O	
21.	<chem>FC(F)(F)Oc1ccccc1C(=O)O</chem>	C8H5F3O3	
22.	<chem>O=C(Cn1ccc(n1)C(F)(F)F)NC(C)(C)C</chem>	C10H14F3N3O	
23.	<chem>CC1(C)CN(CCO1)C(=O)CC(F)(F)F</chem>	C9H14F3NO2	

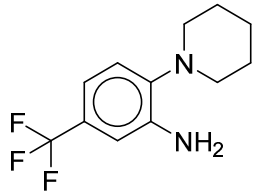
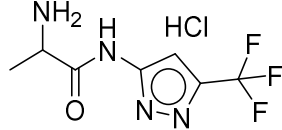
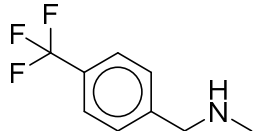
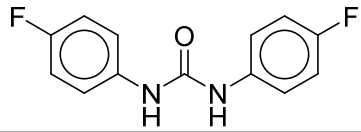
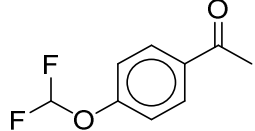
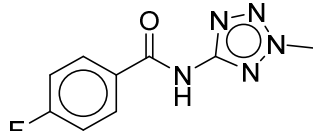
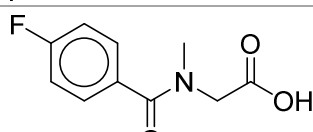
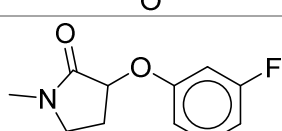
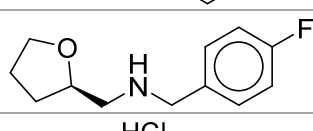
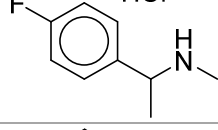
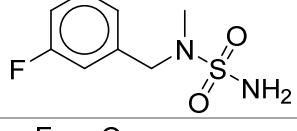
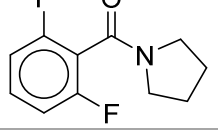
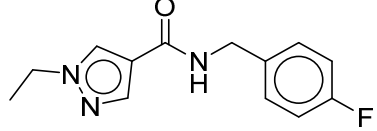
24.	<chem>FC(F)(F)c1ncc(cn1)C(N)=O</chem>	C ₆ H ₄ F ₃ N ₃ O	
25.	<chem>O=C(Nc1cc(ccc1)C(C)=O)C(F)(F)F</chem>	C ₁₀ H ₈ F ₃ NO ₂	
26.	<chem>Cl.Cl.Cl.FC(F)N2CCNC2CN1CCNCC1</chem>	C ₉ H ₁₇ Cl ₃ F ₂ N ₄	
27.	<chem>O=C(c1ccc(F)cc1Br)N(C)C</chem>	C ₉ H ₉ BrFNO	
28.	<chem>Fc2ccc(CNCc1ccc(C#N)cc1)cc2</chem>	C ₁₅ H ₁₃ FN ₂	
29.	<chem>FC=1CN(CCC=1)Cc2nnc(C)s2</chem>	C ₉ H ₁₂ FN ₃ S	
30.	<chem>O=C(c1c(F)cccc1F)N2CCCCC2</chem>	C ₁₃ H ₁₅ F ₂ NO	
31.	<chem>FC=1CN(CCC=1)S(=O)(=O)C2(C)CC2</chem>	C ₉ H ₁₄ FNO ₂ S	
32.	<chem>O=C(N)c1ccc(C)cc1F</chem>	C ₈ H ₈ FNO	
33.	<chem>O=C(C)Nc1cccc(F)c1C(=O)O</chem>	C ₉ H ₈ FNO ₃	
34.	<chem>O=C(Nc1ccc(F)cc1)Cc2ccc(F)cc2</chem>	C ₁₄ H ₁₁ F ₂ NO	
35.	<chem>O=C(Nc1ccc(F)cc1F)CC(C)(C)C</chem>	C ₁₂ H ₁₅ F ₂ NO	

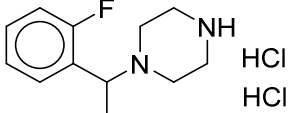
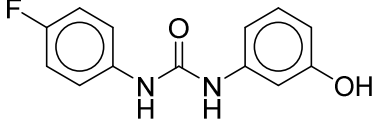
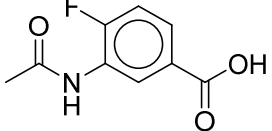
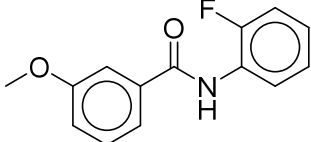
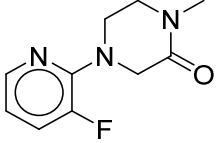
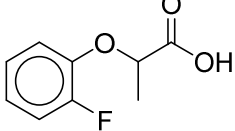
36.	<chem>Fc2ccccc2CNc1cnnc1</chem>	C10H10FN3	
37.	<chem>Fc1cc(ccc1N(C)C)C(=O)NCC</chem>	C11H15FN2O	
38.	<chem>Cc2ncc(CNc1cc(F)cnc1)s2</chem>	C10H10FN3S	
39.	<chem>Fc1cc(ccc1NCC)S(=O)(C)=O</chem>	C9H12FNO2S	
40.	<chem>Fc2ccnc2NCC1CCOCC1</chem>	C11H15FN2O	
41.	<chem>Clc1cc(cnc1NC)C(F)(F)F</chem>	C7H6ClF3N2	
42.	<chem>CC1CCN(CC1)C(=O)CC(F)(F)F</chem>	C9H14F3NO	
43.	<chem>O=C(Nc1ccc(cc1)C(=O)OC)C(F)(F)F</chem>	C10H8F3NO3	
44.	<chem>FC1(F)CCC(CC1)C(N)=O</chem>	C7H11F2NO	
45.	<chem>Fc1ccc(cc1)Cn3cnc2ccccc23</chem>	C14H11FN2	
46.	<chem>CNC(=O)c2c(C)onc2c1ccc(F)cc1</chem>	C12H11FN2O2	

47.	<chem>O=C(Nc1cccc(F)c1)c2cncn2</chem>	C11H8FN3O	
48.	<chem>O=C(NCc1ccccc1)c2cc(F)ccc2</chem>	C14H12FNO	
49.	<chem>Fc1cc(C)c(CS(N)(=O)=O)cc1</chem>	C8H10FNO2S	
50.	<chem>FC(F)c1nn(C)cc1C(N)=O</chem>	C6H7F2N3O	
51.	<chem>Fc1ccccc1C(=O)NCCc2ccncc2</chem>	C14H13FN2O	
52.	<chem>O=C(Nc1ccc(F)cc1)CN2CCCC2</chem>	C12H15FN2O	
53.	<chem>O=C(NCCc1ccc(F)cc1)c2ccco2</chem>	C13H12FNO2	
54.	<chem>O=C(C)Nc1cc(c(F)cc1)C(=O)O</chem>	C9H8FNO3	
55.	<chem>Cl.Fc1ccccc1N2CCC(NC)C2=O</chem>	C11H14ClFN2O	
56.	<chem>O=C(C1CC1)N2CCN(CC2)c3ccc(F)cc3</chem>	C14H17FN2O	
57.	<chem>COC1ccc(cc1)C(=O)Nc2ccccc2F</chem>	C14H12FNO2	

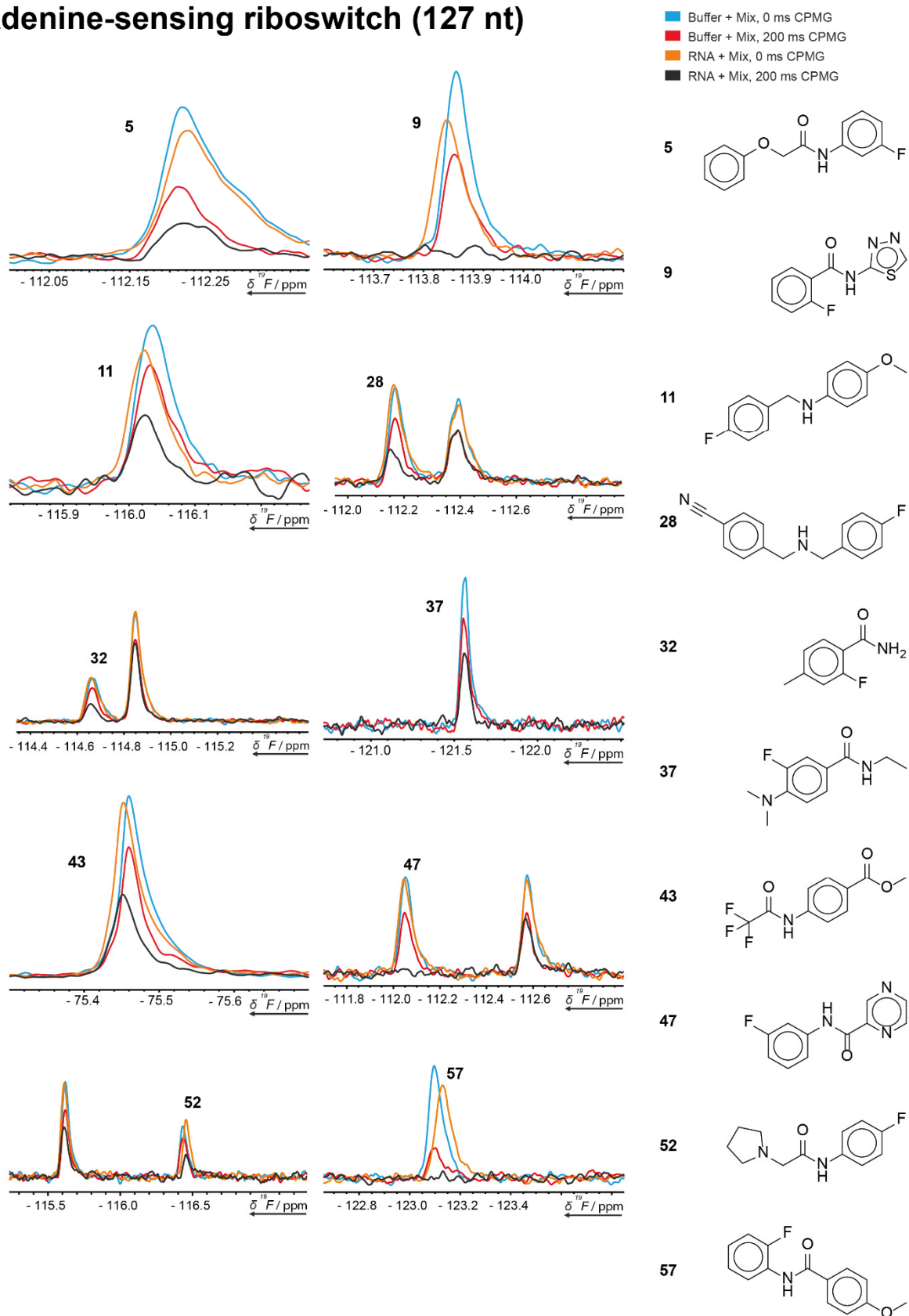
58.	<chem>CC(=O)NCCc2cnc1ccc(F)cc12</chem>	C12H13FN2O	
59.	<chem>CCN2CCC(Nc1ccccc1F)CC2</chem>	C13H19FN2	
60.	<chem>O=C(N1CCN(C)CC1)c2ccc(F)c(F)c2</chem>	C12H14F2N2O	
61.	<chem>FC(F)(F)Oc1ccccc1NC(N)=S</chem>	C8H7F3N2OS	
62.	<chem>FC(F)(F)C=1CC(=O)N(CC)N=1</chem>	C6H7F3N2O	
63.	<chem>Fc2ccc(CNCc1ccccc1)cc2</chem>	C14H14FN	
64.	<chem>O=C(Nc1nnc(CC)s1)C(F)(F)F</chem>	C6H6F3N3OS	
65.	<chem>FC(F)OCC(=O)N2CCCC1(CC1)C2</chem>	C10H15F2NO2	
66.	<chem>Fc1c(C)cccc1S(=O)(=O)NC</chem>	C8H10FNO2S	
67.	<chem>CC(O)COc1cccc(F)c1</chem>	C9H11FO2	
68.	<chem>Cl.N[C@@H]1CCCN(C1)C(=O)CCC(F)(F)F</chem>	C9H16ClF3N2O	
69.	<chem>Cc2ccc(NC(=O)c1cccc(F)c1)cc2O</chem>	C14H12FNO2	
70.	<chem>O=C(N1CC(F)=CCC1)[C@H]2CC[C@@H](C)O2</chem>	C11H16FNO2	

71.	<chem>Fc2ccccc2CN1CCOCC1C</chem>	C ₁₂ H ₁₆ FNO	
72.	<chem>O=C(NC1CCCCC1)Cc2ccc(F)cc2</chem>	C ₁₄ H ₁₈ FNO	
73.	<chem>Fc1ccc(CCNS(C)=O)cc1</chem>	C ₉ H ₁₂ FNO ₂ S	
74.	<chem>O=C(NCc1ccco1)Nc2ccc(F)cc2</chem>	C ₁₂ H ₁₁ FN ₂ O ₂	
75.	<chem>N#Cc1cc(F)c(cc1)N2CCNCC2</chem>	C ₁₁ H ₁₂ FN ₃	
76.	<chem>O=C(N)N1CCN(CC1)c2ccc(F)cc2</chem>	C ₁₁ H ₁₄ FN ₃ O	
77.	<chem>O=C(COc1ccc(F)cc1)N2CCCC2</chem>	C ₁₂ H ₁₄ FNO ₂	
78.	<chem>O=S(=O)(Nc1ccccc1F)CC</chem>	C ₈ H ₁₀ FNO ₂ S	
79.	<chem>Fc1cccnc1N2CC(C)(C)C(O)C2</chem>	C ₁₁ H ₁₅ FN ₂ O	
80.	<chem>C[C@@H](CO)Nc1ncc(Cl)cc1F</chem>	C ₈ H ₁₀ ClFN ₂ O	
81.	<chem>FC(F)(F)Oc1ccccc1CN</chem>	C ₈ H ₈ F ₃ NO	
82.	<chem>FC(F)(F)Oc1ccc(cc1)NC(N)=S</chem>	C ₈ H ₇ F ₃ N ₂ OS	

83.	<chem>FC(F)(F)c1cc(N)c(cc1)N2CCCC2</chem>	C ₁₂ H ₁₅ F ₃ N ₂	
84.	<chem>Cl.FC(F)(F)c1cc(NC(=O)C(C)N)nn1</chem>	C ₇ H ₁₀ ClF ₃ N ₄ O	
85.	<chem>CNCc1ccc(cc1)C(F)(F)F</chem>	C ₉ H ₁₀ F ₃ N	
86.	<chem>O=C(Nc1ccc(F)cc1)Nc2ccc(F)cc2</chem>	C ₁₃ H ₁₀ F ₂ N ₂ O	
87.	<chem>FC(F)Oc1ccc(cc1)C(C)=O</chem>	C ₉ H ₈ F ₂ O ₂	
88.	<chem>O=C(Nc1nn(C)nn1)c2ccc(F)cc2</chem>	C ₉ H ₈ FN ₅ O	
89.	<chem>O=C(c1ccc(F)cc1)N(C)CC(=O)O</chem>	C ₁₀ H ₁₀ FN ₃ O ₃	
90.	<chem>CN2CCC(Oc1cc(F)ccc1)C2=O</chem>	C ₁₁ H ₁₂ FN ₂ O ₂	
91.	<chem>Fc2ccc(CNC[C@H]1CCCO1)cc2</chem>	C ₁₂ H ₁₆ FN ₂ O	
92.	<chem>Cl.Fc1ccc(cc1)C(C)NC</chem>	C ₉ H ₁₃ ClFN	
93.	<chem>Fc1cc(CN(C)S(N)(=O)=O)ccc1</chem>	C ₈ H ₁₁ FN ₂ O ₂ S	
94.	<chem>O=C(c1c(F)cccc1F)N2CCCC2</chem>	C ₁₁ H ₁₁ F ₂ N ₂ O	
95.	<chem>O=C(NCc1ccc(F)cc1)c2cn(CC)nc2</chem>	C ₁₃ H ₁₄ FN ₃ O	

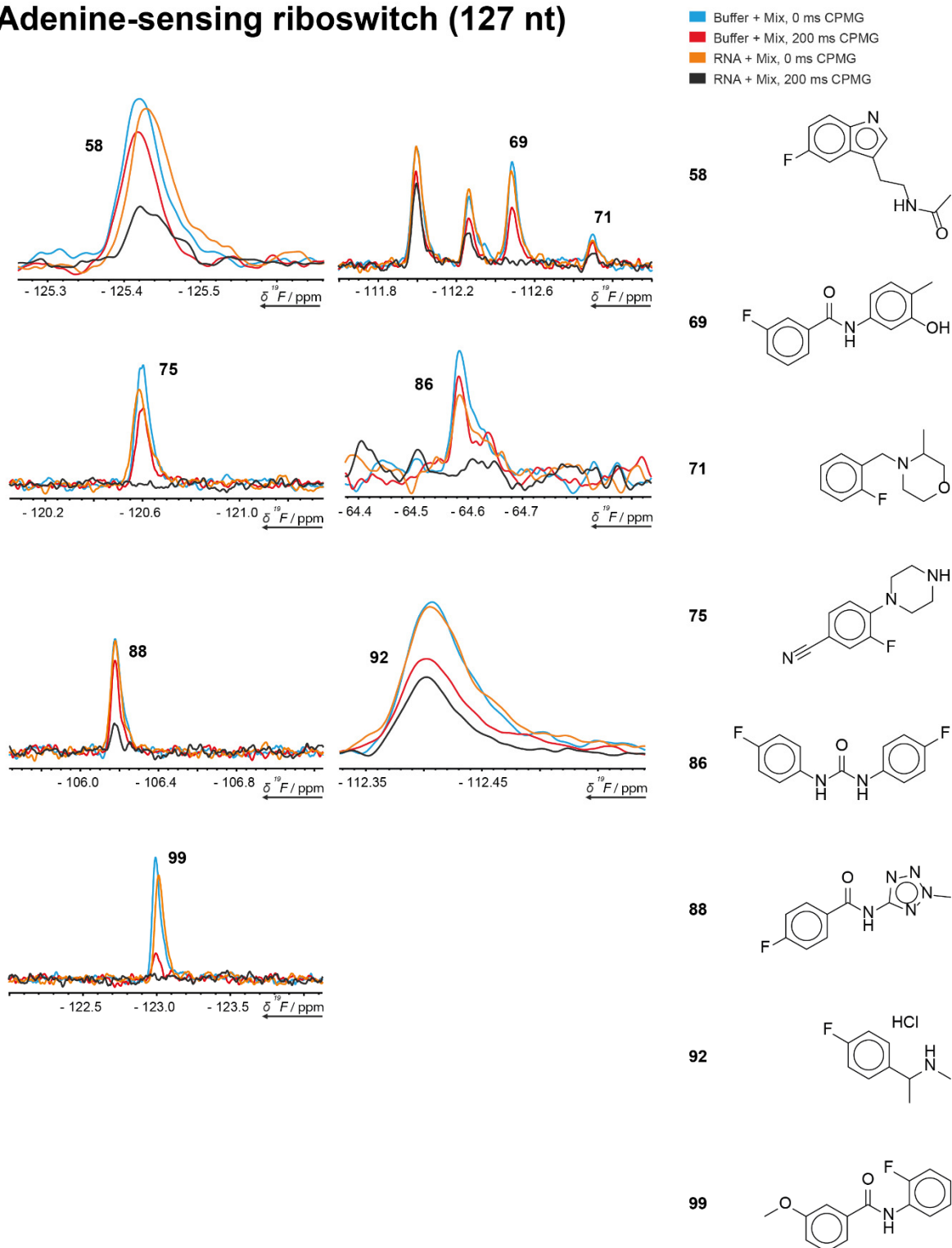
96.	<chem>Cl.Cl.CC(c1ccccc1F)N2CCNCC2</chem>	C ₁₂ H ₁₉ Cl ₂ FN ₂	
97.	<chem>Oc2cccc(NC(=O)Nc1ccc(F)cc1)c2</chem>	C ₁₃ H ₁₁ FN ₂ O ₂	
98.	<chem>O=C(C)Nc1cc(ccc1F)C(=O)O</chem>	C ₉ H ₈ FN ₂ O ₃	
99.	<chem>COc1cc(ccc1)C(=O)Nc2ccccc2F</chem>	C ₁₄ H ₁₂ FN ₂ O ₂	
100.	<chem>CN1CCN(CC1=O)c2ncccc2F</chem>	C ₁₀ H ₁₂ FN ₃ O	
101.	<chem>CC(Oc1ccccc1F)C(=O)O</chem>	C ₉ H ₉ FO ₃	

Adenine-sensing riboswitch (127 nt)



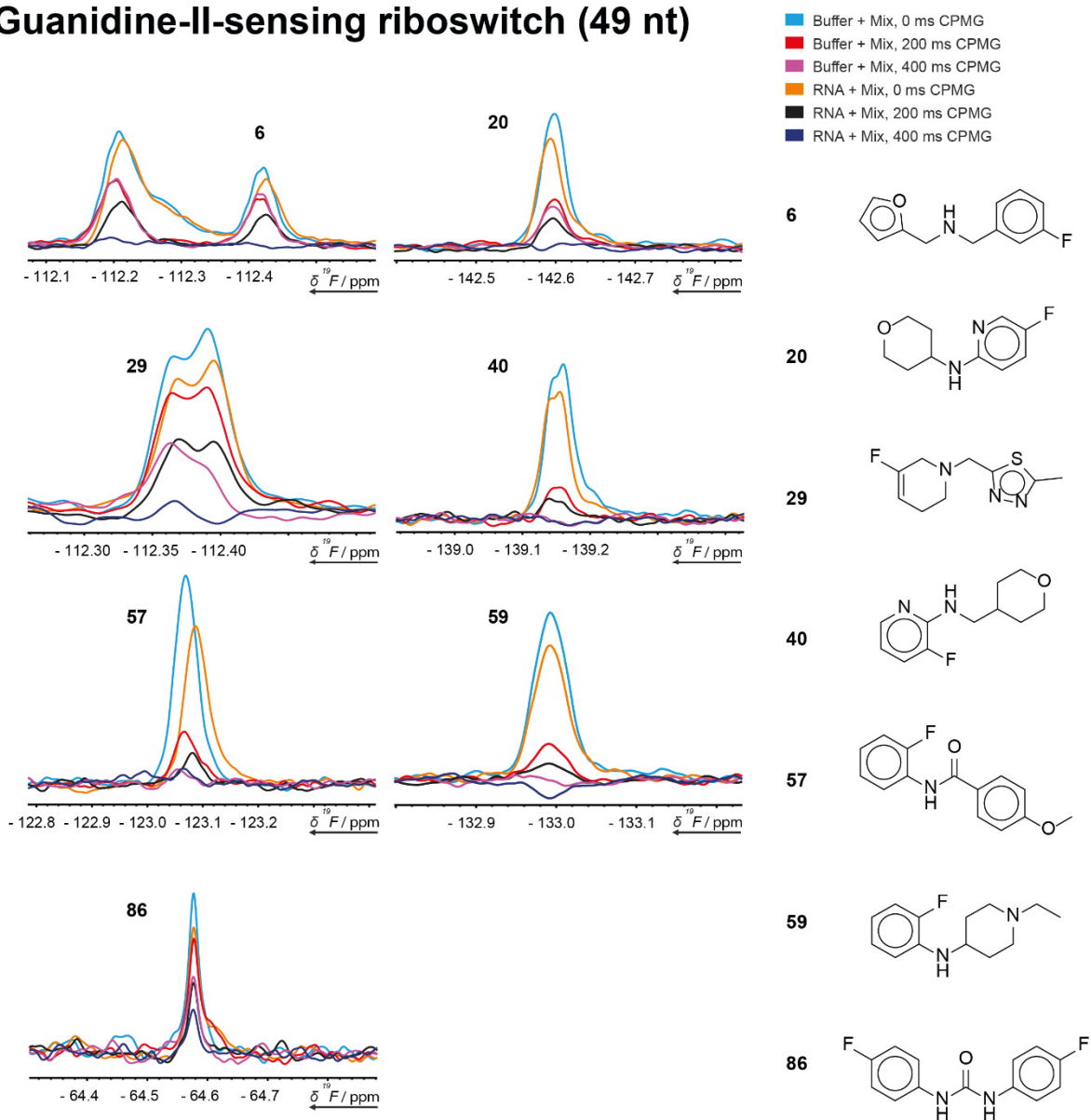
Supplementary Figure 5: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the adenine-sensing riboswitch (127 nt) from *Vibrio vulnificus*.

Adenine-sensing riboswitch (127 nt)



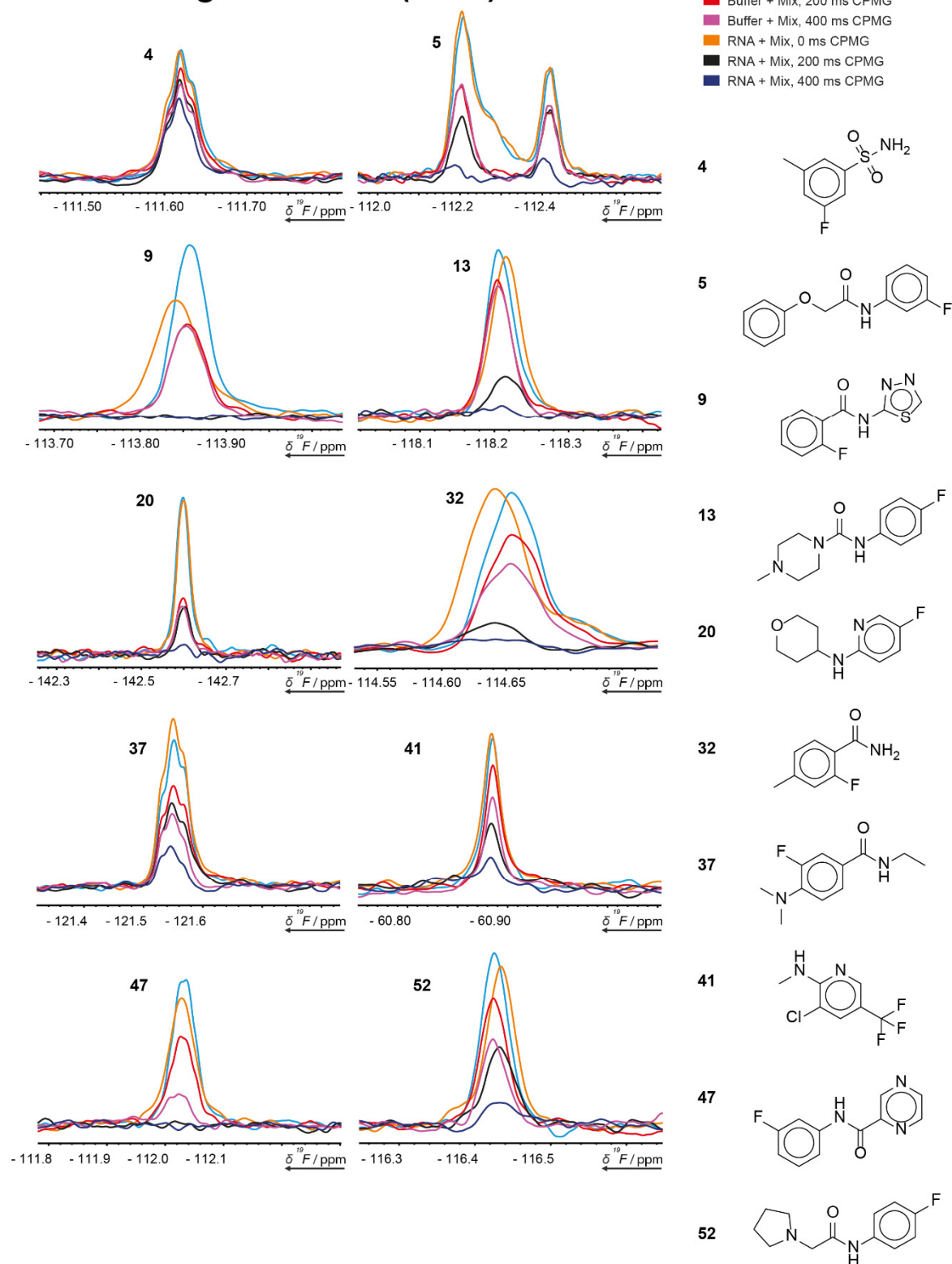
Supplementary Figure 6: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the adenine-sensing riboswitch (127 nt) from *Vibrio vulnificus*.

Guanidine-II-sensing riboswitch (49 nt)



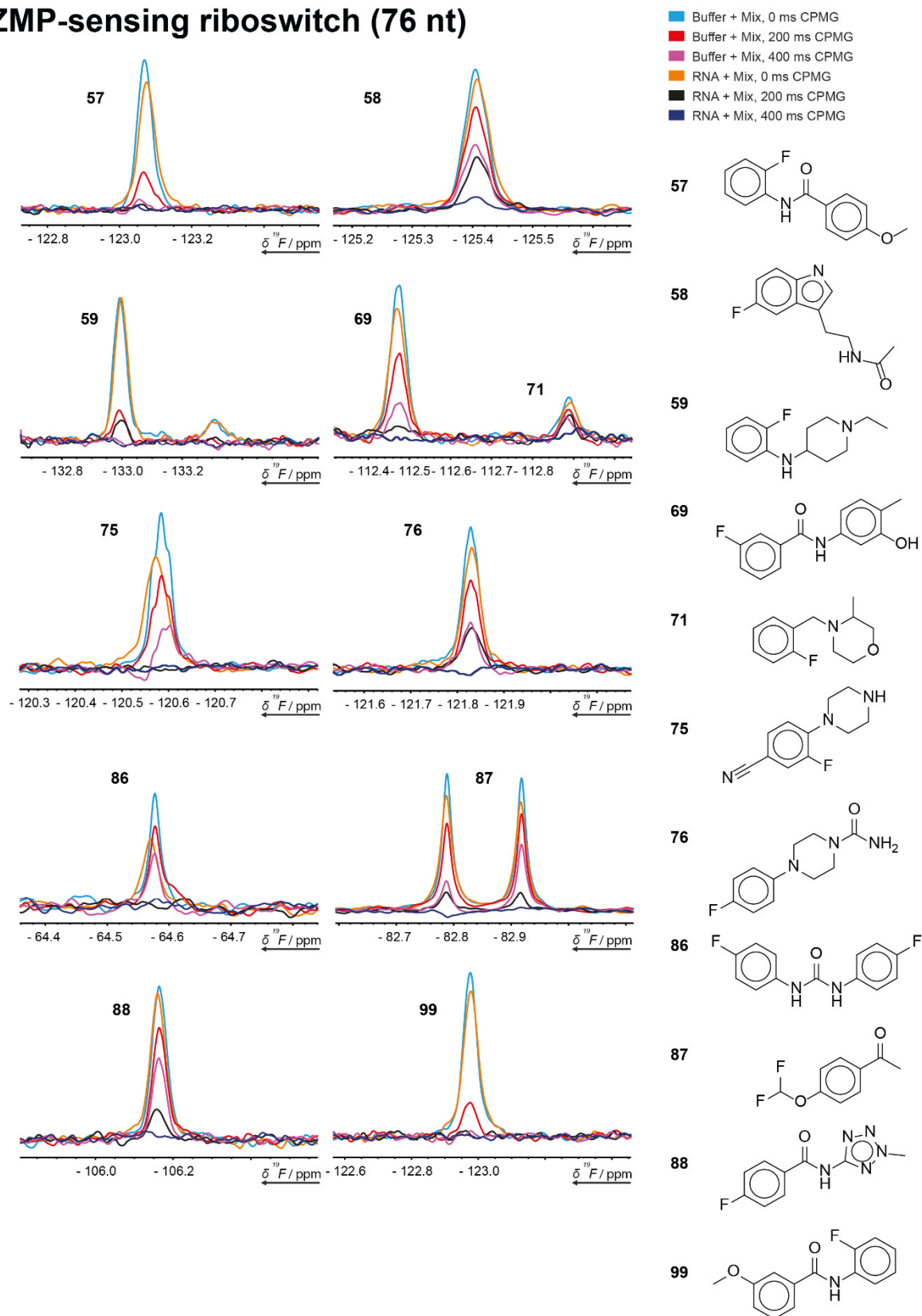
Supplementary Figure 7: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Guanidine-II riboswitch (49 nt) from *E. coli*. For further hit validation, the CPMG experiment at 400 ms is shown.

ZMP-sensing riboswitch (76 nt)



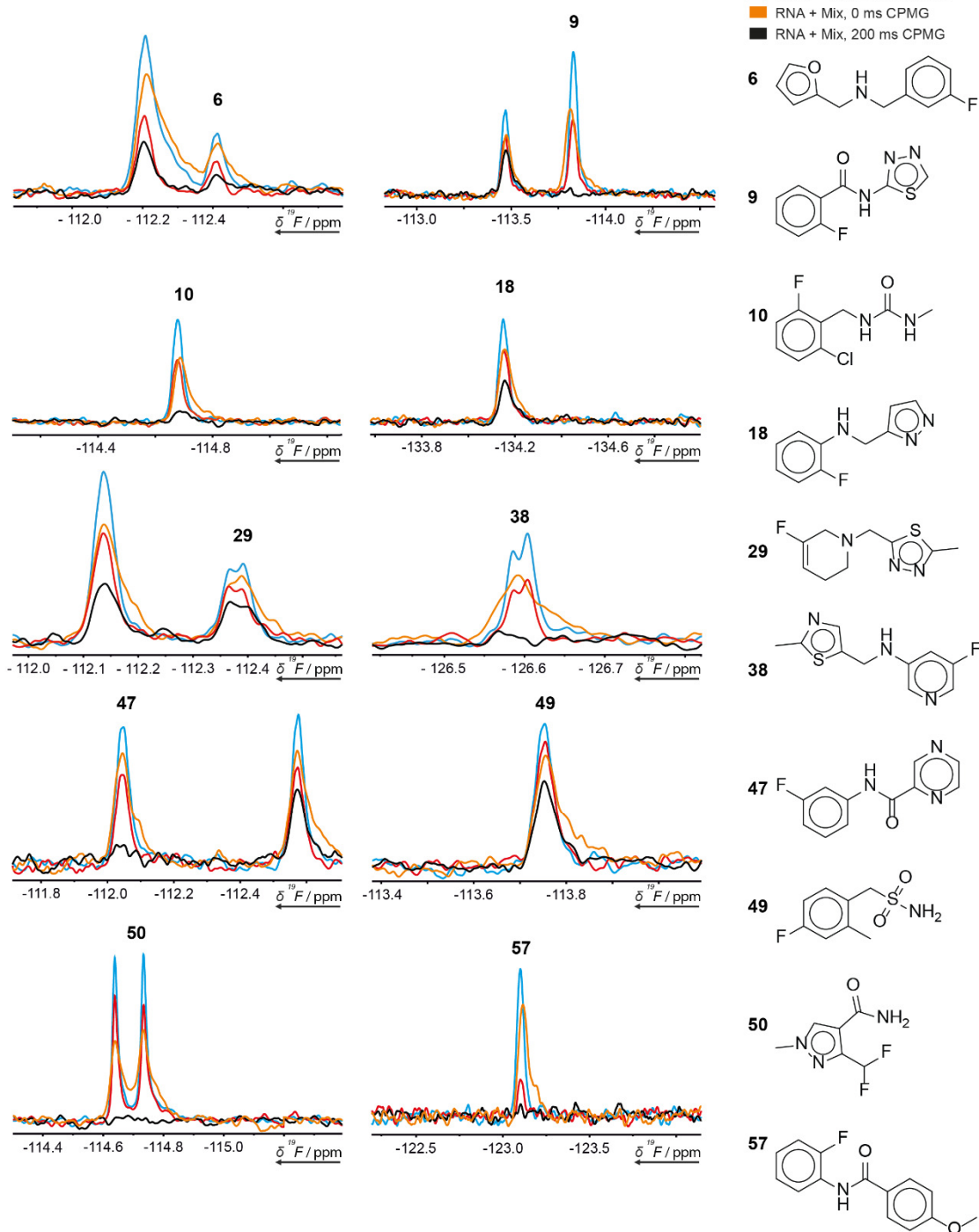
Supplementary Figure 8: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the ZMP-sensing riboswitch (76 nt) from *Thermosinus carboxydivorans*. For further hit validation, the CPMG experiment at 400 ms is shown.

ZMP-sensing riboswitch (76 nt)



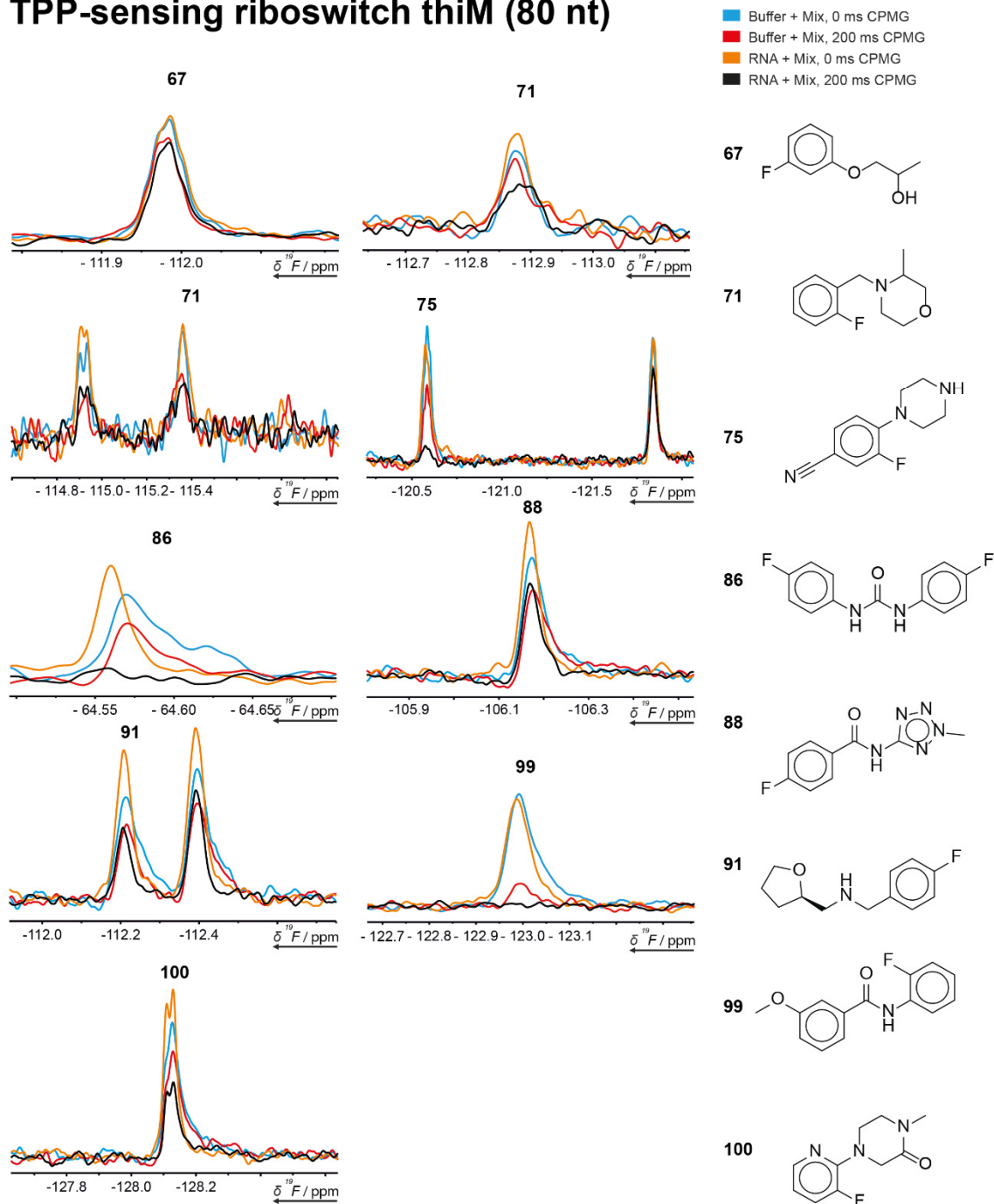
Supplementary Figure 9: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the ZMP-sensing riboswitch (76 nt) from *Thermosinus carboxydivorans*. For further hit validation, the CPMG experiment at 400 ms is shown.

TPP-sensing riboswitch thiM (80)



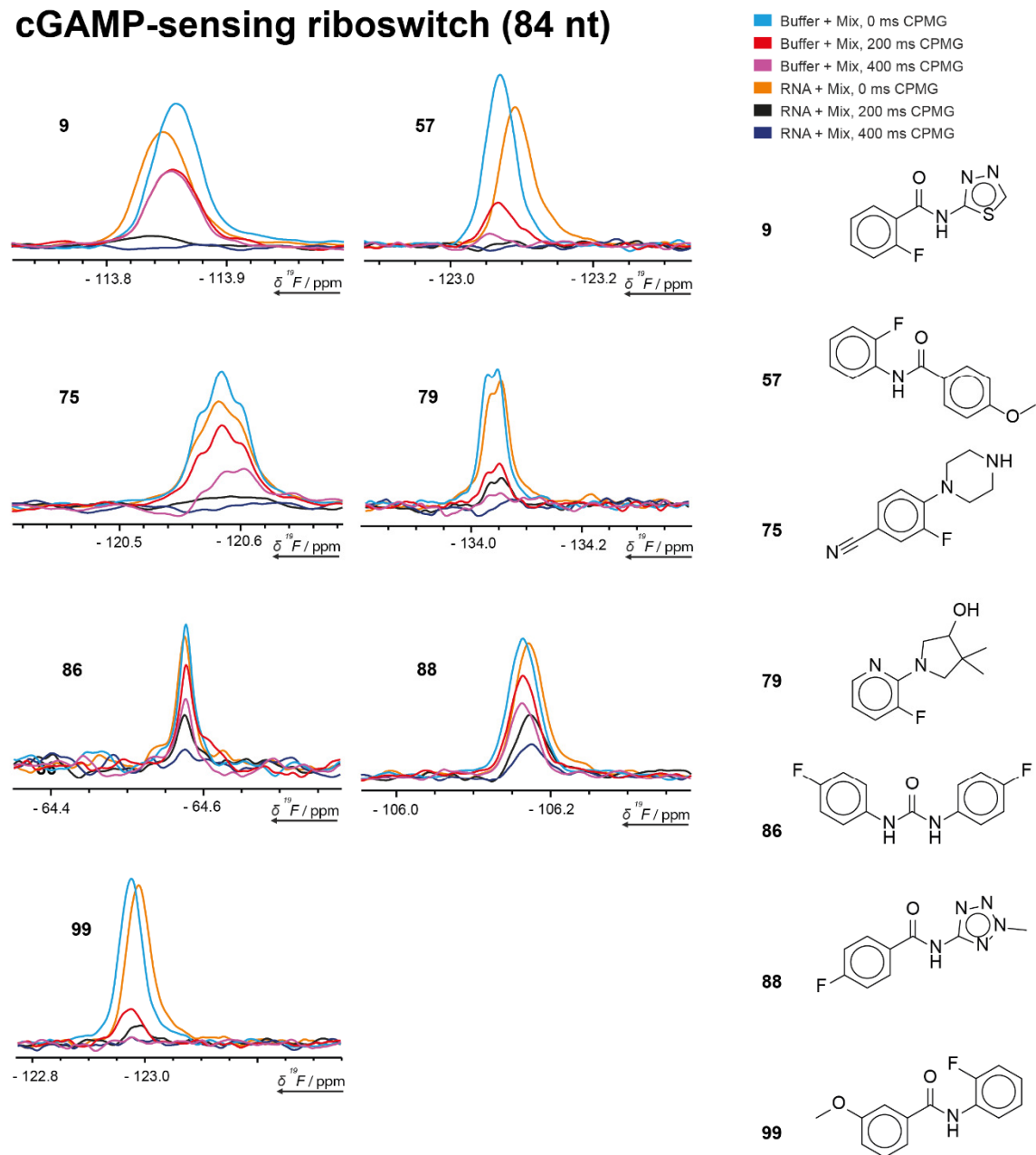
Supplementary Figure 10: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the thiM TPP-sensing riboswitch (80 nt).

TPP-sensing riboswitch thiM (80 nt)



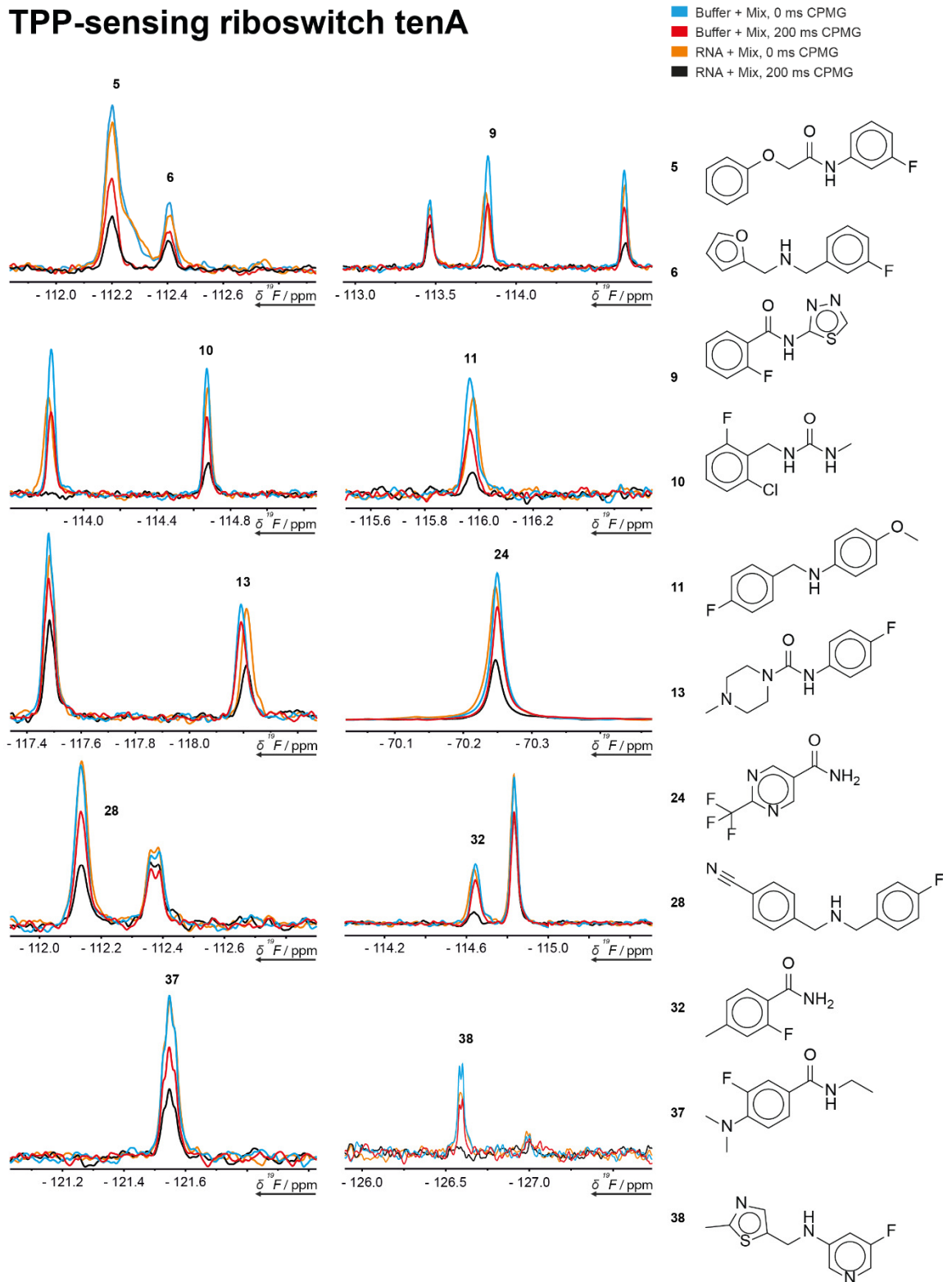
Supplementary Figure 11: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the thiM TPP-sensing riboswitch (80 nt).

cGAMP-sensing riboswitch (84 nt)



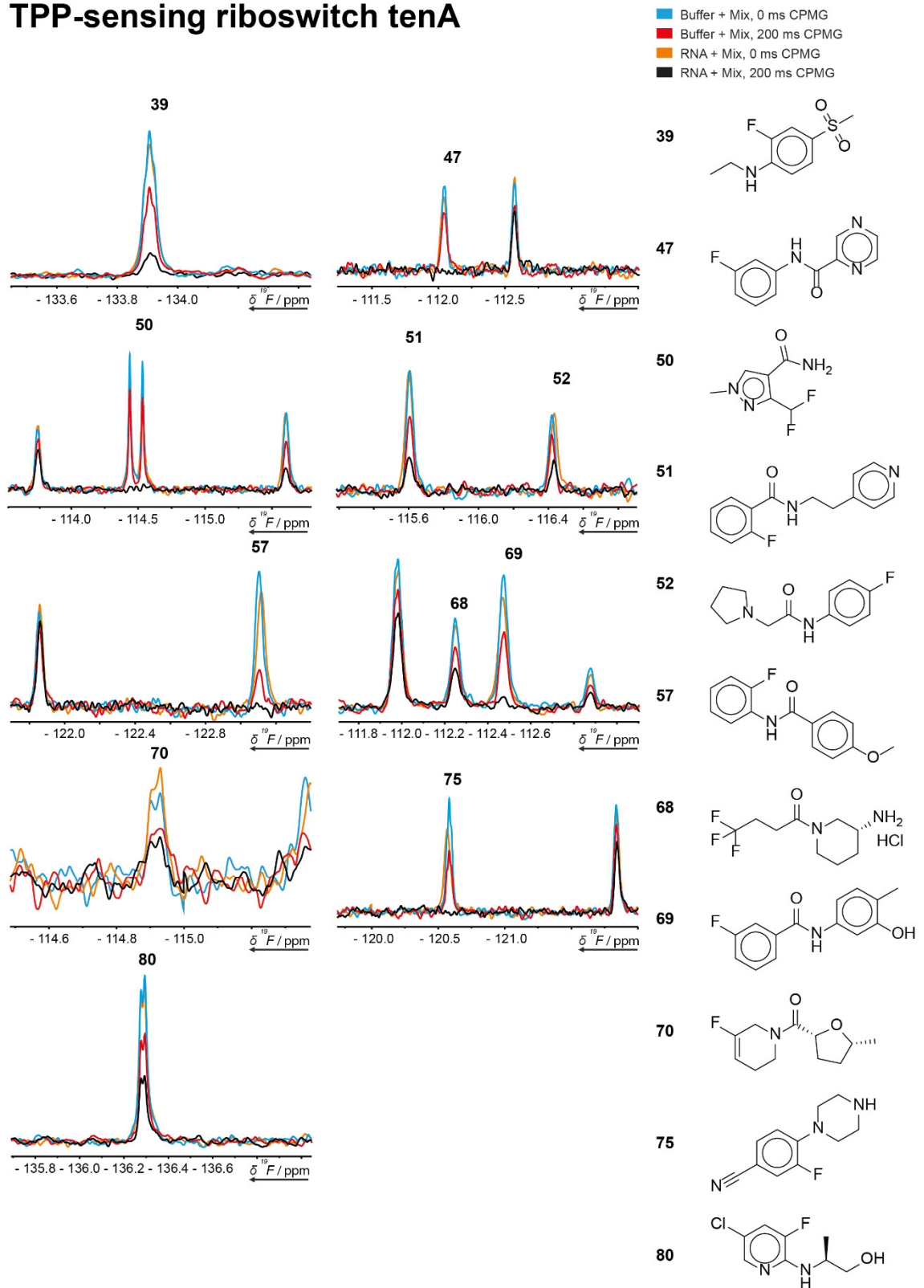
Supplementary Figure 12: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the *pilM* 3',3'-cGAMP-sensing riboswitch (84 nt). For further hit validation, the CPMG experiment at 400 ms is shown.

TPP-sensing riboswitch tenA



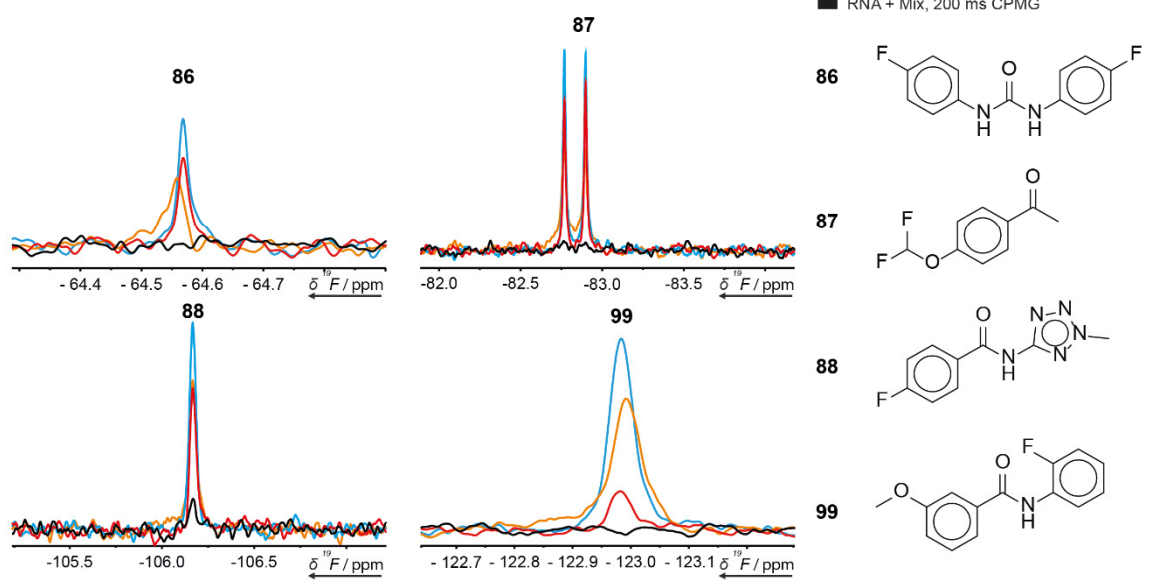
Supplementary Figure 13: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the tenA TPP-sensing riboswitch (94 nt).

TPP-sensing riboswitch tenA



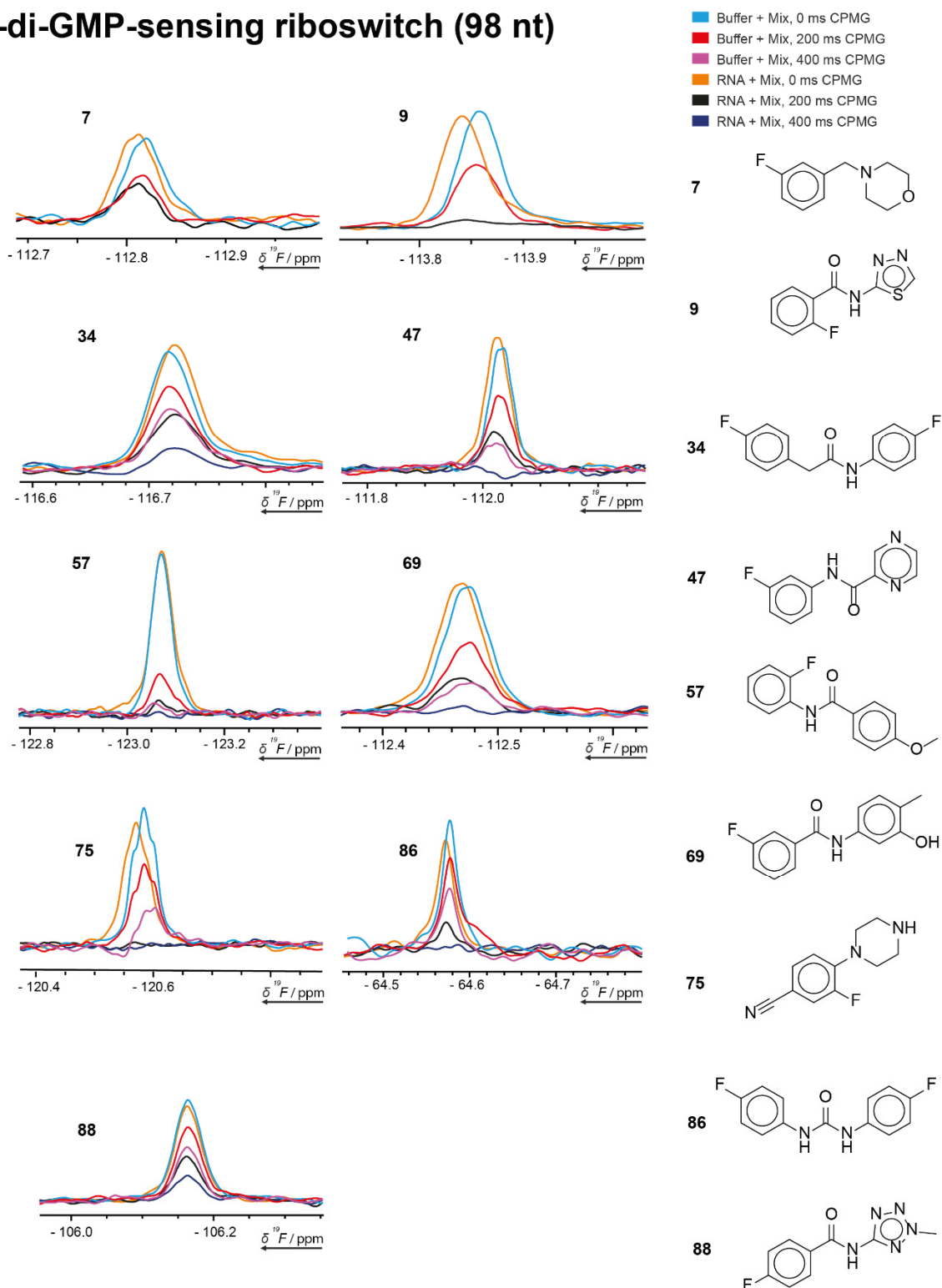
Supplementary Figure 14: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the tenA TPP-sensing riboswitch (94 nt).

TPP-sensing riboswitch tenA



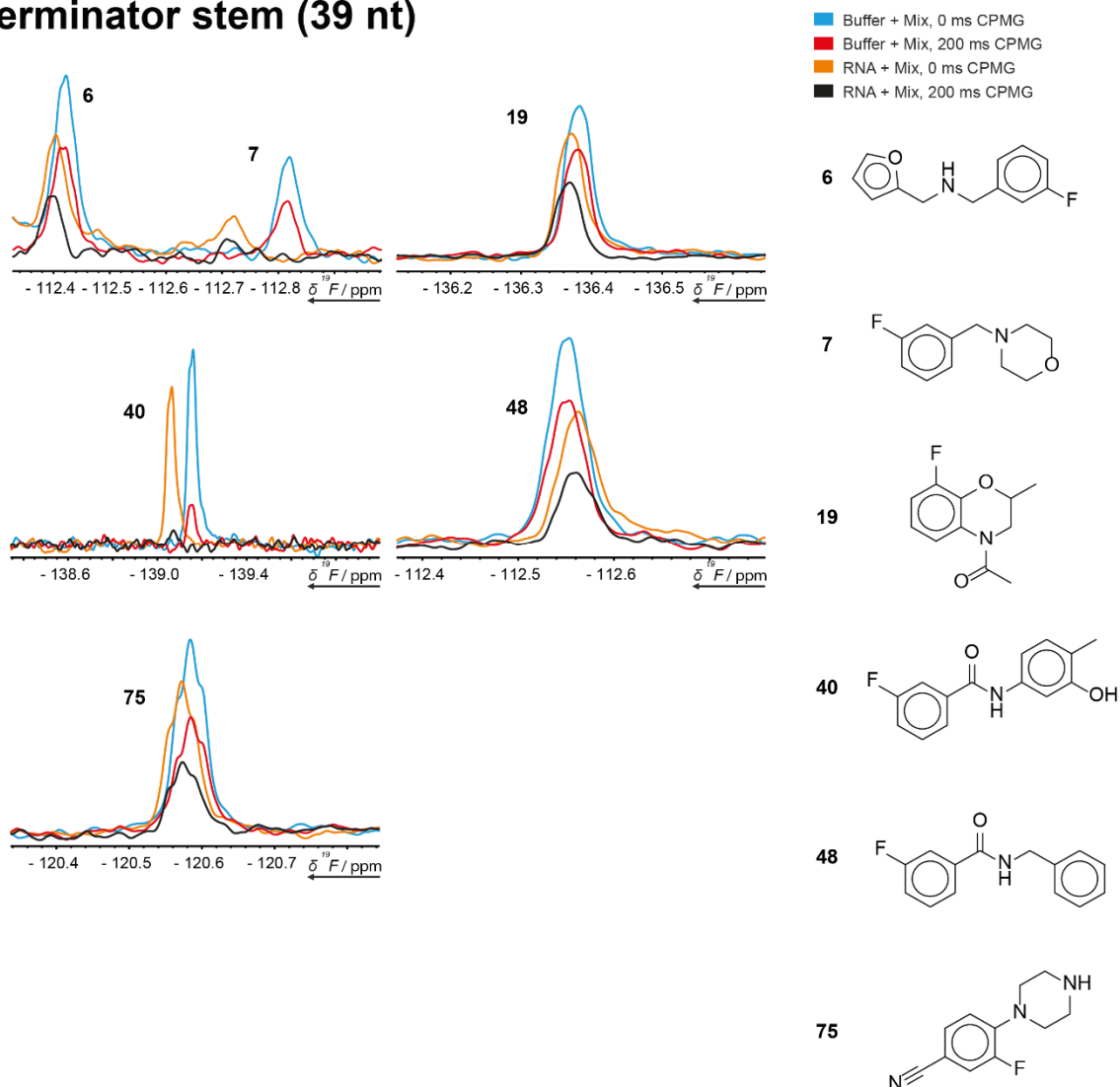
Supplementary Figure 15: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the tenA TPP- sensing riboswitch (94 nt).

c-di-GMP-sensing riboswitch (98 nt)



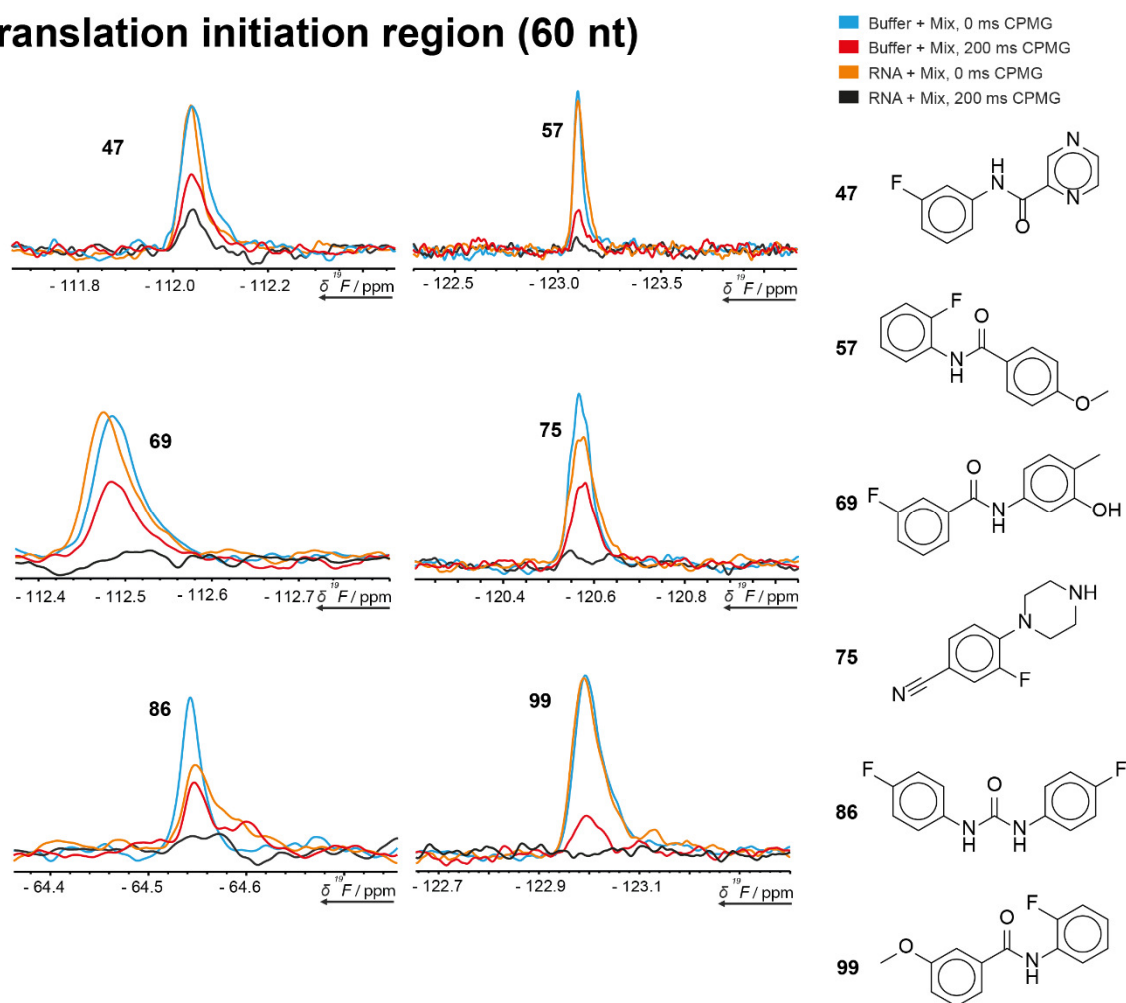
Supplementary Figure 16: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the cyclic di-GMP-1 riboswitch (98 nt). For further hit validation, the CPMG experiment at 400 ms is shown.

Terminator stem (39 nt)



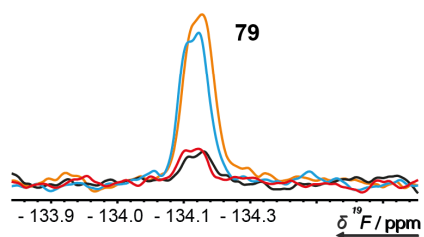
Supplementary Figure 17: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the 2'-deoxyguanosine-sensing-riboswitch terminator (39 nt).

Translation initiation region (60 nt)

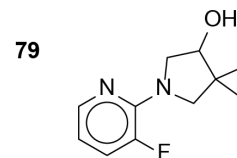


Supplementary Figure 18: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Adenine-sensing riboswitch expression platform (60 nt).

CUUG hairpin (14 nt)

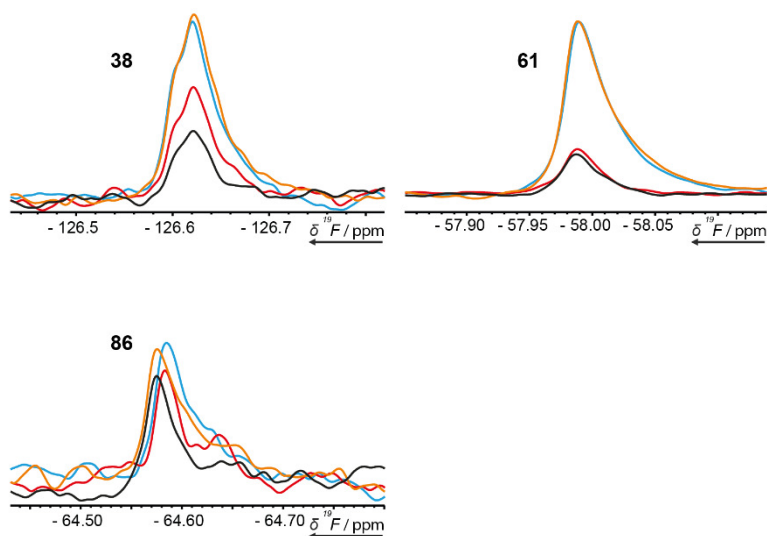


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG

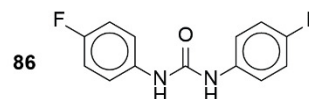
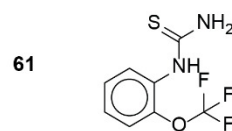
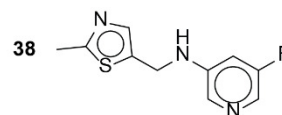


Supplementary Figure 19: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the RNA with CUUG tetraloop (14 nt).

GAAG hairpin (14 nt)

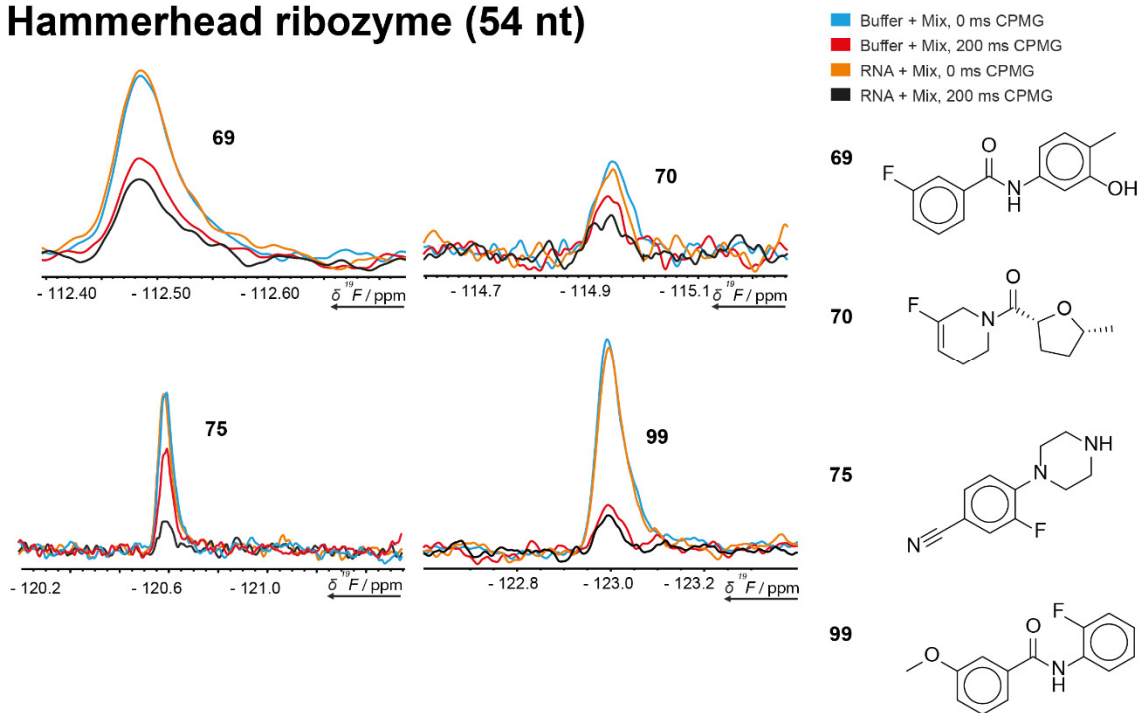


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG



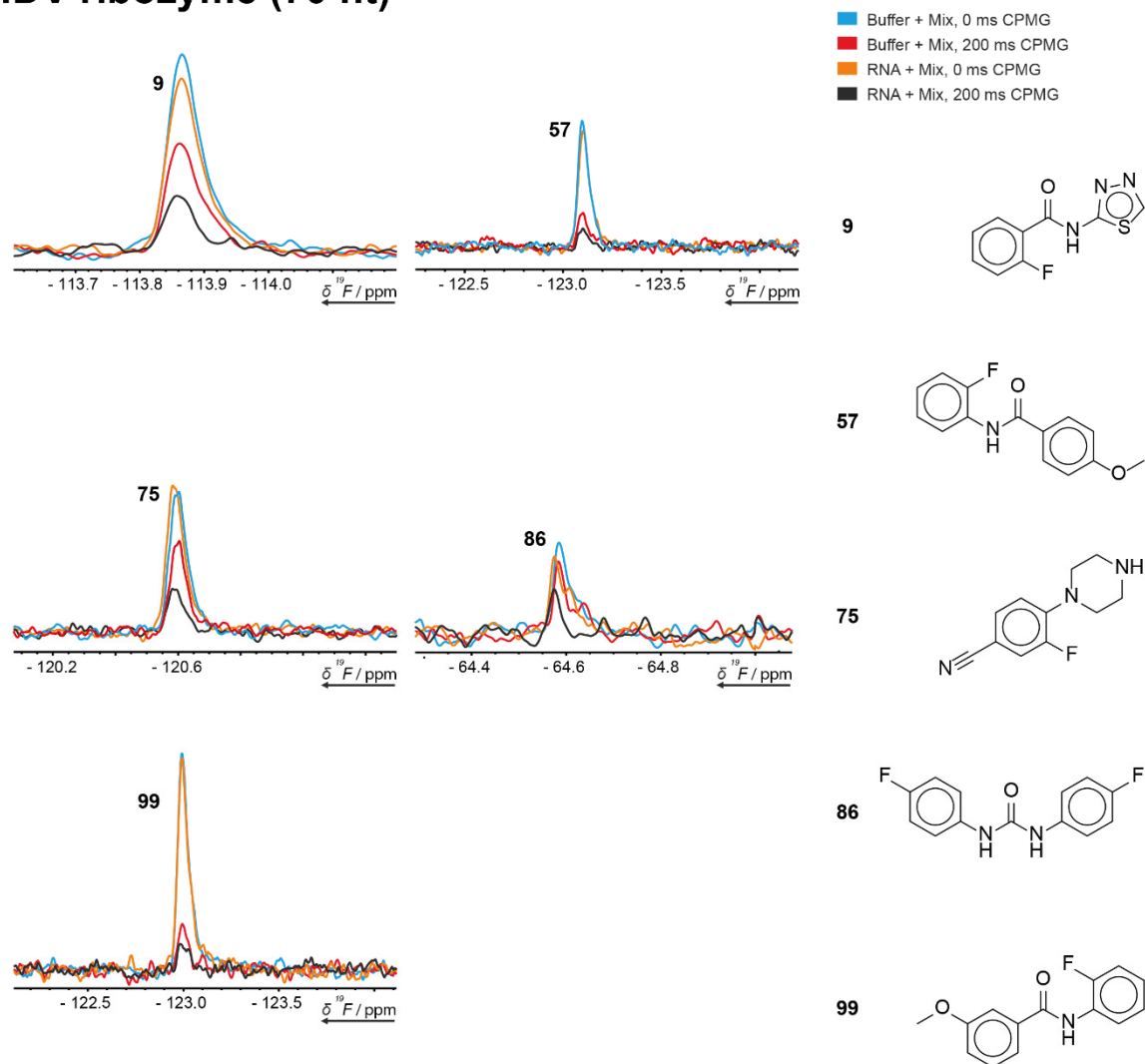
Supplementary Figure 20: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the RNA with GAAG tetraloop (14 nt).

Hammerhead ribozyme (54 nt)



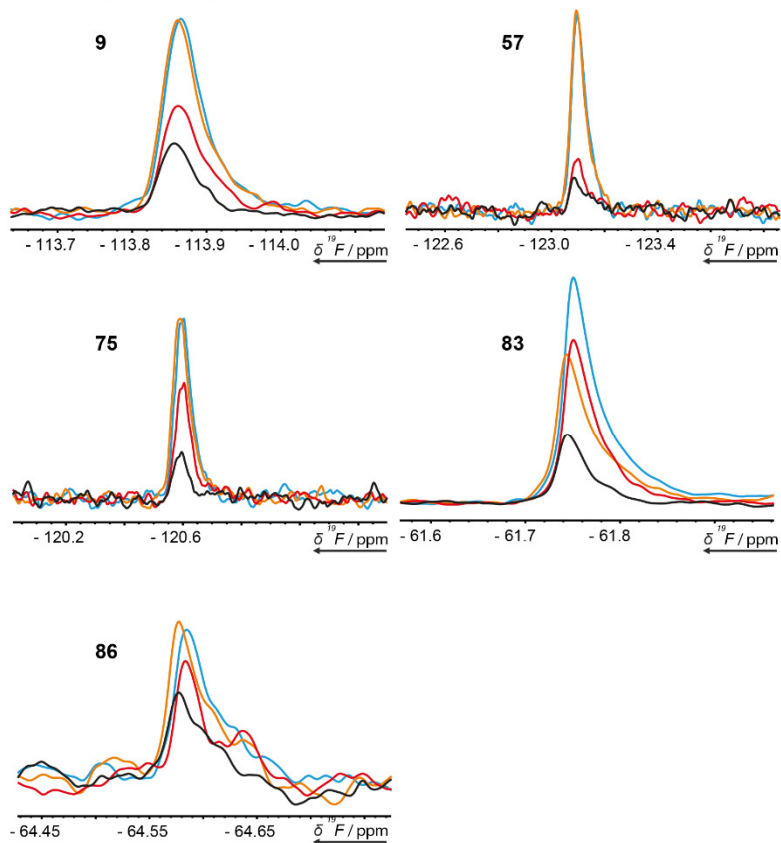
Supplementary Figure 21: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Hammerhead ribozyme (54 nt).

HDV ribozyme (70 nt)

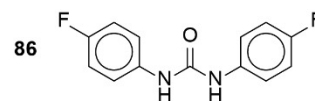
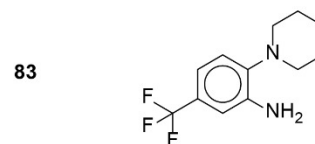
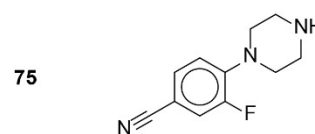
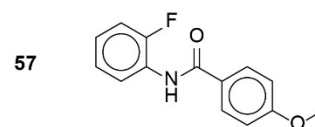
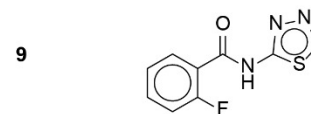


Supplementary Figure 22: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Hepatitis delta virus ribozyme (70 nt).

tRNA (77 nt)

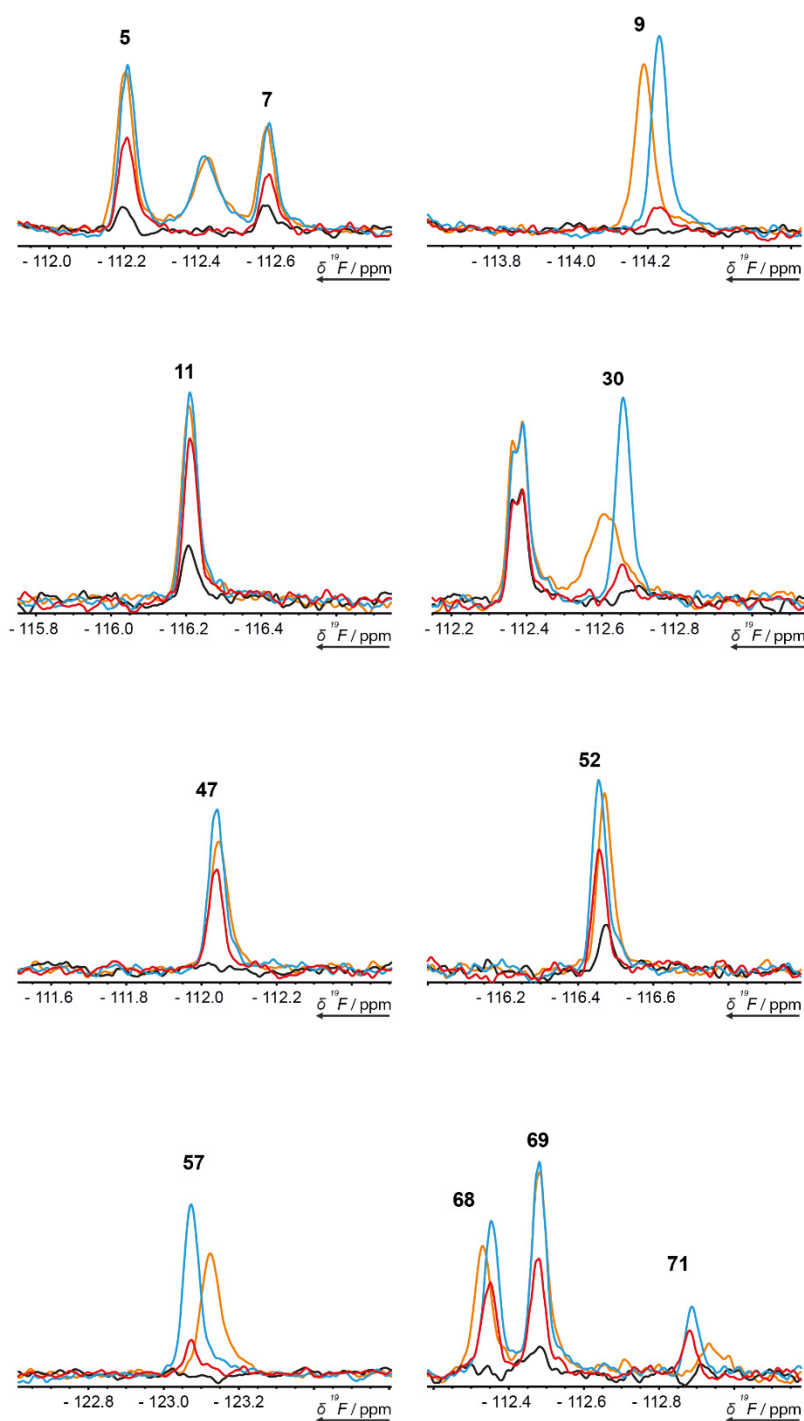


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG

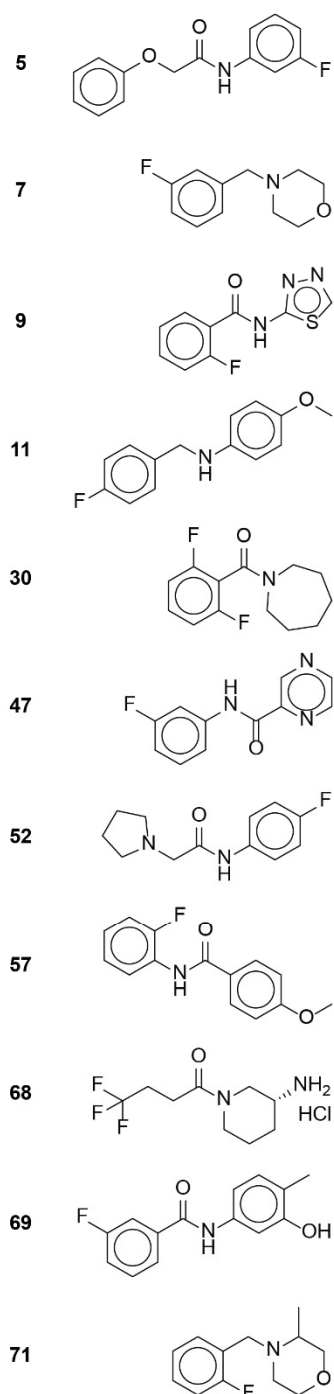


Supplementary Figure 23: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the initiator tRNA^{fMet} (77 nt).

cMyc G-Quadruplex (22 nt)

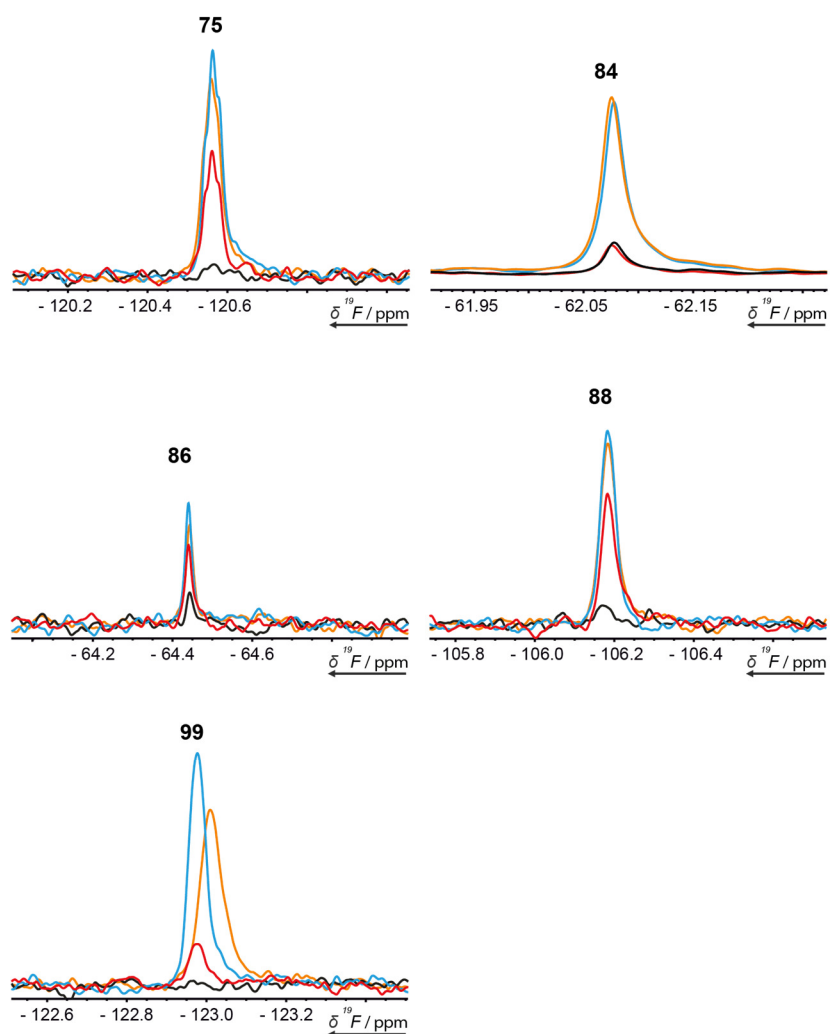


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG

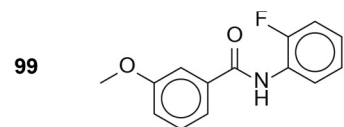
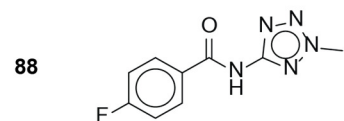
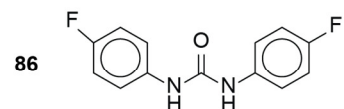
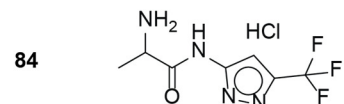
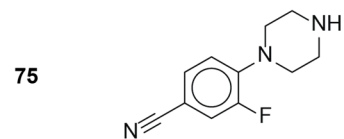


Supplementary Figure 24: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the cMyc G-Quadruplex (22 nt).

cMyc G-Quadruplex (22 nt)

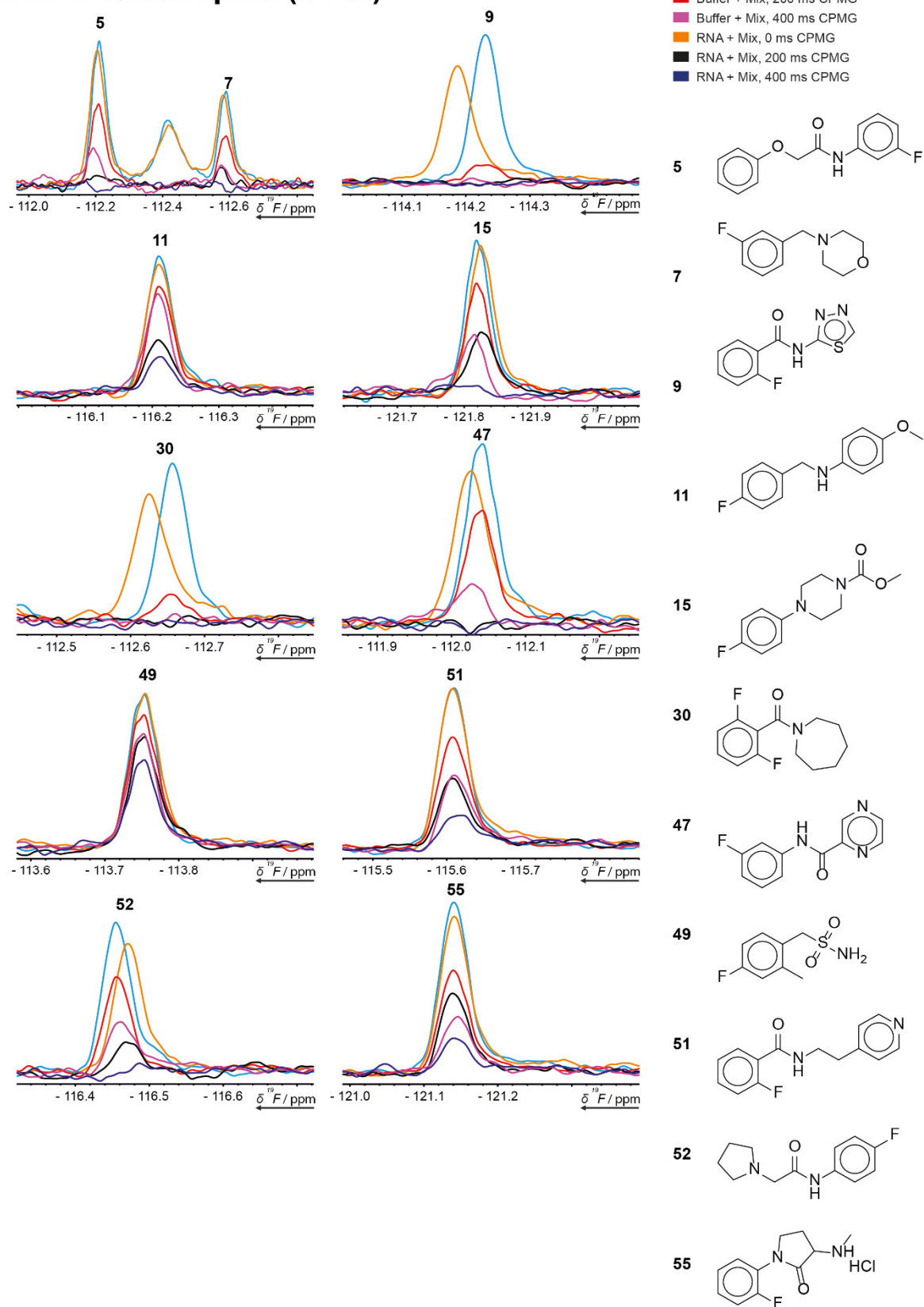


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG



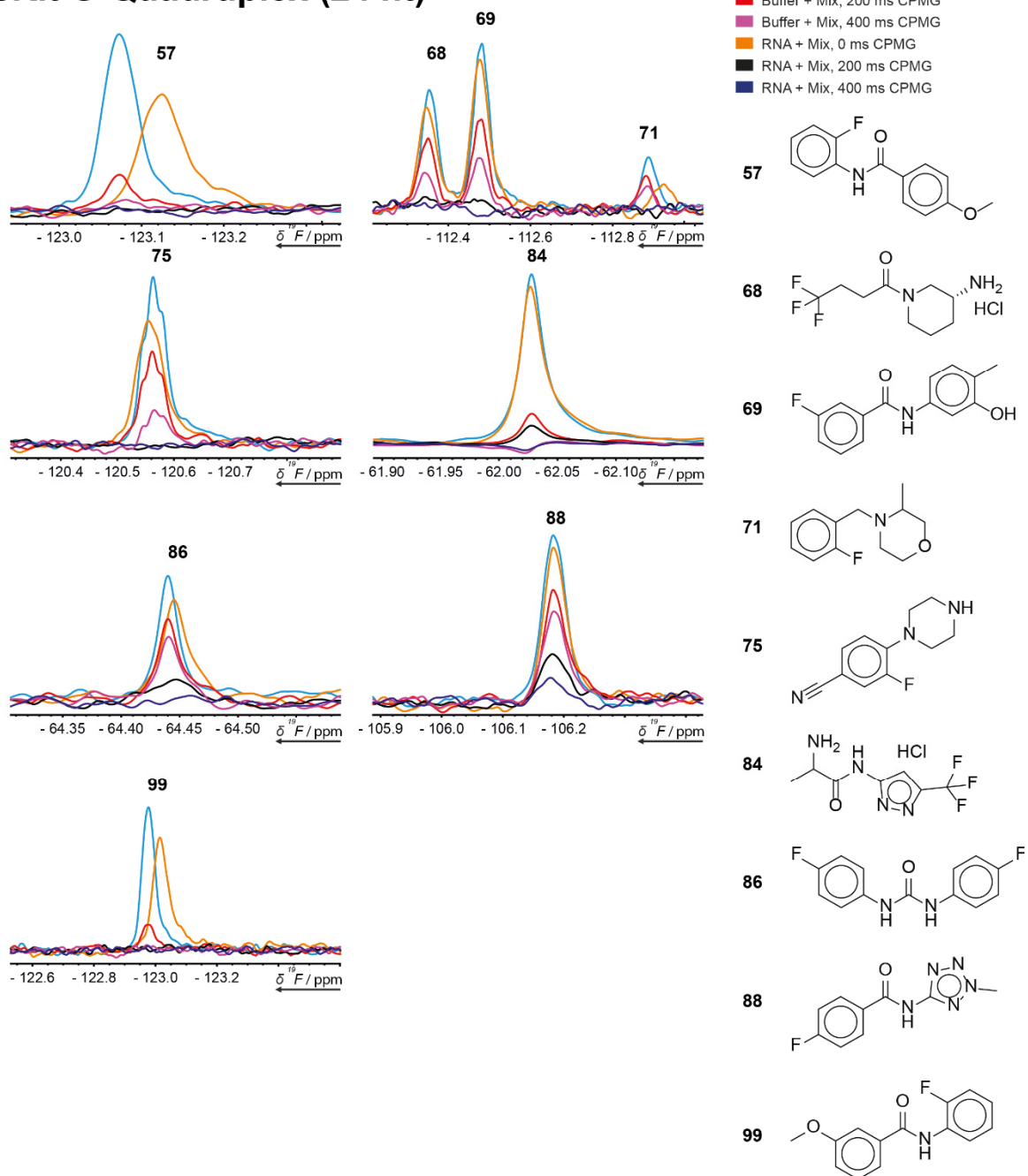
Supplementary Figure 25: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the cMyc G-Quadruplex (22 nt).

cKit G-Quadruplex (24 nt)



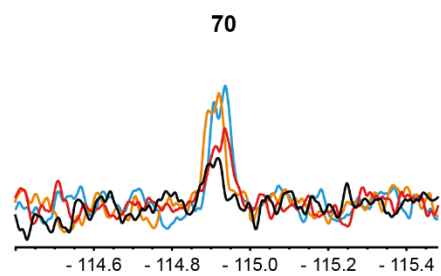
Supplementary Figure 26: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the cKit G-Quadruplex (24 nt). For further hit validation, the CPMG experiment at 400 ms is shown.

cKit G-Quadruplex (24 nt)

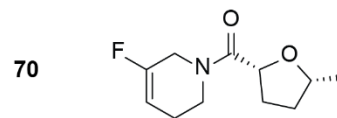


Supplementary Figure 27: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the cKit G-Quadruplex (24 nt). For further hit validation, the CPMG experiment at 400 ms is shown.

DNA duplex (24 nt)

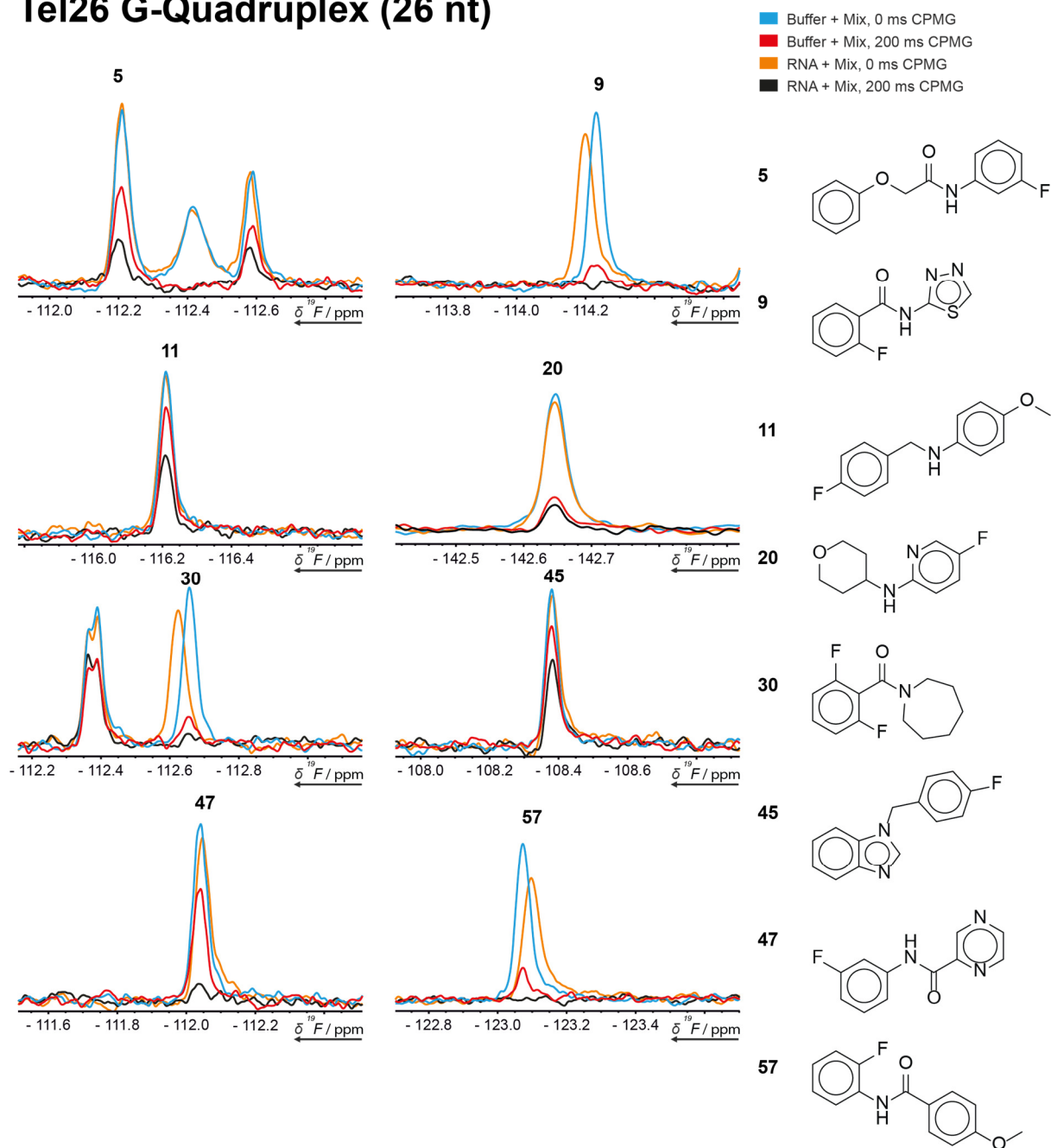


- Buffer + Mix, 0 ms CPMG
- Buffer + Mix, 200 ms CPMG
- RNA + Mix, 0 ms CPMG
- RNA + Mix, 200 ms CPMG



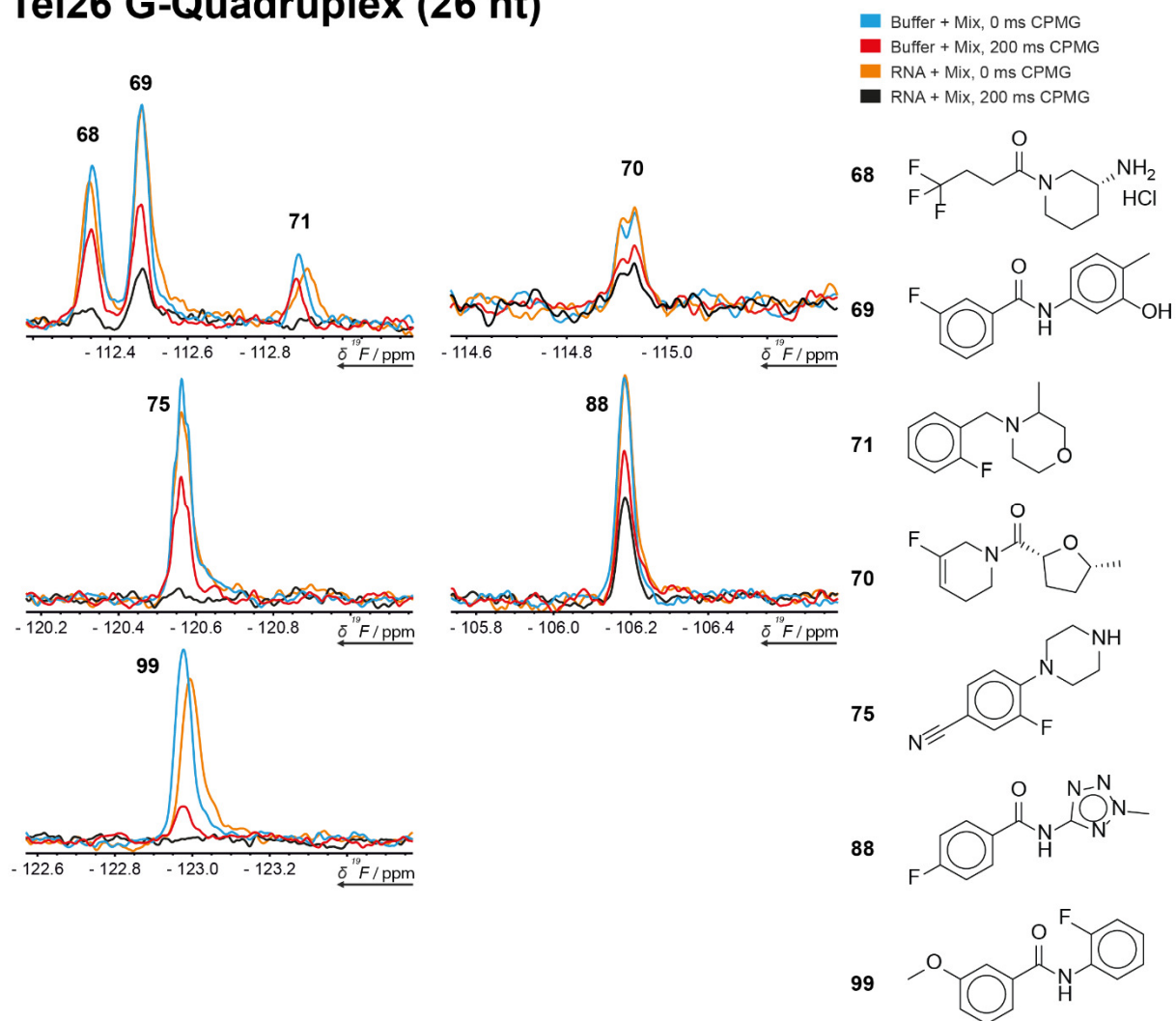
Supplementary Figure 28: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the DNA Duplex (24 nt).

Tel26 G-Quadruplex (26 nt)

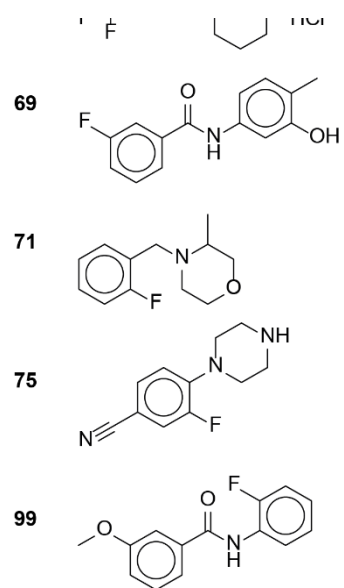
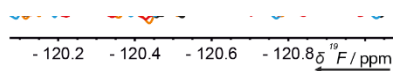
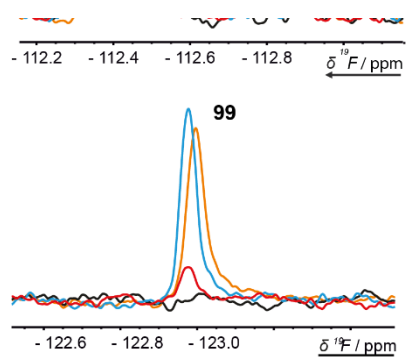


Supplementary Figure 29: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Tel26 G-Quadruplex (26 nt).

Tel26 G-Quadruplex (26 nt)

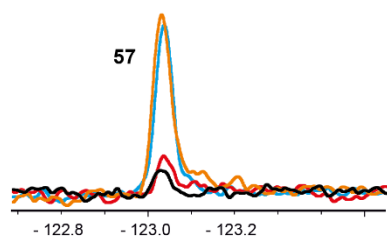


Supplementary Figure 30: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Tel26 G-Quadruplex (26 nt).

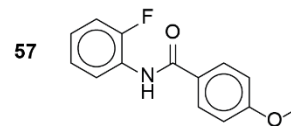


Supplementary Figure 31: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the wtTel26 G-Quadruplex (26 nt).

MptpA (18 kDa)

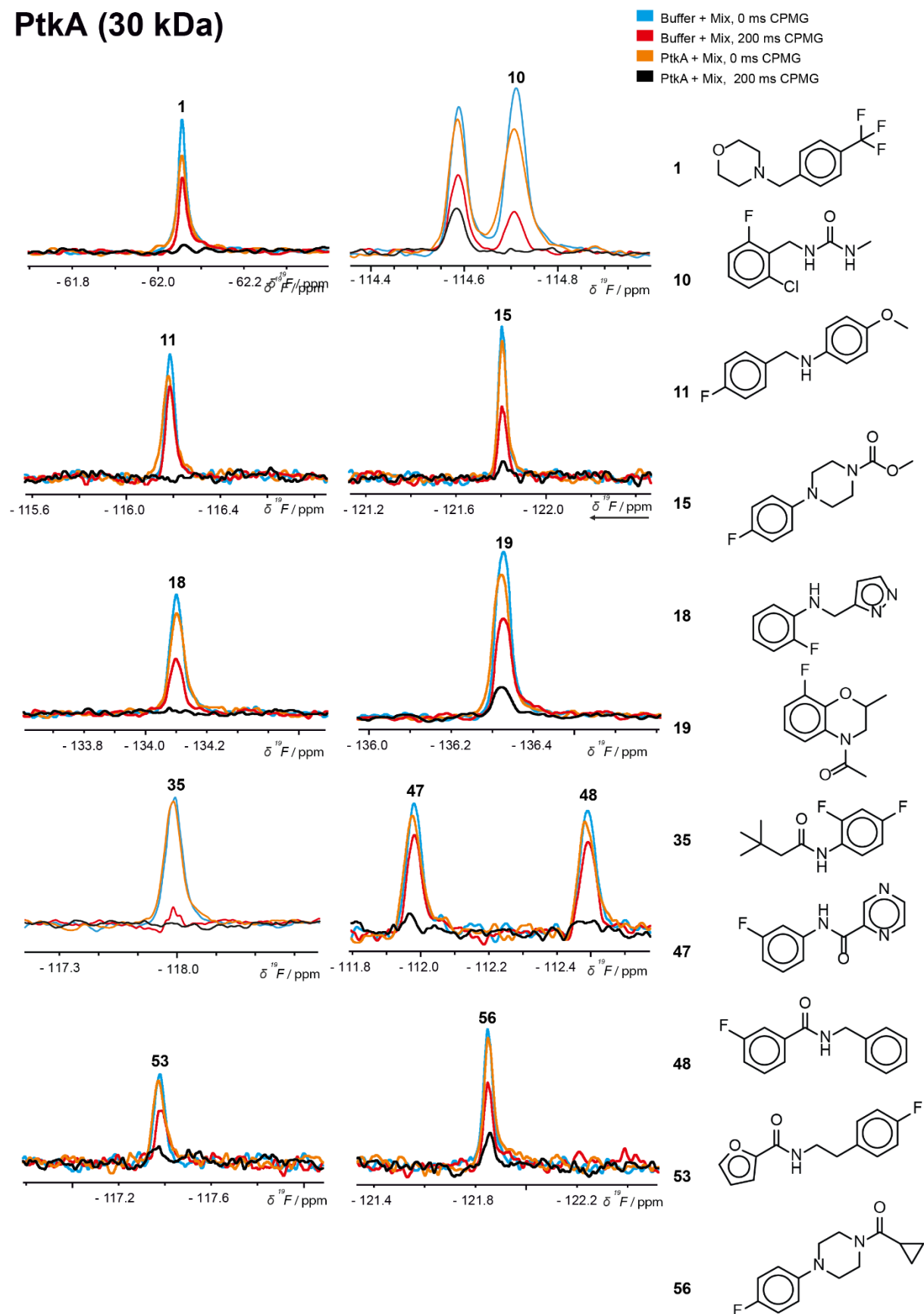


- Buffer + Mix C, 0 ms CPMG
- Buffer + Mix C, 200 ms CPMG
- MptpA + Mix C, 0 ms CPMG
- MptpA + Mix C, 200 ms CPMG



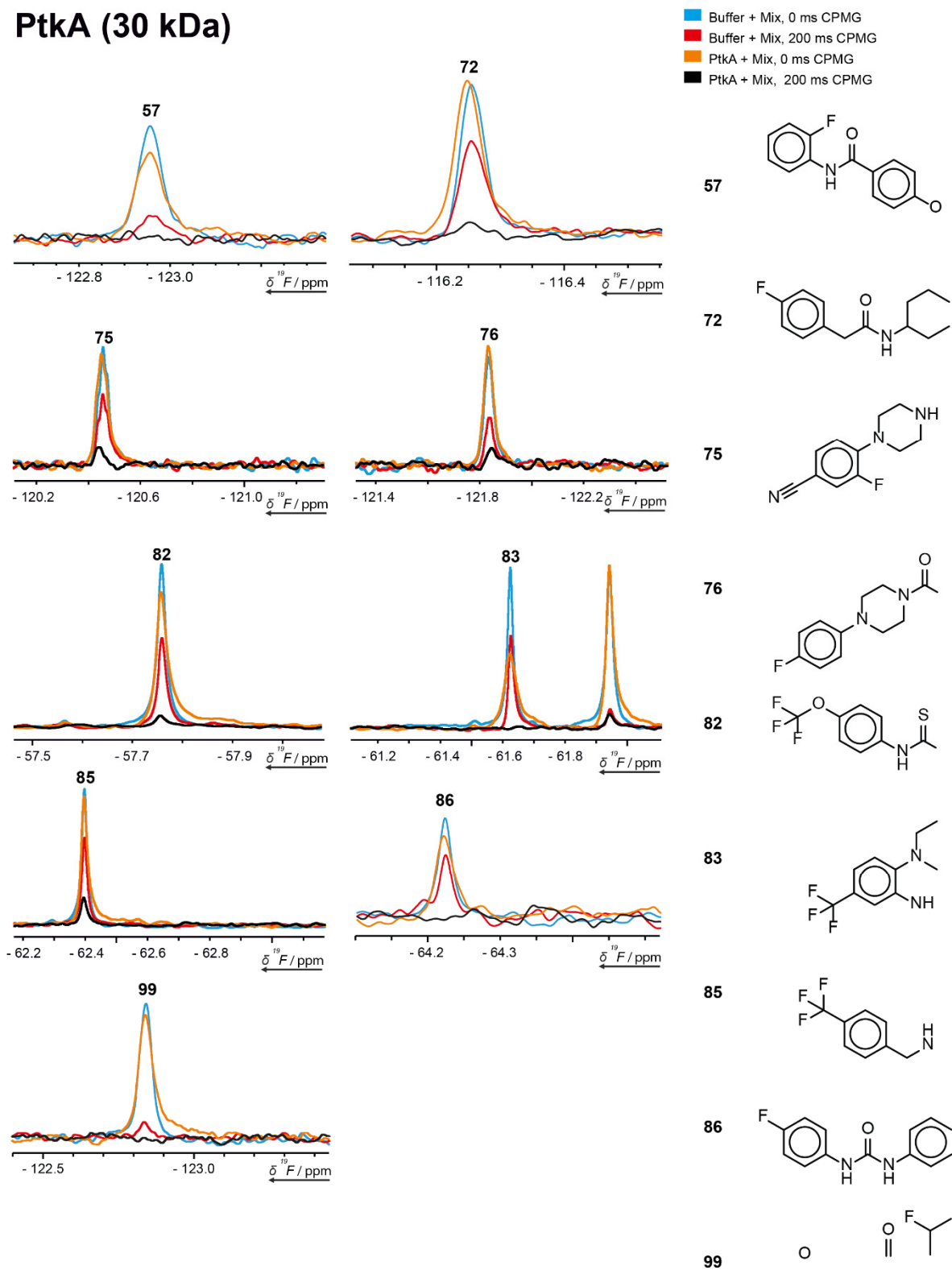
Supplementary Figure 32: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the *Mycobacterium tuberculosis* Protein Tyrosine Phosphatase A (MptpA, 18 kDa).

PtkA (30 kDa)



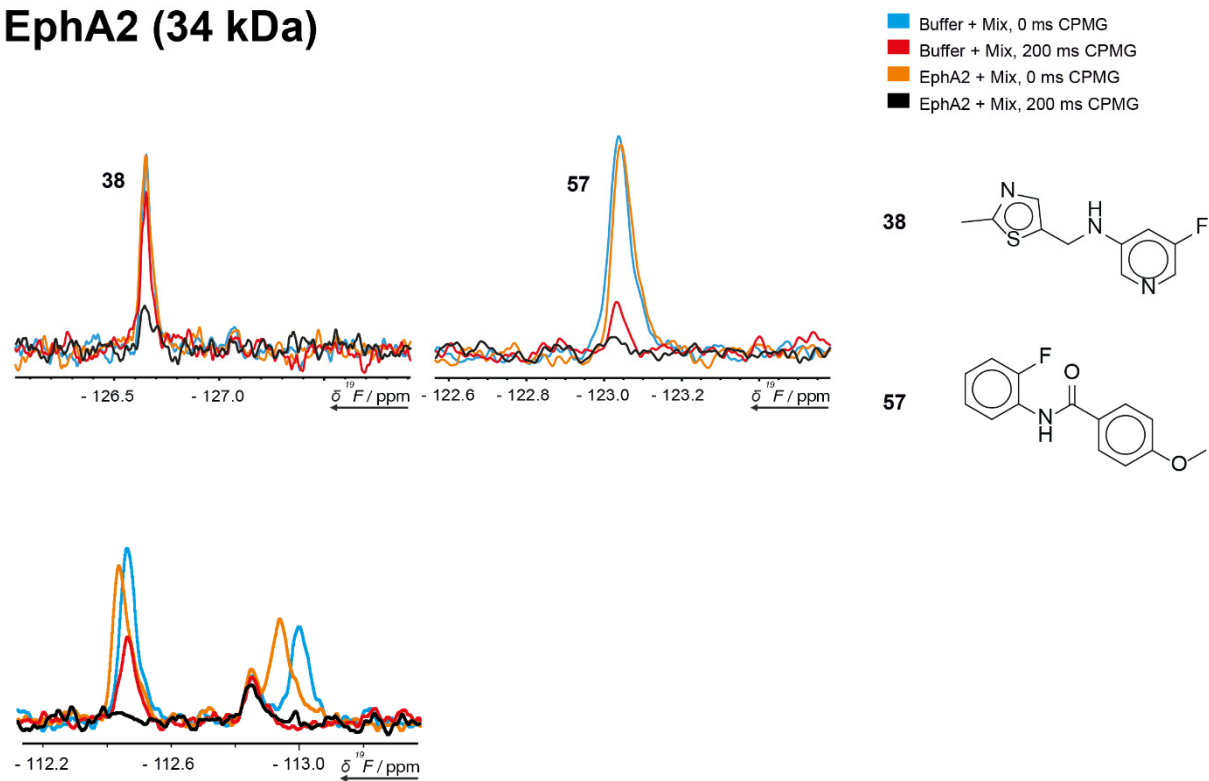
Supplementary Figure 33: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Protein tyrosine Kinase A (PtkA, 30 kDa).

PtkA (30 kDa)



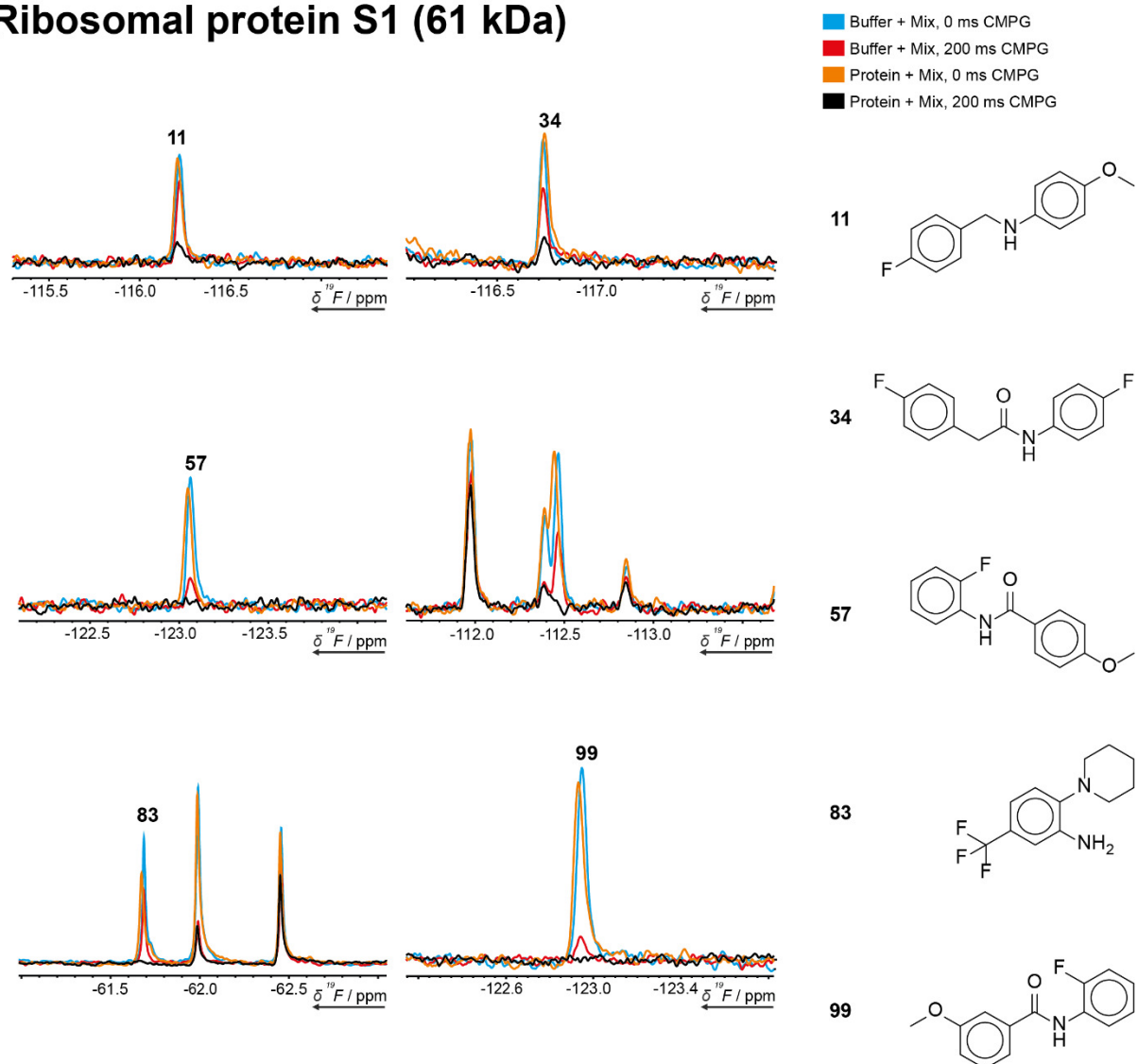
Supplementary Figure 34: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the *Mycobacterium tuberculosis* Protein Tyrosine Phosphatase A (MptpA, 18 kDa).

EphA2 (34 kDa)



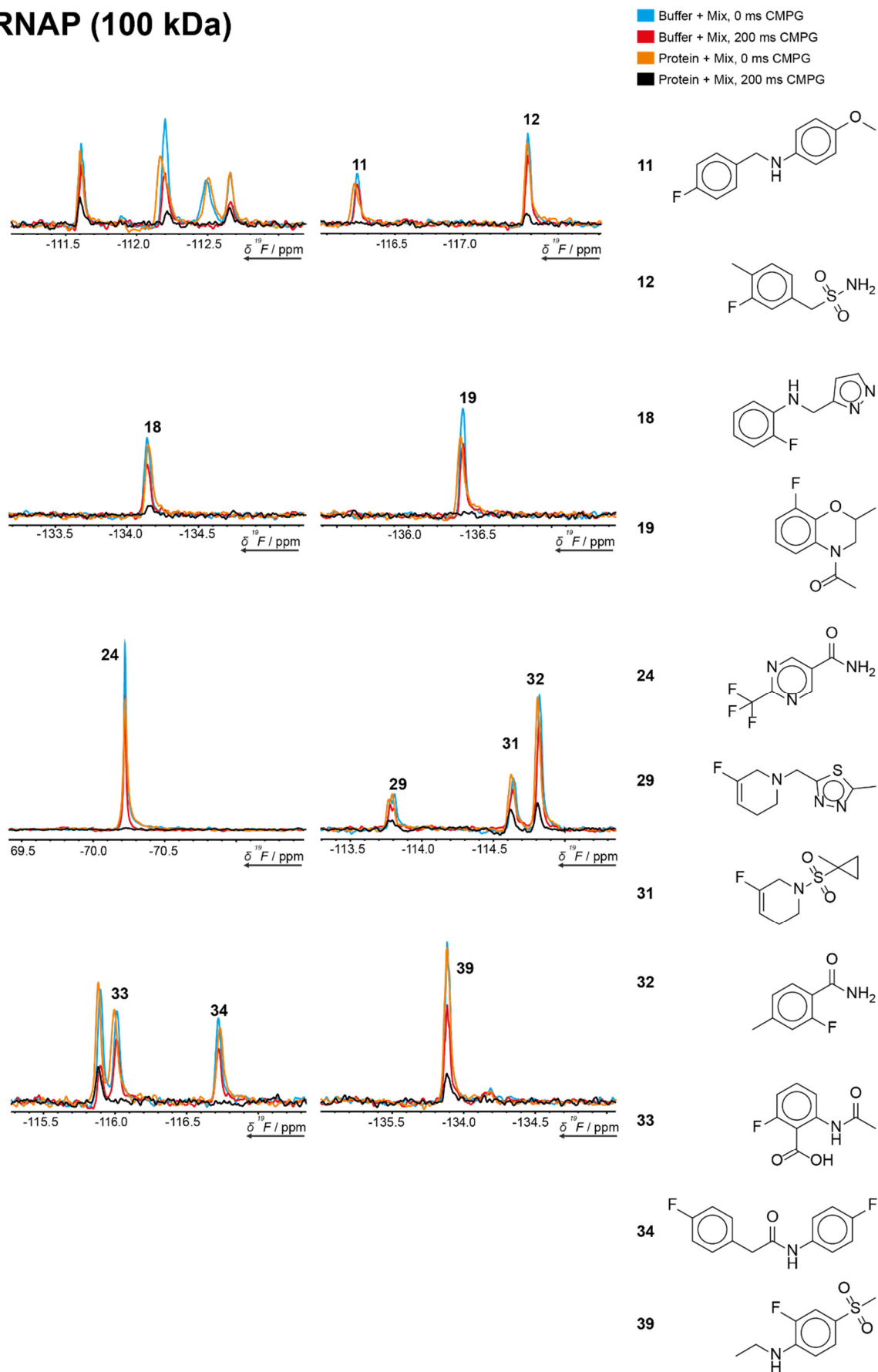
Supplementary Figure 35: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Receptor tyrosine kinase EphA2 (61 kDa).

Ribosomal protein S1 (61 kDa)



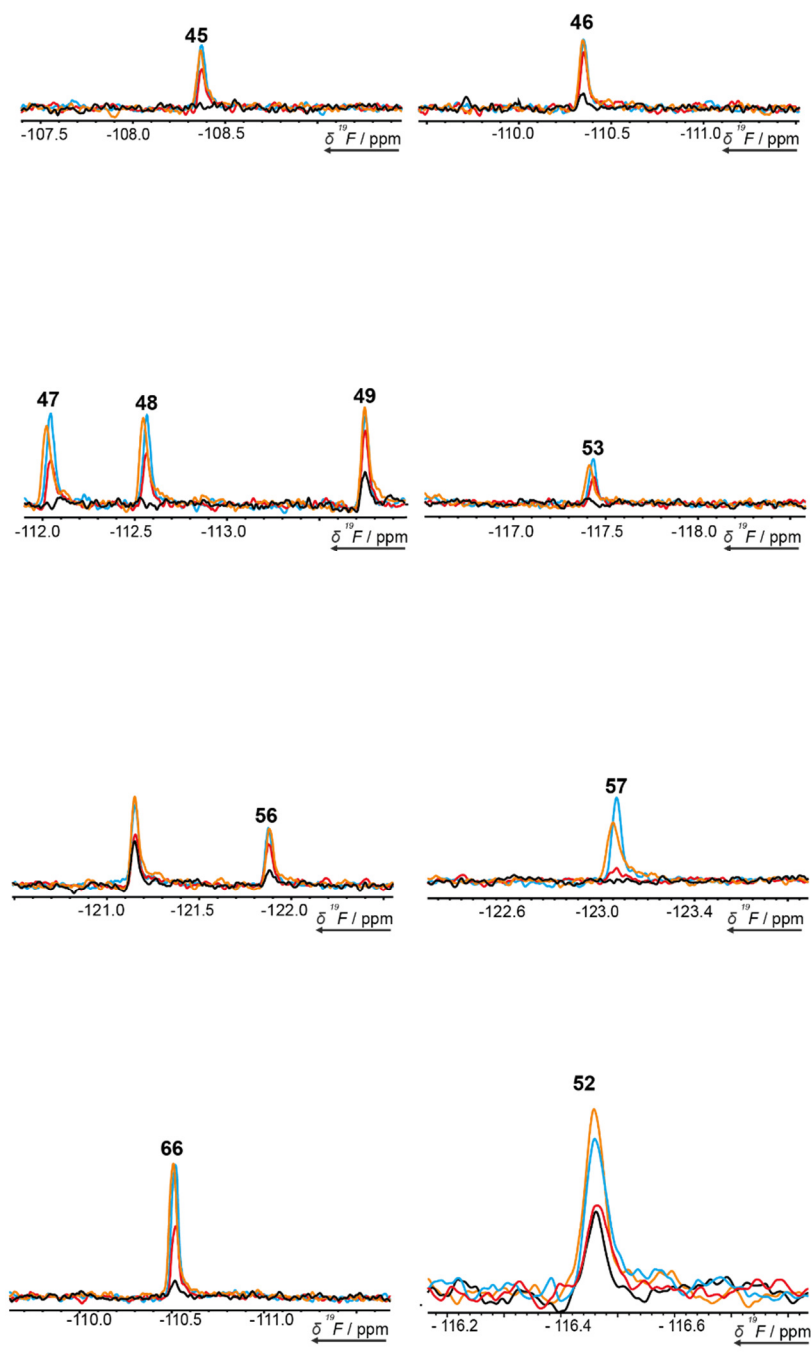
Supplementary Figure 36: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the Ribosomal protein S1 (61 kDa).

T7 RNAP (100 kDa)

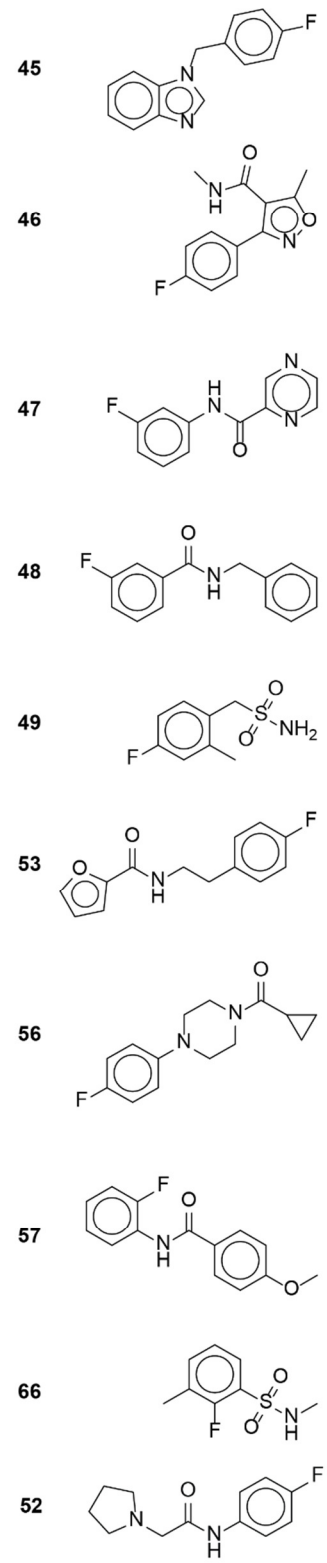


Supplementary Figure 37: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the T7 RNA polymerase (100 kDa).

T7 RNAP (100 kDa)

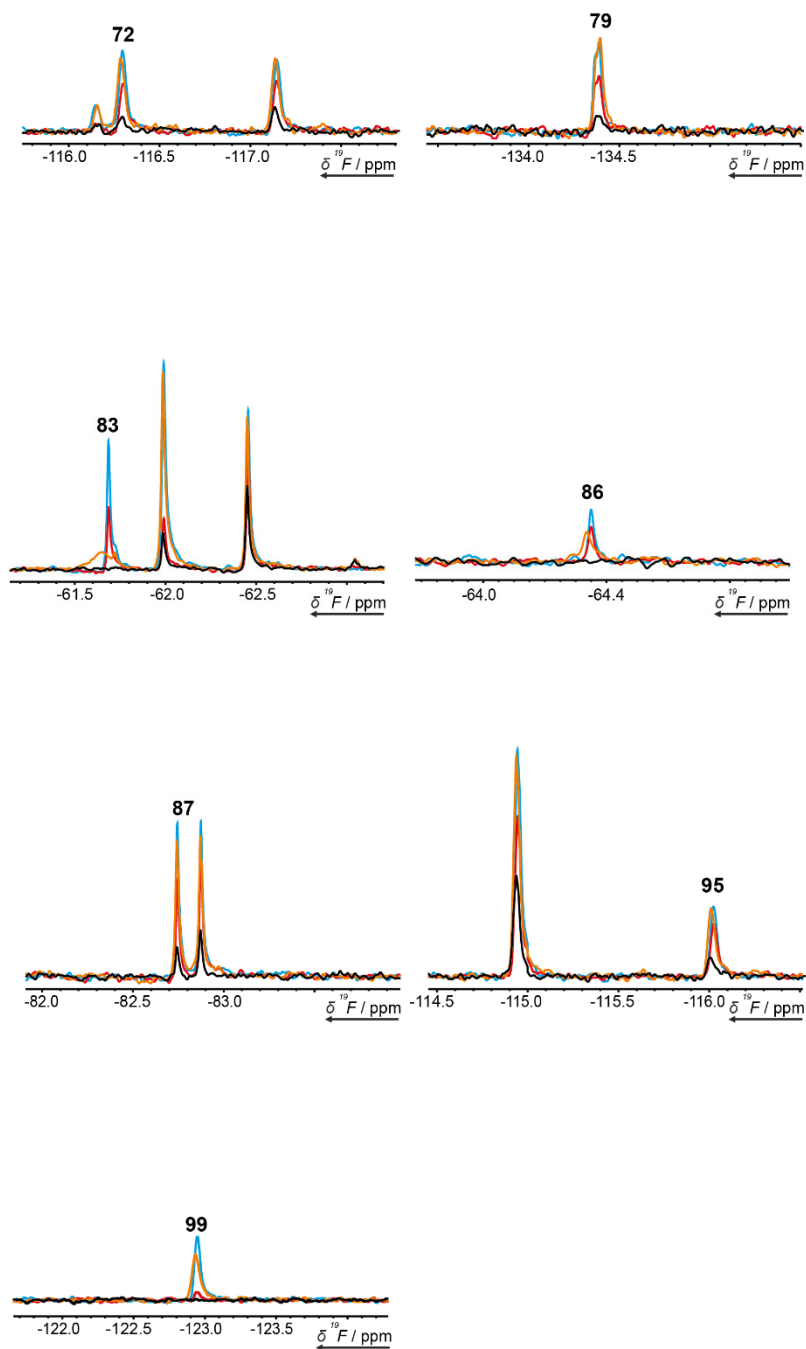


■ Buffer + Mix, 0 ms CPMG
■ Buffer + Mix, 200 ms CPMG
■ Protein + Mix, 0 ms CPMG
■ Protein + Mix, 200 ms CPMG

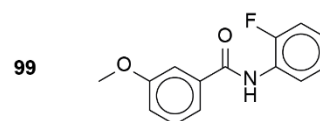
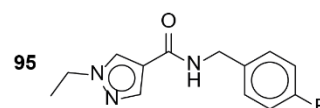
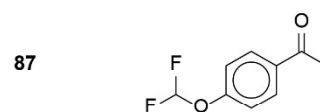
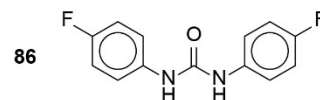
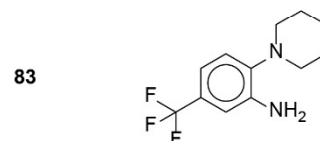
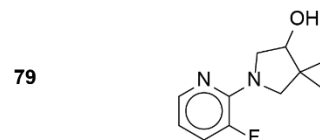
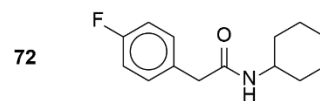


Supplementary Figure 38: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the T7 RNA polymerase (100 kDa).

T7 RNAP (100 kDa)



■ Buffer + Mix, 0 ms CPMG
■ Buffer + Mix, 200 ms CPMG
■ Protein + Mix, 0 ms CPMG
■ Protein + Mix, 200 ms CPMG



Supplementary Figure 39: Spectral excerpts displaying the intensity modulation obtained in the 200 ms CPMG experiment against 0 ms CPMG identifying the respective fragments as target hits for the T7 RNA polymerase (100 kDa).

Follow up Chemistry

We report here the general strategy for targeting the terminator and antiterminator structural elements found in riboswitches acting at the level of transcription. As a proof of concept, we chose the 39 nt terminator stem as drug target, since it represents the smallest RNA structure containing a druggable bulge motif. For technical reasons, we complemented the screening of the ^{19}F -library with a ^1H -library covering 768 analyzed fragments. Aside from the 6 weak and 2 strong hit found in ^{19}F screening, we found 27 additional hits via ^1H screening. For the ease of chemistry and commercial availability and affordability of precursors, we focused on a fragment hit containing a benzamide (P2D11) for further modification, which is a close homologue of fragment 48, which showed effect during the ^{19}F -screening. This fragment was linked to an acridine moiety to enhance to binding affinity. Furthermore, the fluorescence of acridine enables fluorescence-based binding assays (Supplementary Figure 56 - Supplementary Figure 61). We performed fluorescence titration assays of the coupled derivative with the herein investigated 39 nt terminator stem. In order to assess selectivity variations towards different RNA structures, we further included a 38 nt antiterminator stem with a large loop as putative drug binding site and a 51 nt terminator stem containing an internal loop. The coupled fragment exhibited a K_D of 15.3 μM towards the 39 nt terminator stem, illustrating that low μM affinity can be established by initial hit modification. Moreover, the coupled fragment showed up to a ~ 15 -fold selectivity for different RNA structures (K_D (38 nt) = 1.1 μM ; K_D (51 nt) = 1.5 μM).

Synthesis

General experimental procedures

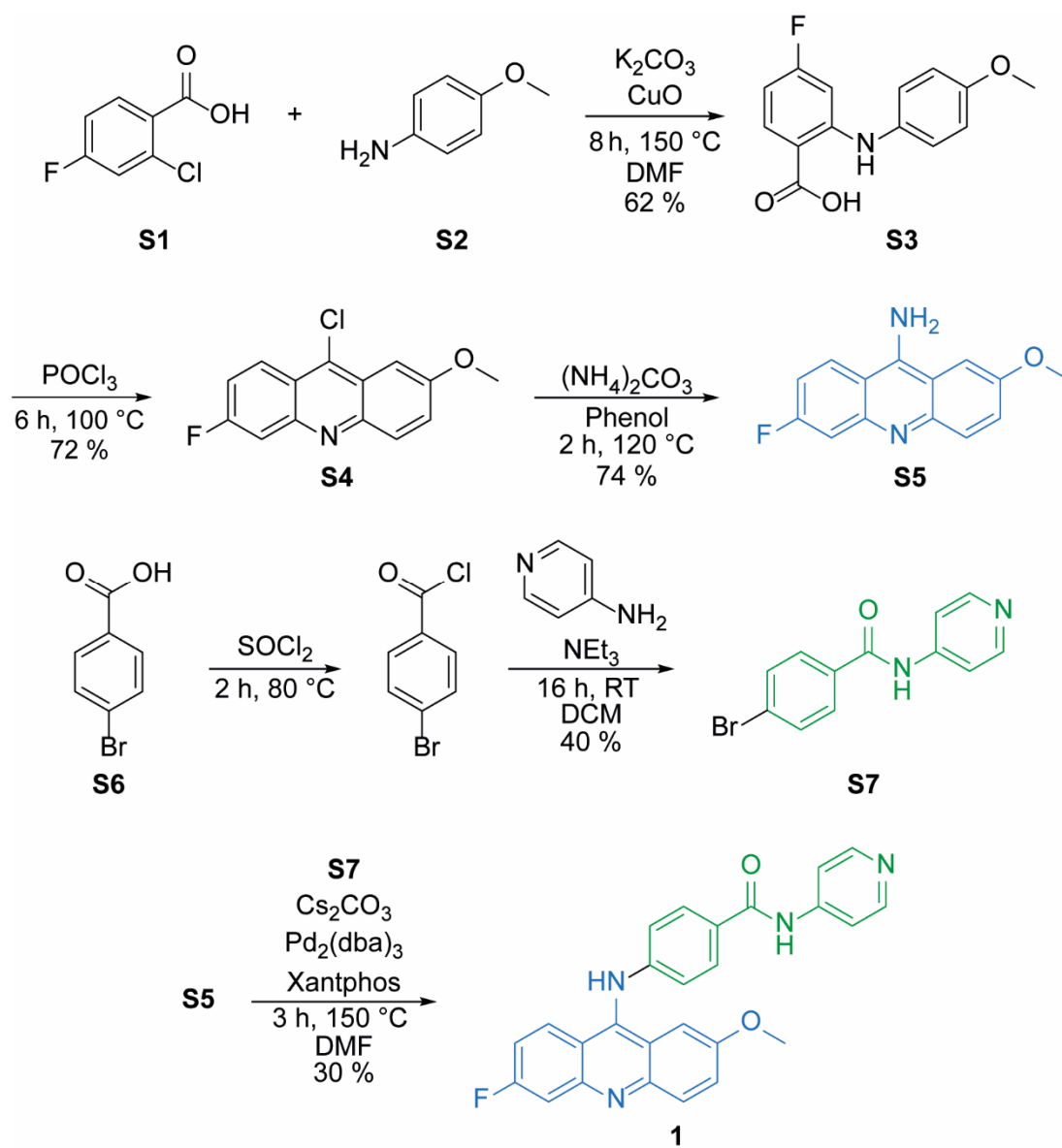
Reactions were conducted under inert conditions if necessary, glassware was dried beforehand. Starting materials were commercially available and used without further purification. POCl₃ und SOCl₂ were distilled before usage. Anhydrous solvents were purchased in crown-capped bottles under argon atmosphere, 4 Å molecular sieves were used for storage when needed.

Alugram® Xtra Sil G UV254 silica gel plates from *Macherey-Nagel* were used for performing thin layer chromatography (TLC). For column chromatography silica gel (Silica 60, 0.04-0.063 mm) from *Macherey-Nagel* was used.

Mass spectra were recorded on a *Thermo Fisher* Surveyor MSQ spectrometer.

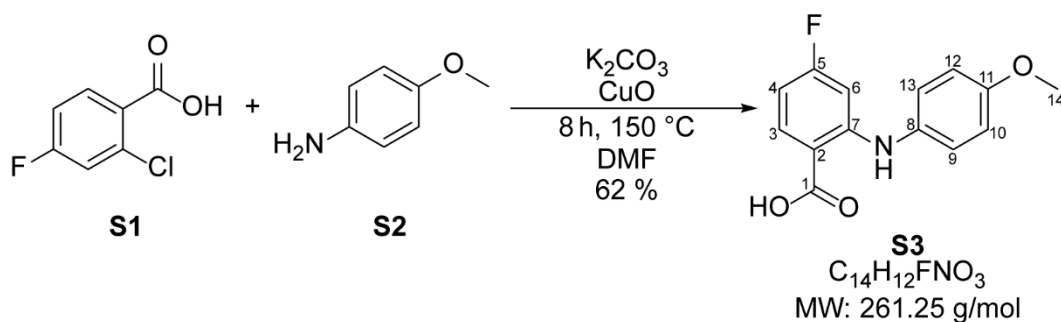
All shown NMR-spectra were recorded at room temperature using on an AV500HD or DRX-600 spectrometer from *Bruker*. Calibration of the NMR spectra were done by referencing the chemical shift values to residual solvent signals DMSO-*d*₆: δ (¹H) = 2.50 ppm, δ (¹³C) = 39.52 ppm. Following abbreviations were used: s = singlet, d = doublet, dd = doublet of doublet, t = triplet, q = quartet, m = multiplet, b = broad.

Synthesis and analytics



Supplementary Figure 40: Synthesis route of compound 1.

4-Fluoro-2-((4-methoxyphenyl)amino)benzoic acid (**S3**)



The synthesis of **S3** was carried out according to a procedure from Haider *et al.*¹ for the synthesis of 2-(phenylamino)benzoic acid. Potassium carbonate (6.33 g, 45.8 mmol) was dried in oil-pump vacuum for 30 min, followed by the addition of **S1** (4.00 g, 22.9 mmol), **S2** (3.10 g, 25.2 mmol) and copper(II) oxid (0.73 g, 9.2 mmol). All solid starting materials were dried for additional 10 min in oil-pump vacuum and were suspended in dry. DMF (10 mL). The reaction mixture was heated to reflux for 8 h. After stirring at room temperature overnight the mixture was diluted with H₂O and filtered over celite. The filtrate was acidified with hydrochloric acid. The formed precipitate was collected by filtration and washed with H₂O. The solid was dissolved in EtOAc and the acidic filtrate was extracted three times with EtOAc. The organic layers were combined and dried over MgSO₄, filtered and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (silica gel, Cy:EtOAc = 1:3). The product was obtained as a brown solid (3.75 g, 62%).

$R_f = 0.85$ (silica gel, Cy:EtOAc = 1:3).

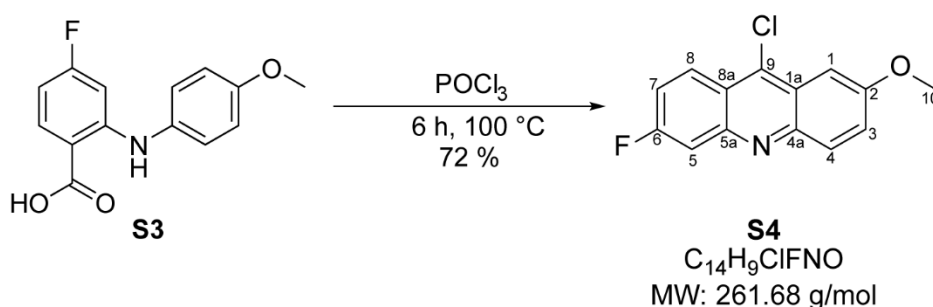
MS (ESI-): m/z calculated: [M] = 261.25; measured: [M-H]⁻ = 259.99.

¹H-NMR: (500.18 MHz, DMSO-*d*₆) δ [ppm] = 13.06 (bs, 1H, COOH); 9.64 (bs, 1H, NH); 7.93 (dd, 1 H, J = 8.8 Hz, J = 7.1 Hz, Pos. 3); 7.23 – 7.19 (m, 2H, Pos. 9 & 13); 7.00 – 6.97 (m, 2H, Pos. 10 & 12); 6.53 – 6.47 (m, 2H, Pos. 4 & 6); 3.77 (s, 3H, Pos. 14).

¹³C-NMR: (125.77 MHz, DMSO-*d*₆) δ [ppm] = 169.3 (Pos. 1); 165.9 (d, ¹J_{CF} = 248.8 Hz, Pos. 5); 156.6 (Pos. 11); 151.2 (d, ³J_{CF} = 12.2 Hz, Pos. 7); 134.8 (d, ³J_{CF} = 11.6 Hz, Pos. 3); 132.0 (Pos. 8); 125.7 (Pos. 9 & 13); 114.8 (Pos. 10 & 12); 108.0 (Pos. 2); 103.6 (²J_{CF} = 22.6 Hz, Pos. 4 or Pos. 6); 98.3 (²J_{CF} = 26.3 Hz, Pos. 4 or Pos. 6); 55.3 (Pos. 14).

¹⁹F-NMR: (470.64 MHz, DMSO-*d*₆) δ [ppm] = -104.25 – -104.31 (m, Pos. 9).

9-Chloro-6-fluoro-2-methoxyacridine (S4)



S4 was synthesized according a procedure of Mohammadi-Khanaposhtani *et al.*² for the synthesis of 6,9-dichloro-2-methoxyacridine. **S3** (3.75 g, 14.4 mmol) was suspended in freshly distilled POCl₃ (15.00 mL, 164.35 mmol) and heated to reflux for 6 h. The mixture was carefully poured under rigorous stirring on ice water. The pH was adjusted to 11 with aqueous NH₃ solution after the ice melted. The precipitate was removed by filtration and the filtrate was extracted five times with DCM. The combined organic layers were dried over MgSO₄, filtered and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (silica gel, Cy:EtOAc = 4:1). Desired product was obtained as a yellow solid (2.68 g, 72%).

R_f = 0.61 (silica gel, Cy:EtOAc = 1:3).

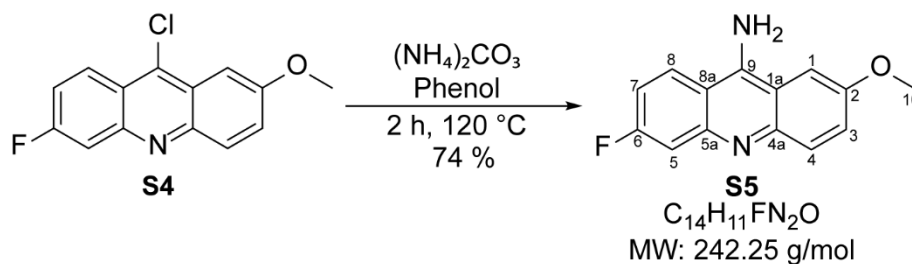
MS (ESI⁺): m/z calculated: [M] = 261.68; measured: [M-H]⁺ = 261.98.

¹H-NMR: (500.18 MHz, DMSO-*d*₆) δ [ppm] = 8.44 (dd, 1H, J = 9.5 Hz, J = 6.2 Hz, Pos. 8); 8.09 (d, 1H, ³J_{HH} = 9.4 Hz, Pos. 4); 7.90 (dd, 1H, ³J_{HF} = 10.3 Hz, ⁴J_{HH} = 2.5 Hz, Pos.5); 7.74 – 7.70 (m, 1H, Pos. 7); 7.62 (dd, ³J_{HH} = 9.5 Hz, ⁴J_{HH} = 2.8 Hz, Pos. 3); 7.51 (d, 1H, ⁴J_{HH} = 2.8 Hz, Pos. 1); 4.02 (s, 3H, Pos. 10).

¹³C-NMR: (125.77 MHz, DMSO-*d*₆) δ [ppm] = 162.3 (d, ¹J_{CF} = 250.1 Hz, Pos. 6); 158.1 (Pos. 2); 147.1 (d, ³J_{CF} = 13.4 Hz, Pos. 5a); 146.3 (Pos. 4a); 137.5 (Pos. 9); 131.1 (Pos. 4); 126.9 (d, ³J_{CF} = 10.7 Hz, Pos. 8); 126.7 (Pos. 3); 124.2 (Pos. 1a); 121.2 (Pos. 8a); 119.2 (²J_{CF} = 27.6 Hz, Pos. 7); 111.8 (²J_{CF} = 20.2 Hz, Pos. 5); 99.7 (Pos. 1); 55.8 (Pos.10).

¹⁹F-NMR: (470.64 MHz, DMSO-*d*₆) δ [ppm] = -108.93 – -109.98 (m, Pos. 6)

9-Amino-6-fluoro-2-methoxyacridine (S5)



The following synthesis was performed according to the synthesis von 9-amino-6-chlor-2-methoxyacridine by Bonse *et al.*³. **S4** (1.50 g, 5.7 mmol), phenol (6.00 g, 63.7 mmol) and ammonium carbonate (0.94 g, 9.7 mmol) were dried in oil-pump vacuum and then heated to 120 °C for 2 h. The mixture was cooled to room temperature and 5 M aqueous NaOH solution was added until pH 12. After separating the precipitate by filtration, it was washed with 1 M aqueous NaOH and H₂O and then dissolved in EtOAc. The aqueous layer was extracted with EtOAc and the combined organic layers were dried over MgSO₄ and filtered. After removing the solvent the crude product was purified by column chromatography (silica gel, Cy:EtOAc = 1:1 to DCM:MeOH = 4:1). The product was obtained as a yellow solid (1.02 g, 72%).

R_f = 0.74 (silica gel, DCM:MeOH = 4:1).

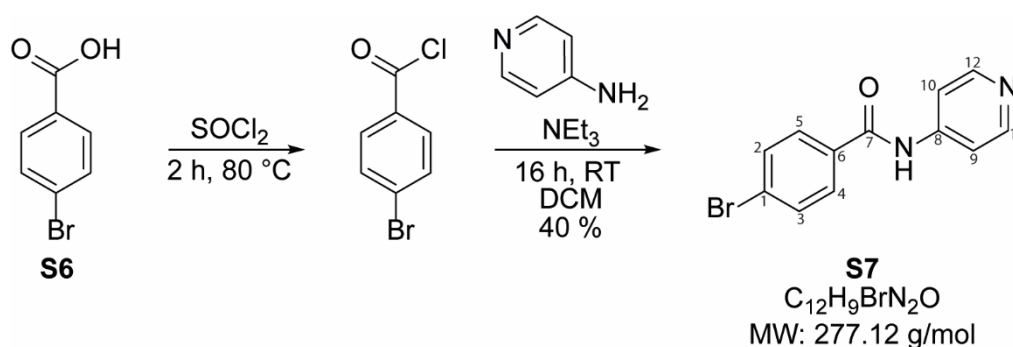
MS (ESI-): m/z calculated: [M] = 242.25; measured: [M-H]⁺ = 243.04.

¹H-NMR: (500.18 MHz, DMSO-*d*₆) δ [ppm] = 8.44 (dd, 1H, J = 9.4 Hz, J = 6.5 Hz, Pos. 8); 7.74 (d, 1H, ³J_{HH} = 9.0 Hz, Pos. 4); 7.67 – 7.66 (m, 3H, Pos. 1 & NH₂); 7.44 (dd, ³J_{HF} = 11.6 Hz, ⁴J_{HH} = 2.4 Hz, Pos. 5); 7.36 (dd, ³J_{HH} = 9.3 Hz, ⁴J_{HH} = 2.3 Hz, Pos. 3); 7.24 – 7.21 (m, 1H, Pos. 7); 3.91 (s, 3H, Pos. 10).

¹³C-NMR: (125.77 MHz, DMSO-*d*₆) δ [ppm] = 162.4 (d, ¹J_{CF} = 246.6 Hz, Pos. 6); 154.3 (Pos. 2); 148.8 (Pos. 9); 148.4 (d, ³J_{CF} = 13.1 Hz Pos. 5a); 146.0 (Pos. 4a); 130.2 (Pos. 4); 126.1 (d, ³J_{CF} = 11.3 Hz, Pos. 8); 124.0 (Pos. 3); 112.9 (Pos. 1a); 112.4 (d, ²J_{CF} = 25.8 Hz, Pos. 7); 110.5 (d, ²J_{CF} = 19.3 Hz, Pos. 5); 110.2 (Pos. 8a); 100.4 (Pos. 1) 55.7 (Pos. 10).

¹⁹F-NMR: (470.64 MHz, DMSO-*d*₆) δ [ppm] = -111.68 – -111.62 (m, Pos. 6)

4-Bromo-*N*-4-pyridinylbenzamide (**S7**)



S6 (0.50 g, 2.3 mmol) was dried in oil-pump vacuum for 15 min, suspended in fresh distilled SOCl₂ (3.1 mL, 42 mmol) and heated to reflux until the suspension turned to a clear solution. After cooling to room temperature excessive SOCl₂ was removed under reduced pressure and the resulting solid was dried for 1 h in oil-pump vacuum. The solid was dissolved in dry DCM (2 mL), following by the addition of NEt₃ (0.65 mL, 4.6 mmol) under cooling. A solution of 4-aminopyridine (0.22 g, 2.28 mmol) in a mixture of dry DCM (15.00 mL) and dry DMF (1.50 mL) was added dropwise under cooling. The reaction mixture was stirred at room temperature overnight. The reaction was stopped by addition of H₂O and subsequently the aqueous layer was extracted with EtOAc. The combined organic layers were dried over MgSO₄, filtered and the solvent removed under reduced pressure. The crude product was purified by column chromatography (silica gel, Cy:EtOAc = 1:1). Since residues of **S6** were observed after column chromatography, the obtained solid was dissolved in 1 M NaOH and the aqueous layer extracted with EtOAc. After removing the solvent under reduced pressure, the product was obtained as a white solid (251 mg, 40%).

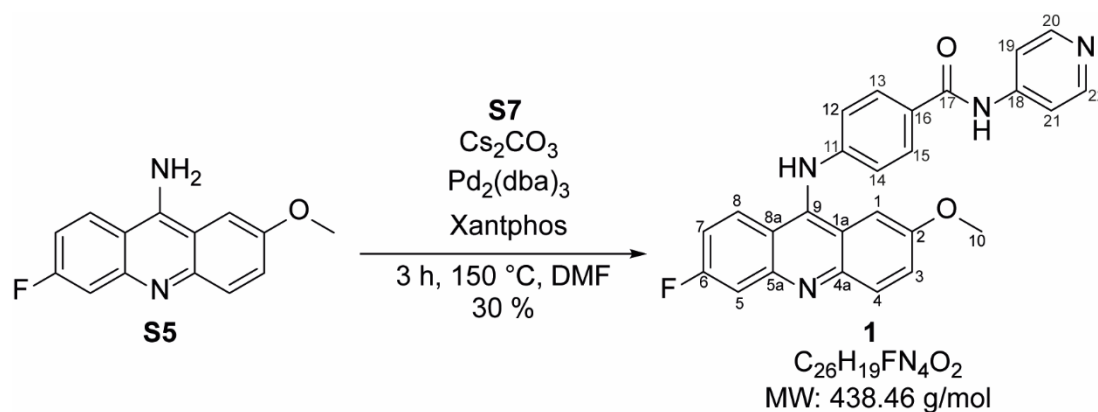
R_f = 0.61 (silica gel, Cy:EtOAc = 1:3).

MS (ESI-): m/z calculated: [M] = 277.12; measured: [M-H]⁻ = 276.93.

¹H-NMR: (500.18 MHz, DMSO-*d*₆) δ [ppm] = 10.64 (bs, 1H, NH); 8.49 – 8.47(m, 2H, Pos. 11 & 12); 7.93 - 7.90 (m, 2H, Pos. 4 & 5), 7.79 – 7.76 (m, 4H, Pos. 2, 3, 9 & 10).

¹³C-NMR: (125.77 MHz, DMSO-*d*₆) δ [ppm] = 165.5 (Pos. 7); 150.3 (Pos. 11 & 12); 145.7 (Pos. 8); 133.3 (Pos. 4); 132.5 (Pos. 2 & 3); 129.9 (Pos. 4 & 5); 126.0 (Pos. 1); 114.0 (Pos. 9 & 10).

4-[6-Fluoro-2-methoxy-9-acridinyl)amino]-N-4-pyridinylbenzamide (1)



The synthesis of **1** was performed according to a procedure of Gellerman *et al.*⁴ for the synthesis of functionalized aminoacridine derivatives. **S5** (110.0 mg, 454.1 μ mol), **S7** (125.8 mg, 454.1 μ mol) and Cs₂CO₃ (74 mg, 227 μ mol) were dried in oil-pump vacuum. The solids were then suspended in dry DMF (4 mL) and heated to 90 °C for 7 h, following by heating for 4 h to reflux. However, no product formation could be observed by TLC. Therefore, Cs₂CO₃ (518 mg, 1.59 mmol), Pd₂(dba)₃ (37.1 mg, 40.9 μ mol) and Xantphos (39.3 mg, 68.1 μ mol) were added to the reaction mixture. The mixture was heated to reflux for 3 h. The solvent was removed by warming to 40 °C under oil-pump vacuum. The crude product was suspended in MeOH and filtered over celite. After the removal of MeOH, the crude product was purified by column chromatography (silica gel, DCM:MeOH = 9:1). The desired product was obtained as an orange solid (58 mg, 30%).

For analytics via NMR spectroscopy 2 vol.-% of either TFA or HCl was added to the NMR sample, since analyzing the normal sample was not possible due to an aminoacridine-acridanimine tautomerism as described in literature⁵.

The isolated product contained two impurities after applying oil-pump vacuum. One of those is DCM and the other could not be identified. ¹H and ¹³C spectra show the presence of these impurities.

R_f = 0.57 (silica gel, DCM/MeOH 9:1).

MS (ESI-): m/z calculated for [M] = 438.46; measured: [M-H]⁻ = 438.08.

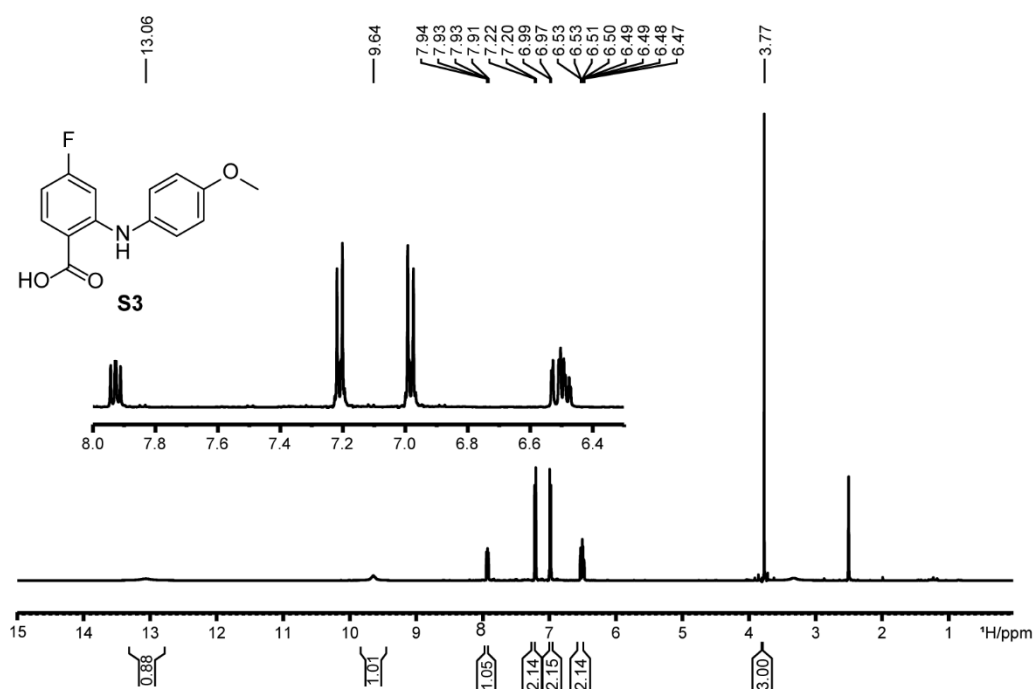
¹H-NMR: (500.18 MHz, DMSO-*d*₆) δ [ppm] = 11.94 (s, 1H, Amid-NH); 8,76 (d, 2H, ³J_{HH} = 7.3 Hz, Pos. 20 & 22); 8.49 (d, 2H, ³J_{HH} = 7.4 Hz, Pos. 19 & 21); 8.27 – 8,21 (m, 4H, Pos. 4, 8, 13 & 15); 8.00 (dd, 1H, ³J_{HF} = 9.6 Hz, ⁴J_{HH} = 2.5 Hz, Pos. 5); 7.92

(d, 1H, $^4J_{\text{HH}} = 2.7$ Hz, Pos. 1); 7.77 (dd, 1H, $^3J_{\text{HH}} = 9.5$ Hz, $^4J_{\text{HH}} = 2.6$ Hz, Pos. 3); 7.55 (d, 2H, $^3J_{\text{HH}} = 8.7$ Hz, Pos. 12 & 14); 7.44 – 7.39 (m, 1H, Pos. 7); 3.79 (s, 3H, Pos. 10);

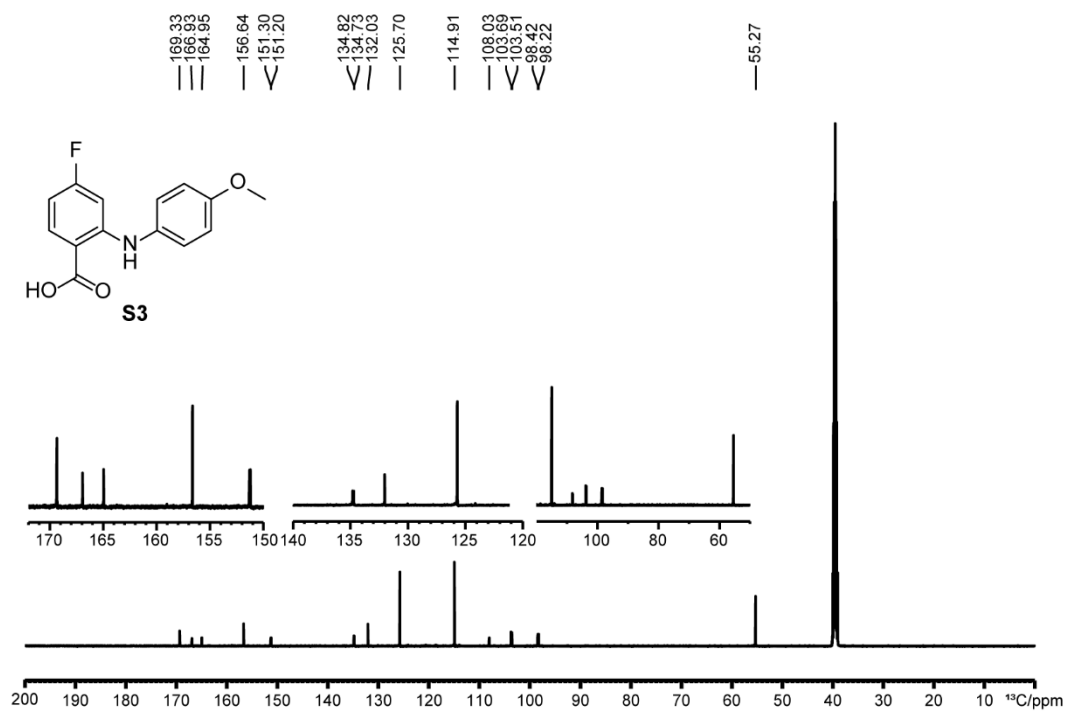
$^{13}\text{C-NMR}$: (125.77 MHz, $\text{DMSO-}d_6$) δ [ppm] = 166.4 (Pos. 17); 164.7 (d, $^1J_{\text{CF}} = 256.3$ Hz, Pos. 6); 156.2 (Pos. 2); 154.8 (Pos. 18); 152.8 (Pos. 9); 146.3 (Pos. 11); 141.9 (Pos. 20 & 22); 140.9 ($^3J_{\text{CF}} = 13.8$ Hz, Pos. 5a); 136.6 (Pos. 4a); 130.3 (Pos. 4); 129.8 ($^3J_{\text{CF}} = 11.5$ Hz, Pos. 8); 129.2 (Pos. 3 & Pos. 16); 122.4 (Pos. 12 & 14); 121.1 (Pos. 13 & 15); 117.1 (Pos. 1a); 115.4 (Pos. 19 & 21); 115.1 ($^2J_{\text{CF}} = 25.9$ Hz, Pos. 7); 112.6 (Pos. 8a); 103.8 (Pos. 1); 103.6 ($^2J_{\text{CF}} = 25.1$ Hz, Pos. 5); 56.2 (Pos. 10).

$^{19}\text{F-NMR}$: (470.64 MHz, $\text{DMSO-}d_6$) δ [ppm] = -100.79 – -100.83 (m, Pos. 6)

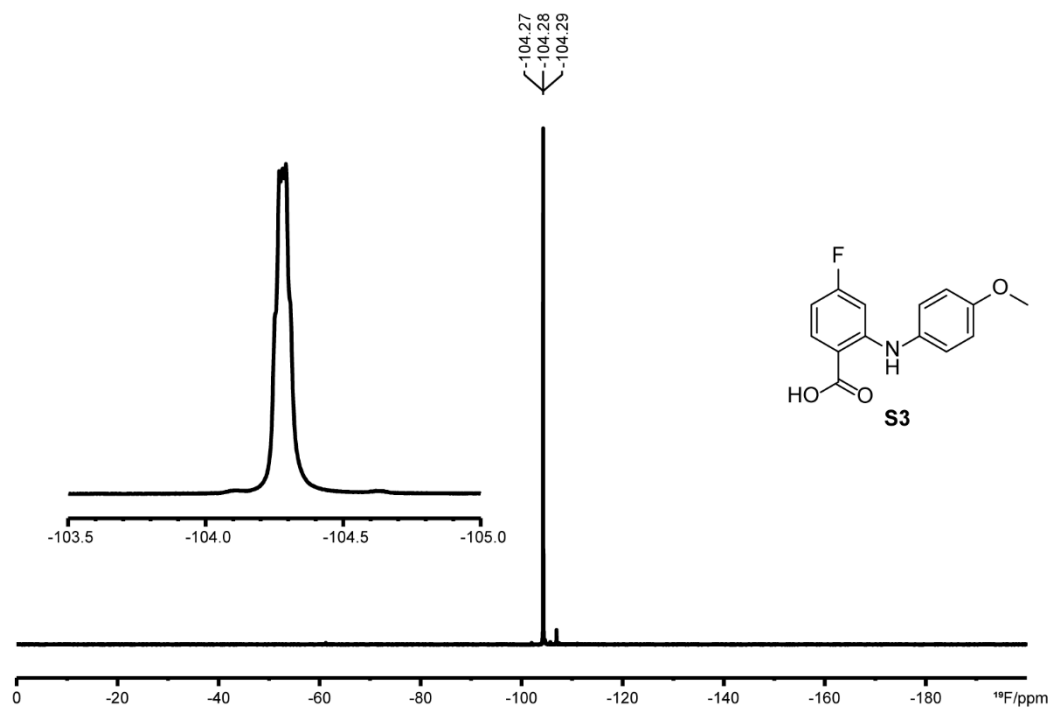
NMR-spectra



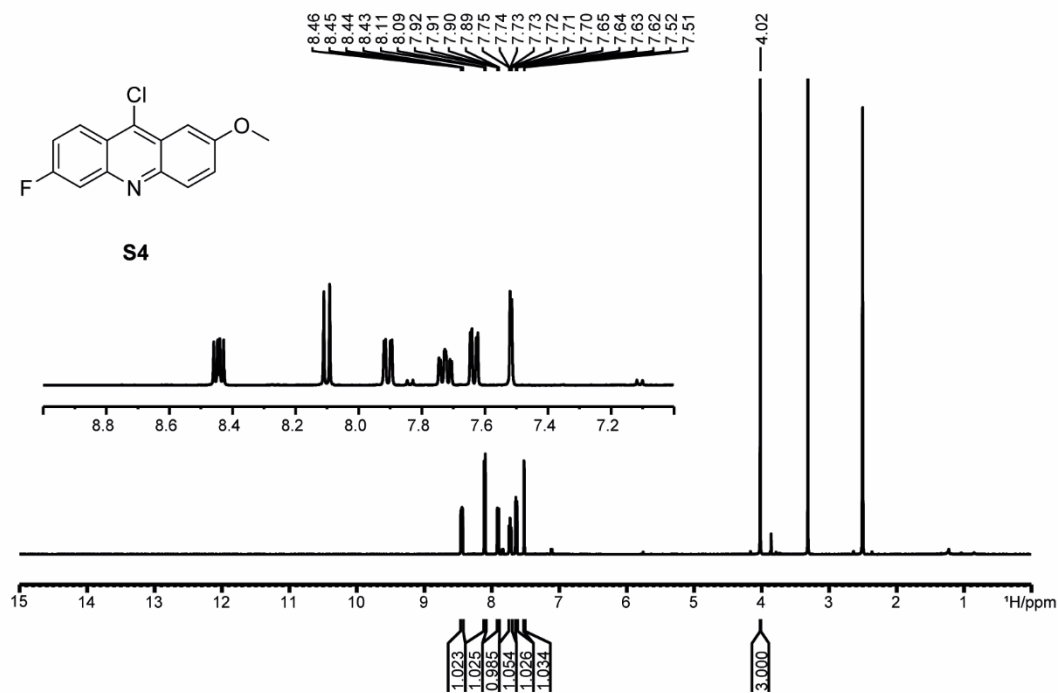
Supplementary Figure 41: $^1\text{H-NMR}$ spectrum of **S3** ($\text{DMSO-}d_6$, 298 K, 500.18 MHz).



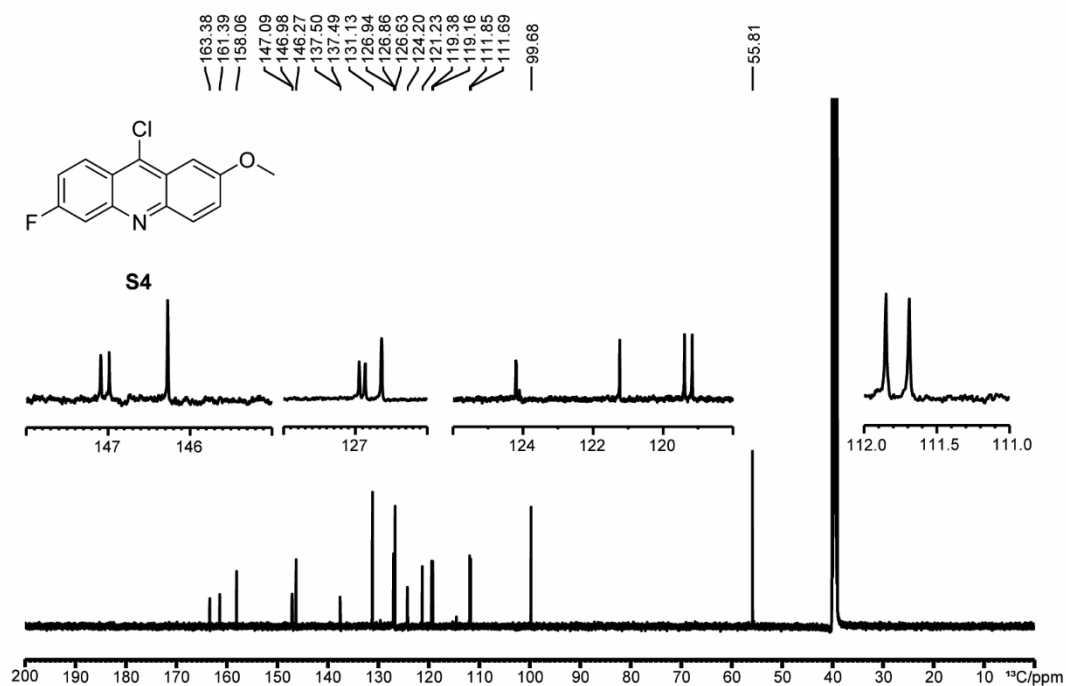
Supplementary Figure 42: ¹³C-NMR spectrum of S3 (DMSO-d₆, 298 K, 157.77 MHz).



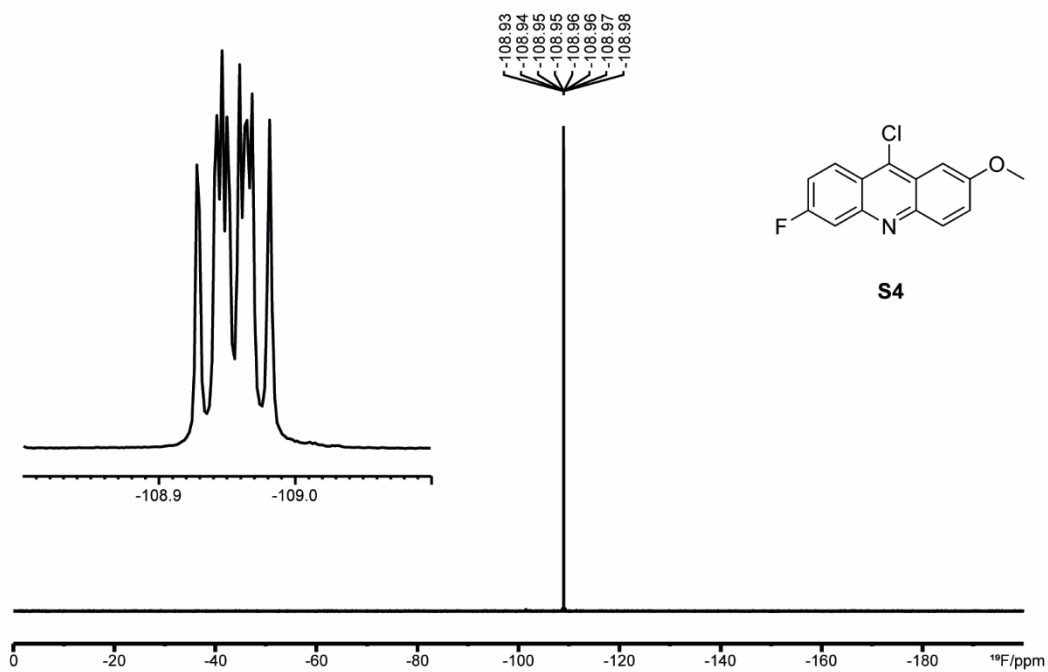
Supplementary Figure 43: ¹⁹F-NMR spectrum of S3 (DMSO-d₆, 298 K, 470.64 MHz).



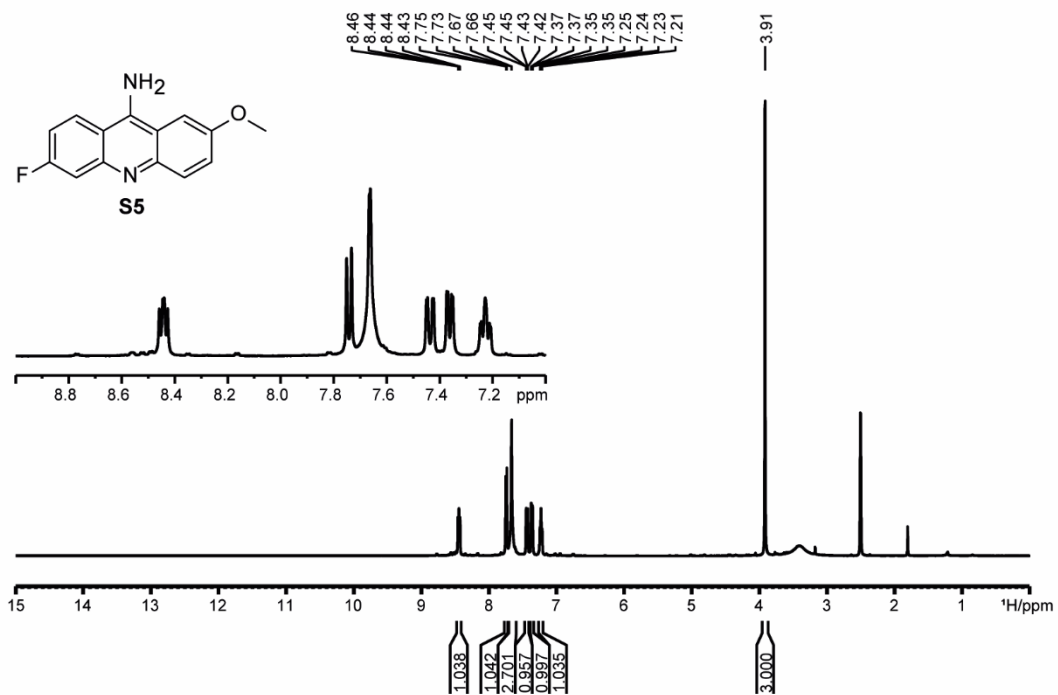
Supplementary Figure 44: ¹H-NMR spectrum of S4 (DMSO-d₆, 298 K, 500.18 MHz).



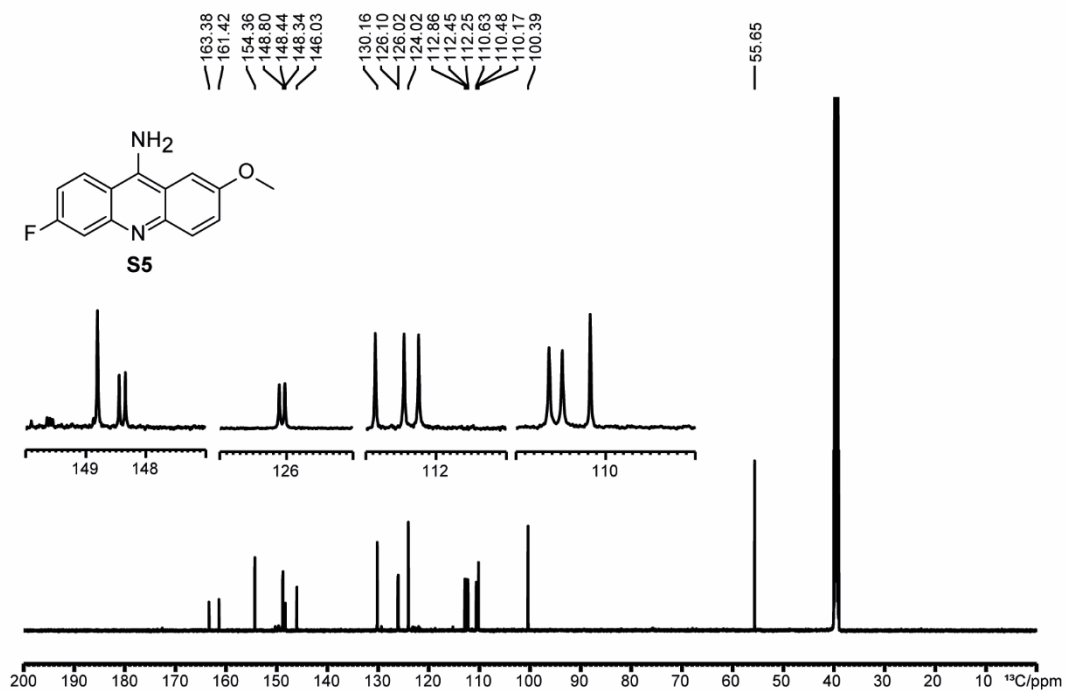
Supplementary Figure 45: ¹³C-NMR spectrum of S3 (DMSO-d₆, 298 K, 157.77 MHz).



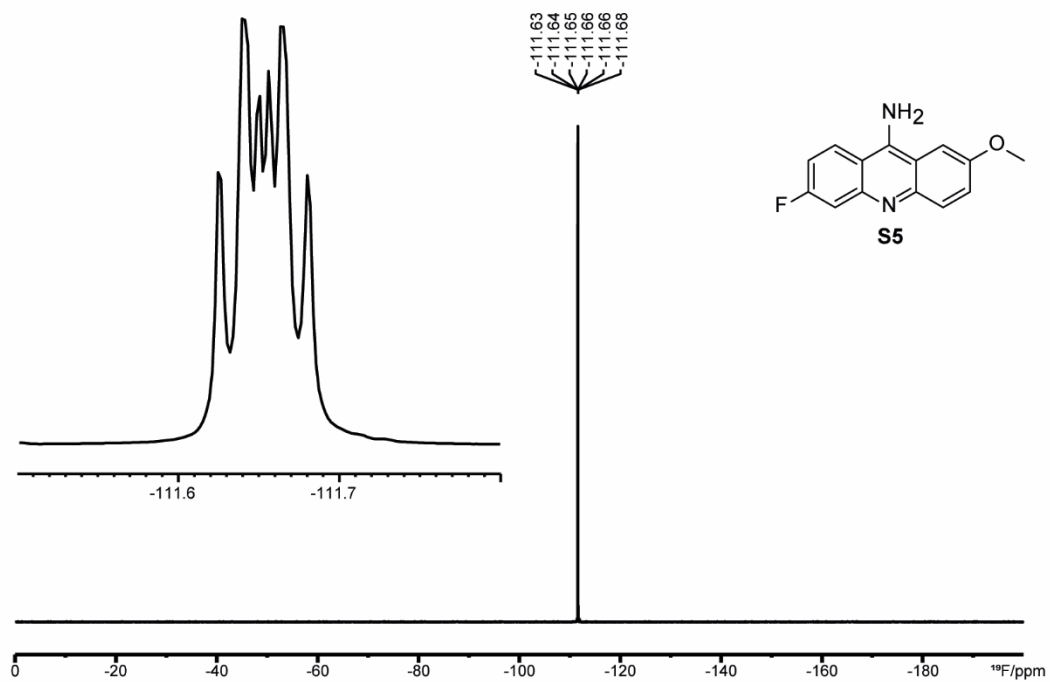
Supplementary Figure 46: ¹⁹F-NMR spectrum of *S4* (DMSO-*d*₆, 298 K, 470.64 MHz).



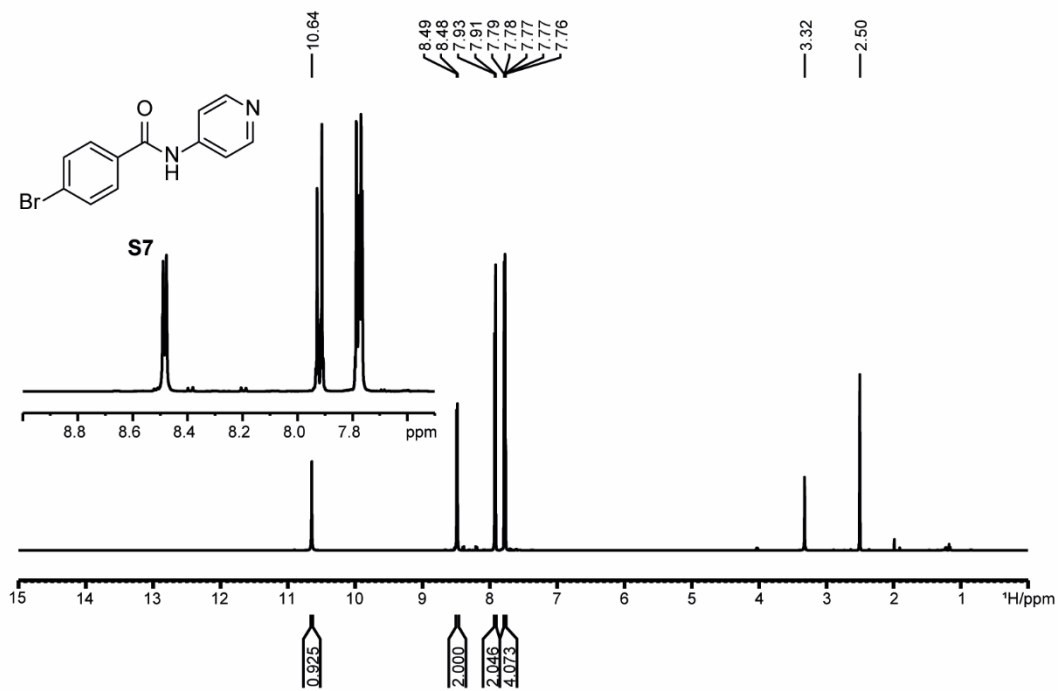
Supplementary Figure 47: ¹H-NMR spectrum of *S5* (DMSO-*d*₆, 298 K, 500.18 MHz).



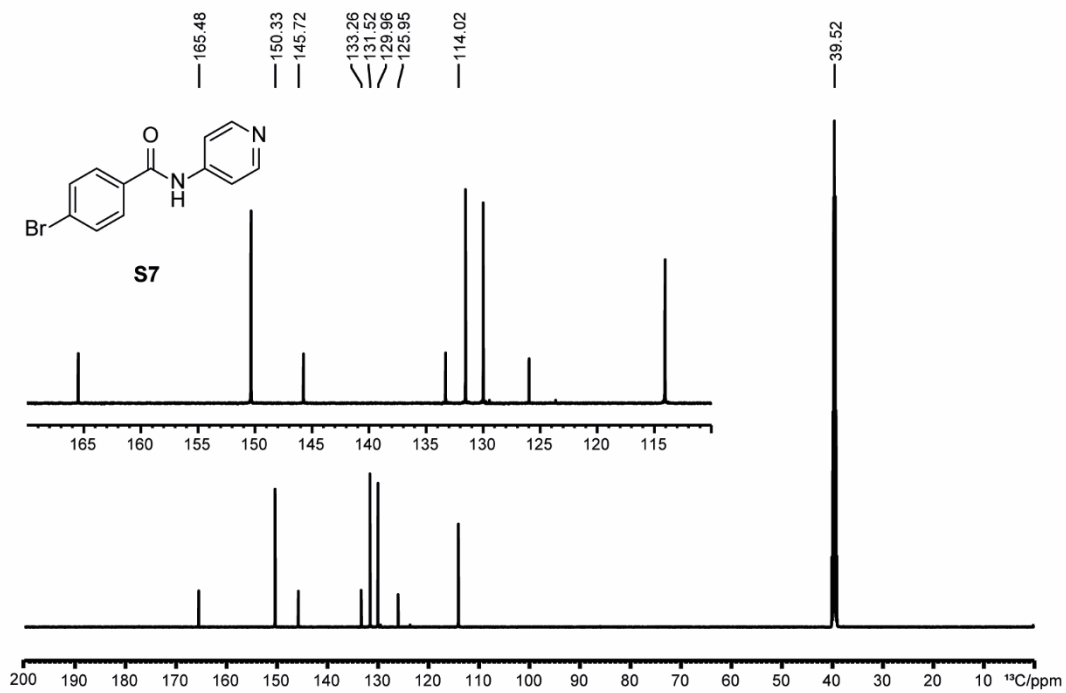
Supplementary Figure 48: ^{13}C -NMR spectrum of *S5* (DMSO- d_6 , 298 K, 157.77 MHz).



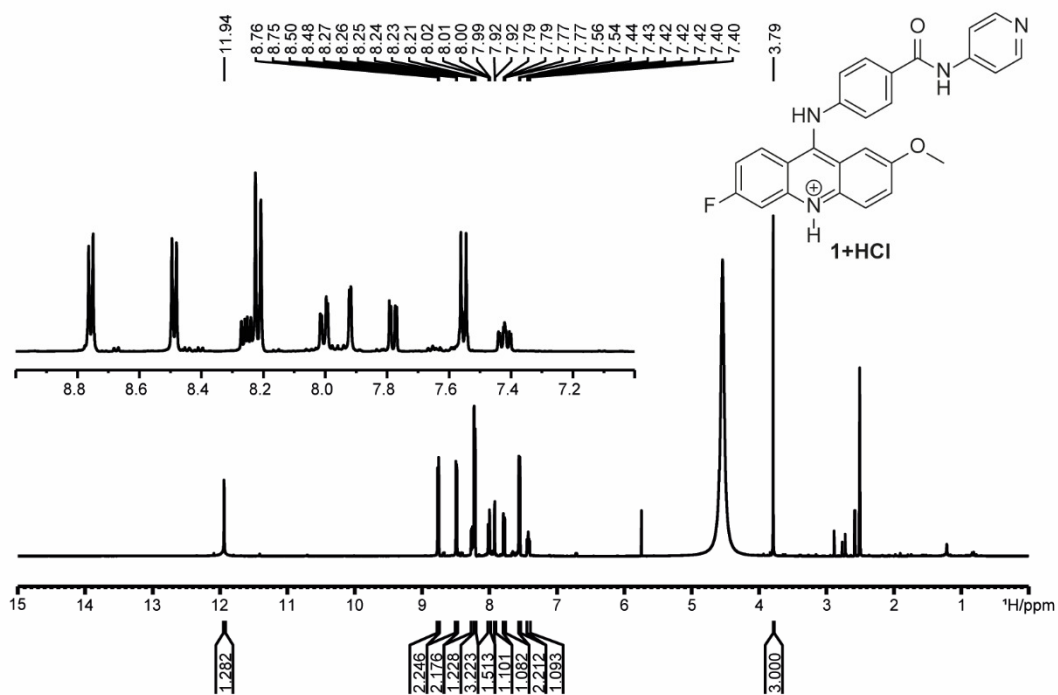
Supplementary Figure 49: ^{19}F -NMR spectrum of *S5* (DMSO- d_6 , 298 K, 470.64 MHz).



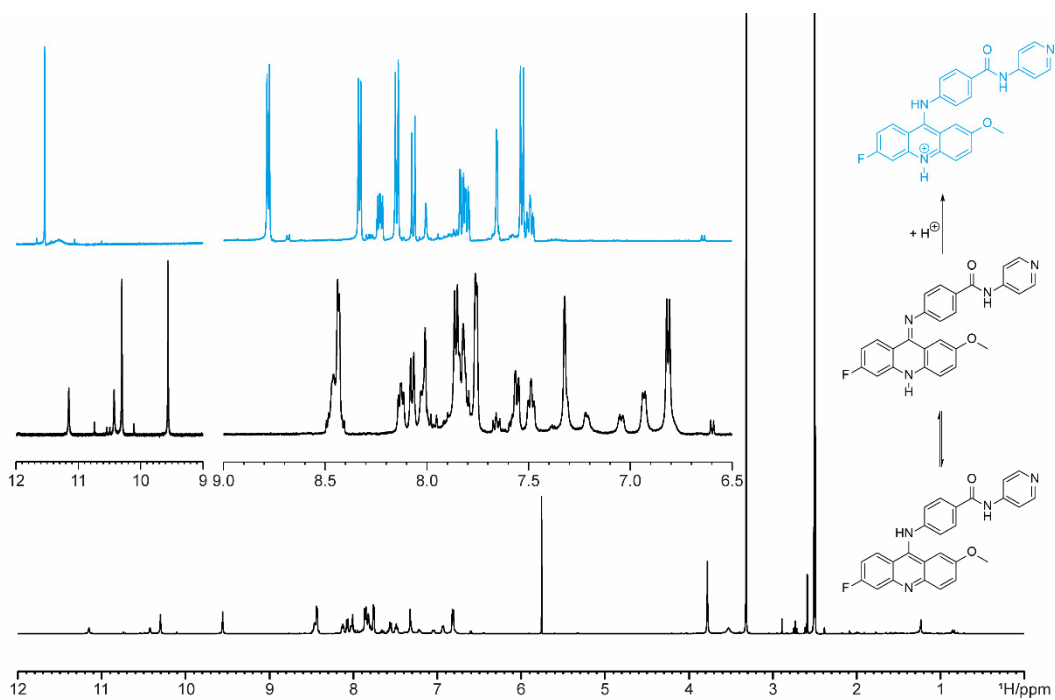
Supplementary Figure 50: ¹H-NMR spectrum of S7 (DMSO-d₆, 298 K, 500.18 MHz).



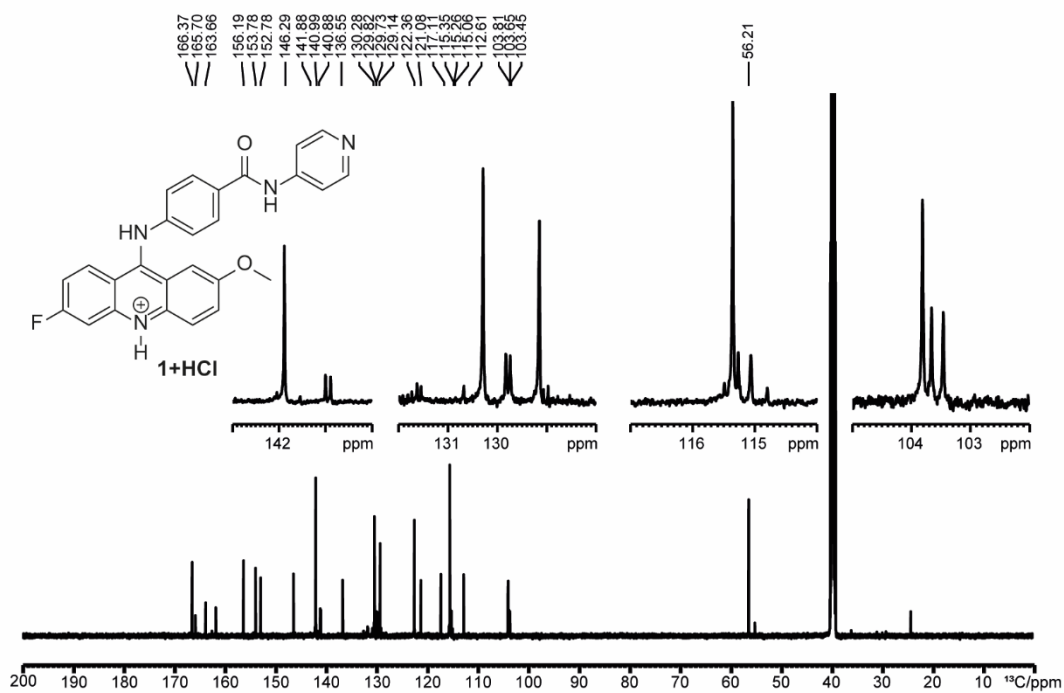
Supplementary Figure 51: ¹³C-NMR spectrum of S7 (DMSO-d₆, 298 K, 157.77 MHz).



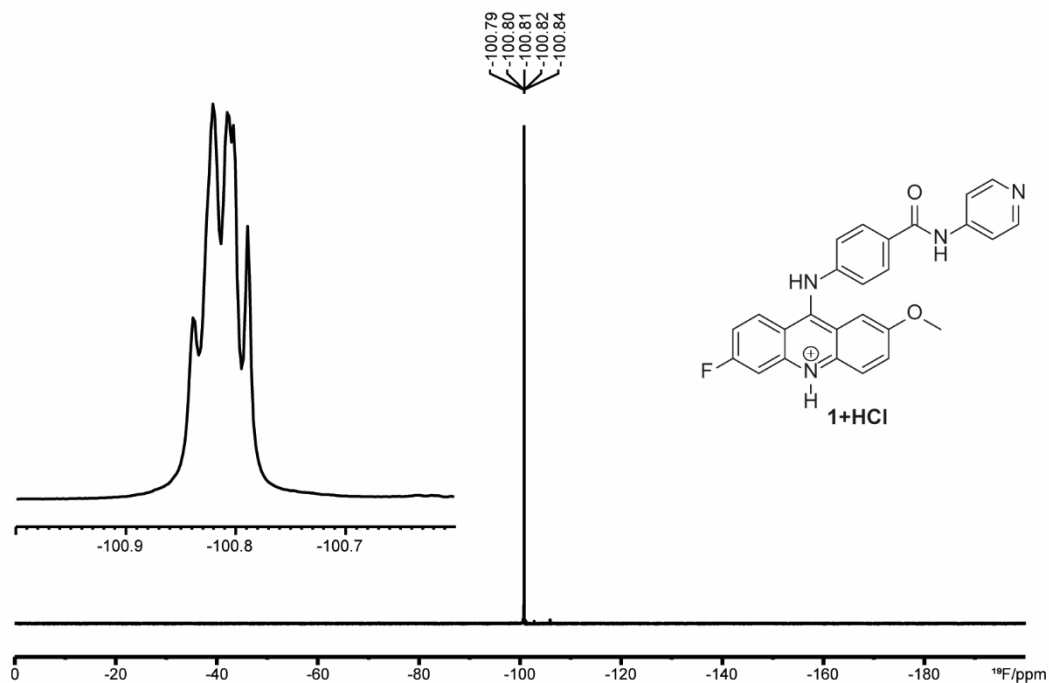
Supplementary Figure 52: $^1\text{H-NMR}$ spectrum of 1 with 2 vol.-% HCl added (DMSO- d_6 , 298 K, 500.18 MHz).



Supplementary Figure 53: NMR spectra of compound 1 in addition (blue) and without acid (black). The spectra were recorded in DMSO- d_6 at 298 K (600.31 MHz). The isolated product contained two impurities after drying in oil-pump vacuum. One of those is DCM and the other could not be identified. ^1H and ^{13}C spectra show the presence of these impurities.

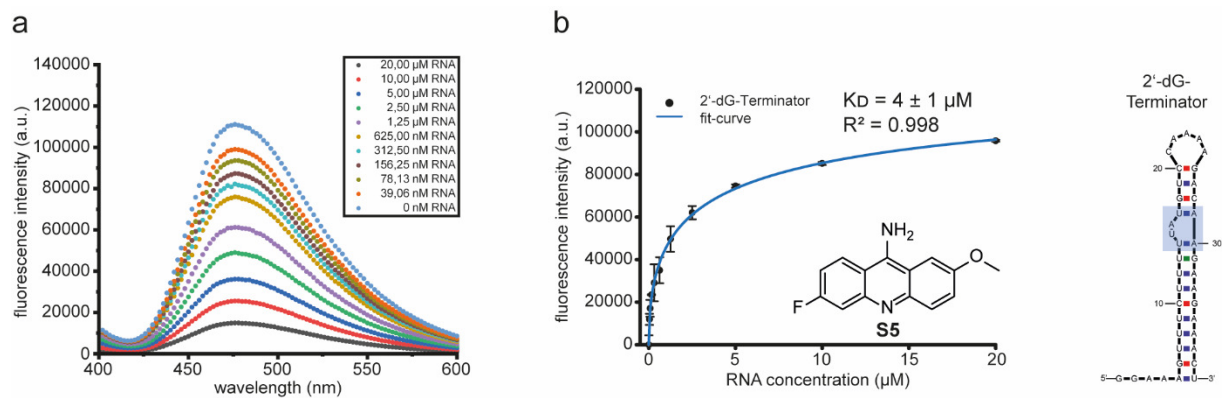


Supplementary Figure 54: ^{13}C -NMR spectrum of **1** with 2 vol.-% HCl added (DMSO- d_6 , 298 K, 157.77 MHz).

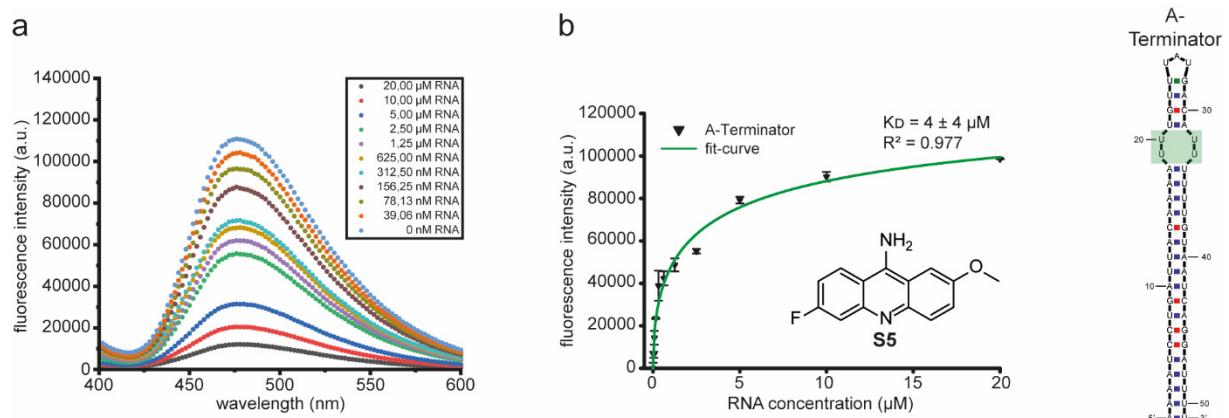


Supplementary Figure 55: ^{19}F -NMR spectrum of **1** with 2 vol.-% HCl added (DMSO- d_6 , 298 K, 470.64 MHz).

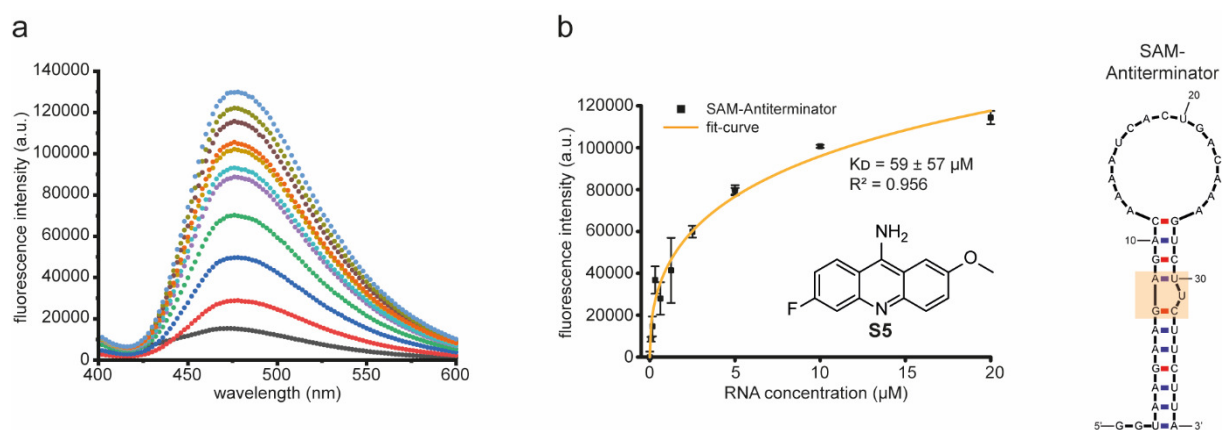
Fluorescence binding assay



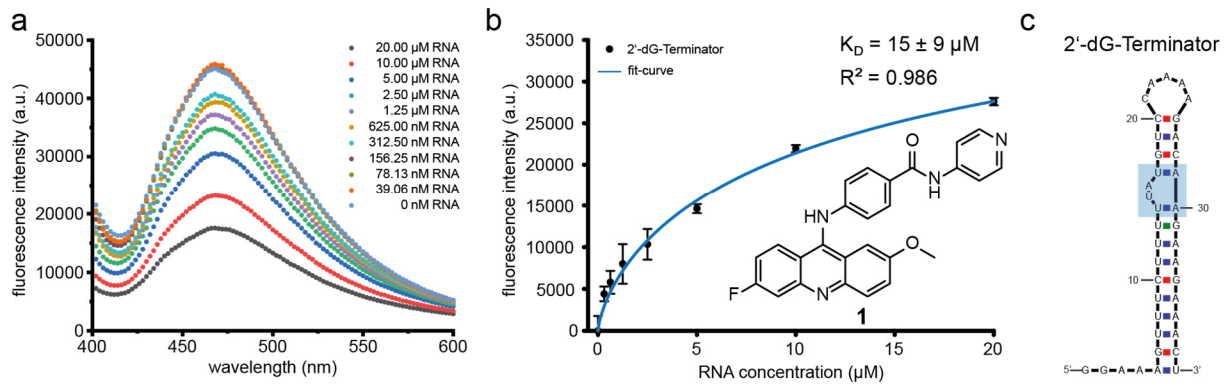
Supplementary Figure 56: Determination of the dissociation constant. (a) Fluorescence titration and (b) binding curve of compound S5 to the 2'-dG-Terminator.



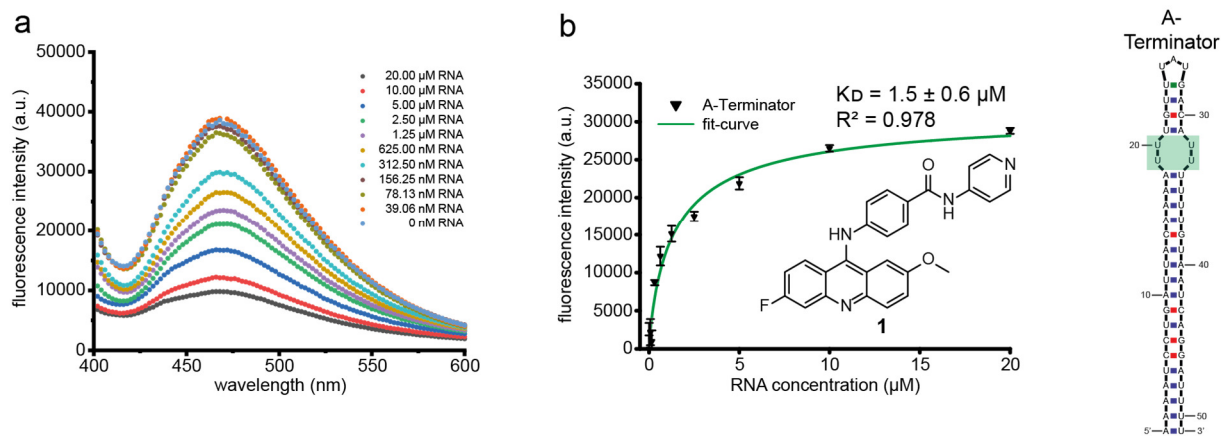
Supplementary Figure 57: Determination of the dissociation constant. (a) Fluorescence titration and (b) binding curve of compound S5 to the A-Terminator.



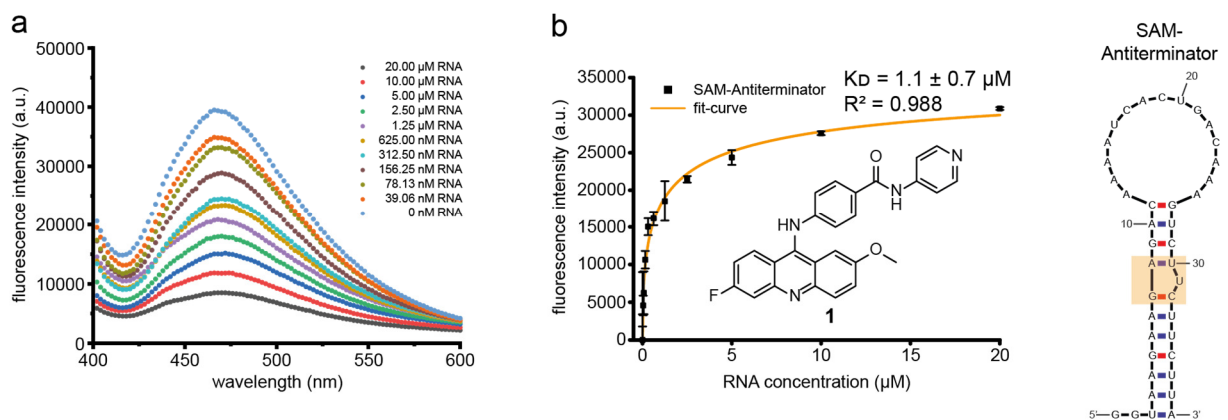
Supplementary Figure 58: Determination of the dissociation constant. (a) Fluorescence titration and (b) binding curve of compound S5 to the SAM-Terminator.



Supplementary Figure 59: Fluorescence-based determination of affinity to the 2'-dG-Terminator stem of a fragment grown with an acridine moiety. (a) Individual UV-VIS spectra, (b) Determination of affinity of compound 1 to the 2'-dG-Terminator stem (c).



Supplementary Figure 60: Determination of the dissociation constant. (a) Fluorescence titration and (b) binding curve of compound 1 to the A-terminator.



Supplementary Figure 61: Determination of the dissociation constant. (a) Fluorescence titration and (b) binding curve of compound 1 to the SAM-anti-terminator stem.

References

- (1) Haider, M. R.; Ahmad, K.; Siddiqui, N.; Ali, Z.; Akhtar, M. J.; Fuloria, N.; Fuloria, S.; Ravichandran, M.; Yar, M. S. Novel 9-(2-(1-Arylethylidene)Hydrazinyl)Acridine Derivatives: Target Topoisomerase 1 and Growth Inhibition of HeLa Cancer Cells. *Bioorg. Chem.* **2019**, *88*, 102962. <https://doi.org/10.1016/j.bioorg.2019.102962>.
- (2) Mohammadi-Khanaposhtani, M.; Rezaei, S.; Khalifeh, R.; Imanparast, S.; Faramarzi, M. A.; Bahadorikhalili, S.; Safavi, M.; Bandarian, F.; Nasli Esfahani, E.; Mahdavi, M.; Larijani, B. Design, Synthesis, Docking Study, α -Glucosidase Inhibition, and Cytotoxic Activities of Acridine Linked to Thioacetamides as Novel Agents in Treatment of Type 2 Diabetes. *Bioorg. Chem.* **2018**, *80* (March), 288–295. <https://doi.org/10.1016/j.bioorg.2018.06.035>.
- (3) Bonse, S.; Santelli-Rouvier, C.; Barbe, J.; Krauth-Siegel, R. L. Inhibition of Trypanosoma Cruzi Trypanothione Reductase by Acridines: Kinetic Studies and Structure–Activity Relationships. *J. Med. Chem.* **1999**, *42* (26), 5448–5454. <https://doi.org/10.1021/jm990386s>.
- (4) Gellerman, G.; Gaisin, V.; Brider, T. One-Pot Derivatization of Medicinally Important 9-Aminoacridines by Reductive Amination and SNAr Reaction. *Tetrahedron Lett.* **2010**, *51* (5), 836–839. <https://doi.org/10.1016/j.tetlet.2009.12.020>.
- (5) Demidov, O. P.; Borovlev, I. V.; Amangasieva, G. A.; Avakyan, E. K. Oxidative SNH Amidation of Acridine and Tautomerism of N-(Acridin-9-Yl)Benzamides. *Chem. Heterocycl. Compd.* **2016**, *52* (2), 104–109. <https://doi.org/10.1007/s10593-016-1841-7>.

14. Research article VI: Book chapter: NMR on Ribozymes in “Ribozymes” (Wiley-VCH)

Bozana Knezic*, Oliver Binas*, Albrecht Völklein and Harald Schwalbe

*authors contributed equally to the work

“Ribozymes”, 2020, *Wiley-VCH*, *accepted*

Ribozymes are a class of functional RNA, particularly prevalent in bacteria, but also existing in lower eukaryotes.^[426] They are characterized by their ability to self-cleave or self-splice, often but not always depending on metal ion catalysis, by providing a structural arrangement which favors in-line RNA cleavage reactions.^[15] The chapter introduces NMR-methods suitable for the investigation of ribozymes ranging from 1D investigation to sophisticated 3D experiments as well as methods for NMR-sample generation covering various RNA preparation methods. NMR-studies on ribozymes are then discussed establishing a concise state-of-the-art in NMR research of the various ribozymes investigated in the literature. The well characterized hepatitis delta virus (HDV) ribozyme is particularly focused, introducing recent and novel data to the scientific community.

The author of the thesis contributed to literature screening and manuscript preparation for the paragraphs introducing NMR-experiments (page 4 to page 8) as well as four of the seven subchapters describing NMR studies on particular ribozymes. Chapters comprising preparatory techniques and HDV as well as Group I intron chapters were prepared by B. Knezic and “Leadzyme” was contributed by A. Völklein.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a highly versatile physical method to study structure and function of proteins, oligonucleotides and their complexes at atomic resolution. With the advent of labeling RNA with NMR-active stable isotopes ¹³C and ¹⁵N, NMR is able to characterize biologically relevant RNAs, RNA-ligand- and RNA-protein-complexes and determine key aspects to further our understanding not only in structural terms, but also to understand enzyme reactivity.

We outline the power of NMR spectroscopy to investigate structure, dynamics and function of catalytically active RNA ribozymes. We will introduce NMR-based methods and also cover practical aspects of NMR in the first part of the chapter. In the second part of the chapter, we will provide a summary of NMR studies of the ribozyme classes shown in Figure 1.

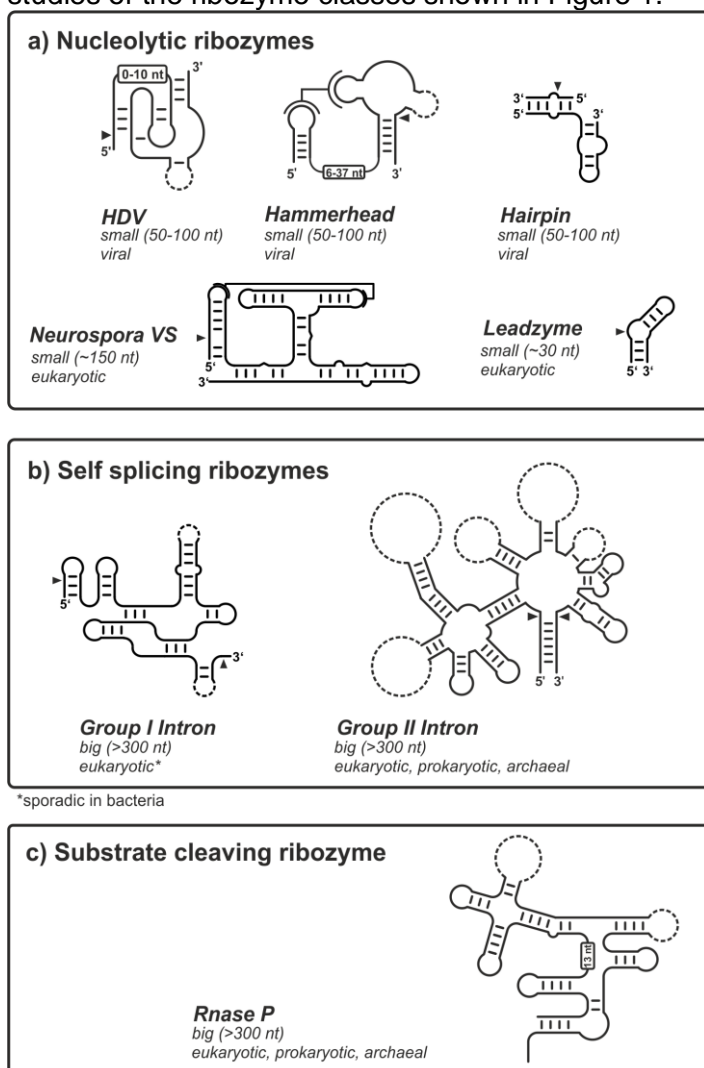


Figure 1 Different classes of ribozymes are shown. Ribozymes can be classified into three different types, namely ribozymes with nucleolytic activity (a), with self-splicing activity (b) and substrate-cleaving activity (c). Occurrence in different kingdoms of life (prokaryotes, eukaryotes and archaea) is pointed out below each Ribozyme. Nucleotides are abbreviated as nt. Base-paired regions are depicted

schematically. Small loops are represented by non-dashed lines. Loops of variable size are represented by dashed lines.

Methods & Preparation

NMR experiments of RNAs^[1] can be conducted in solution, but recently also in the solid state. Solid state NMR (ssNMR) emerged as a new technique for the structural characterization of RNA just a few years ago introducing several new aspects to NMR of RNA. Liquid state NMR is limited in molecular weight of the system under study due to unbeneficial relaxation properties of high molecular weight systems. In ssNMR though, resolution is independent of molecular weight which only affects signal intensity. Resolution in ssNMR is instead dependent on sample preparation. Typical freeze drying methods like lyophilization yield samples, featuring insufficient resolution for structure determination due to structural heterogeneity.^[2] With new advances like micro-crystallization^[3] it is possible to reach line widths of 0.3 ppm, with solution NMR still being far superior. An advantage investigating RNA with ssNMR is the correlation of heteronuclei with dipolar interactions which is not possible in solution NMR due to molecular tumbling.^[4] As RNA is sparse of protons this provides an interesting benefit over solution NMR. In 2015 Carlomagno et al. showed that it is possible to determine a structure of a 26 nt RNA with ssNMR.^[5] Since ssNMR on RNA is an emerging technique, not many studies exist featuring RNA or ribozymes. Accordingly we will focus on NMR studies on ribozymes in solution in this review. Depending on the kind of information, solution NMR experiments of ribozyme require between 50 μ M - 0.5 mM sample in a volume of 0.3 mL. For an initial screening with simple experiments, quantities of 2-5 nmoles can be sufficient while more sophisticated experiments require over 150 nmoles of substance. Experiments can be conducted on unlabeled samples, but labeling of RNAs with ^{13}C and ^{15}N is typically performed.

An NMR setup for studies on biomacromolecules consists of an NMR spectrometer with a helium-cooled superconducting magnet with field strength between 14-21 Tesla, resulting in a ^1H frequency in the range of 400 to 950 MHz. A console with the required radiofrequency (rf) electronics including frequency generator and power amplifier for 4-6 channels allowing to apply radiofrequency pulse on the NMR active nuclei ^1H , ^2H , ^{13}C , ^{15}N , ^{19}F , ^{31}P is required. A multi-channel probe, containing the RNA sample in an NMR tube of 0.3-0.5 ml volume, and the circuit for rf-application at any desired temperature are required as well. In order to increase signal-to-noise for different NMR experiments, the receiving coil can be cooled and different setups for detected nuclei have been implemented.

There are different ways to prepare isotope-labeled RNAs. Enzymatic in vitro transcription of RNA (Scheme 1a) is probably the most straightforward and cost efficient way to generate RNA samples and requires - besides the DNA template and common laboratory chemicals - only RNA polymerase and the respective ribonucleoside-triphosphates (rNTPs). The DNA template contains the necessary promoter sequence combined with the target DNA-sequence. In common practice, DNA oligonucleotides are produced via PCR or linearized DNA plasmids. Isotopic labeled rNTPs, which serve as raw material, are commercially available at reasonable prices.^[6,7] RNA polymerases that provide the required throughput are as well commercially available but we typically prepare it in house.^[8-11] Purification of the transcribed target RNA involves desalting, followed by HPLC- or preparative PAGE purification. Advantages of the method include the high yields of RNA at reasonable cost and the acceptable preparative effort.

Helmling *et al.* established a fast and high throughput compatible native method.^[12] This method involves centrifugal devices, to remove NTPs and salts right after the in-vitro transcription or ligation process. Furthermore, in case of using NMR buffer as washing material, the purified sample may already be concentrated and prepared for the transfer to an NMR tube. Residual enzymes and DNA cannot be removed by this technique. On the other hand, these residues are generally present in such low concentrations or are non-labeled with NMR-active nuclei ^{13}C or ^{15}N , they do not disturb the structure and function of the target ribozyme nor the relevant NMR signals. Unfortunately all these methods do not allow the incorporation of specifically labeled sites in a straightforward manner, limiting the production to uniformly or nucleotide specific labeled samples in most cases.

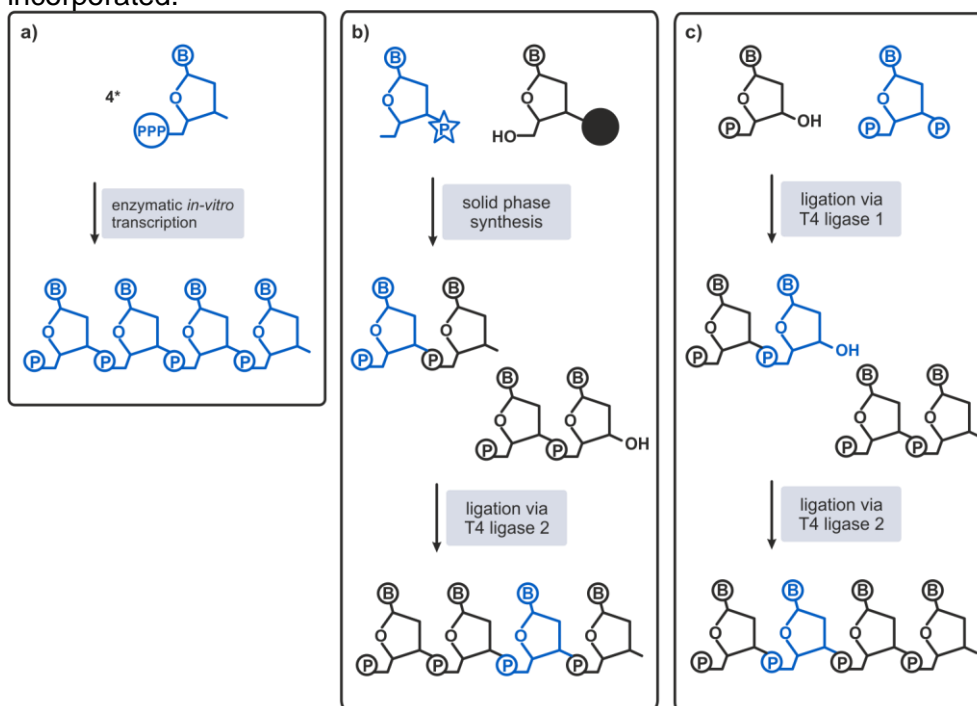
Labeling of particular RNA regions

For detailed NMR investigations exploiting the full potential of this biophysical technique, the ribozyme or particular regions of a ribozyme have to be labeled specifically.

Solid-phase synthesis is a well-known and established method for the site-specific labelling of oligonucleotides. Synthesis proceeds from the 3'- to the 5'-end and the 3'-end is bound to a solid phase. The standard procedure is a fast and automated cycle: deprotection of the 5'-end is followed by coupling with the 3'-end of a nucleoside phosphoramidite and capping of the phosphate.^[13] At this point, a modified nucleoside phosphoramidite can be implemented at the 5' end for modifications at a specific position of a long ribozyme, it is possible to ligate another RNA sequence to the 5'-end of the modified RNA. The *T4* RNA ligase 2 is applicable in this case and accomplishes ligation. The procedure is shown in Scheme 1b. Yet, the length of the target ribozyme is limited in terms of quality and quantity needed for an NMR sample.^[14,15] Side products accumulate during solid phase synthesis with increasing oligonucleotide length. Specific side reactions including 5'-2' coupling are more likely to happen in longer RNAs.^[13] Further, with every synthetic step, parts of the molecule have to be protected and the protection group has to be eliminated, leading to lower synthetic yield. The purification of the target ribozyme is another limiting step towards preparation of mg quantities required for an NMR sample. While preparative PAGE purification offers no limitation regarding the target ribozyme length, HPLC purification can only be done for a maximum length of 60 nt in average, while still maintaining nucleotide resolution.^[16] In conclusion, solid-phase synthesis becomes increasingly expensive for longer RNAs and is practically and routinely used only for short constructs. For the preparation of RNAs longer than ~80 nt, the chemo-enzymatic ligation is more suitable to overcome the solid phase synthesis limitations^[17].

The chemo-enzymatic procedure with *T4* RNA ligase 1 and 2, outlined in scheme 1c, allows the site-specific labeling of RNAs with no limitation. For this purpose, the target RNA sequence is split at certain positions according to labeling requirements. Oligoribonucleotides with an OH-group at C3' at the 3'-terminal nucleotide are ligated with isotopic labeled or modified nucleoside 3',5'-bisphosphates using *T4* RNA ligase 1.^[18,19] In a follow-up step, the ribozyme with implemented modification is dephosphorylated by RNA phosphatase e.g. rSAP. The *T4* RNA ligase 2 catalyzes the formation of a phosphodiester bond between the 3'-terminal hydroxyl from the labeled oligoribonucleotide and a 5'-terminal phosphate of the residual ribozyme sequence.^[20] This ligation step requires support from a DNA-splint, which brings the two oligoribonucleotides in close proximity. Subsequently the site specifically labeled ribozyme has to be purified, to separate the individual

ligation components. HPLC is the fastest and the most common strategy to purify the product ribozyme. However, preparative PAGE purification is often applied for longer ribozymes.^[21] The limitation of this approach is given by the type of labeling, in particular, if non-natural nucleosides have been incorporated.



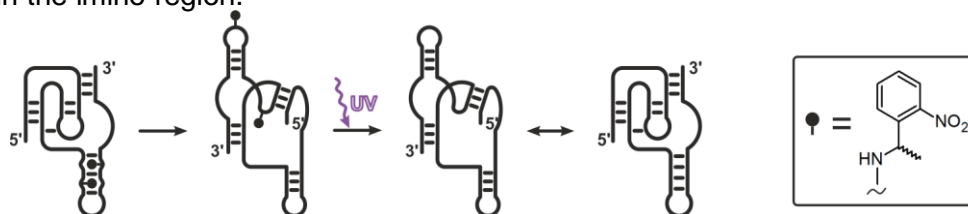
Scheme 1 Outline of three different RNA labeling methods. Labeled nucleotides are highlighted in blue. a) *In-vitro* transcription of RNA with labeled rNTPs. b) Solid-phase synthesis of RNA with side specific incorporation of one labeled nucleotide phosphoramidite and a subsequent ligation via T4 RNA ligase 2. The black dot marks the solid phase e.g. CPG, which binds the 3'-end of the synthesized RNA. c) Chemo-enzymatic ligation of RNA via T4 RNA ligase 1 and 2 with side specific incorporation of a nucleotide bisphosphate.

Photolabile caging of RNAs

The photocaging strategy is based on the incorporation of one or more nucleotides, which have a photolabile group at functional groups of the nucleotide.^[22,23] This photolabile group can prevent the nucleotide from formation of a Watson-Crick base pair with another nucleotide. By irradiation with UV-light of a certain wavelength, the modification can be eliminated, regenerating the original and inhibited function of the ribozyme (scheme 2). In this case, the eliminated photolabile group must not interfere with the structure or function of the original ribozyme.^[24,25] This strategy is widely used in cases, where the ribozyme is able to assume more than one secondary structure, to trap one of these structures. Photolabile groups are named photolabile cages too, as they 'cage' a particular secondary structure of a target ribozyme. Therefore, it is possible to screen through the conformational space of different ribozyme structures.

One of the widely used photolabile cages is the NPE – *ortho*-nitrophenyl ethyl – group.^[26,27] This photolabile cage is bound to the base pair site of each nucleotide. It protects the nucleotide from pairing with the analogous nucleotide and therefore inhibits one potential secondary structure of the ribozyme. For the analysis of such a ribozyme, an NMR sample is prepared, containing the photolabile NPR-group. That kind of samples has to be protected from light, in order to preserve the photolabile NPE-group from detachment. The examination is realized in a special NMR tube with a hollow plunger, which can be connected to a laser, while it is plugged into an NMR

spectrometer. Through this connection, real-time NMR (RT-NMR) experiments become possible.^[28] A laser pulse of 365 nm is sufficient to remove the NPE-group from the ribozyme and as a result, the caged secondary structure of the ribozyme will immediately recover. Observed NMR spectra will reveal potential rearrangements as signal-differences, especially in the imino region.



Scheme 2 Illustration of the effect caused by the incorporation of a photolabile caging group in the target RNA. The caged RNA prefers a secondary structure, where the cage does not disturb any hydrogen bonds. After irradiation with UV light, the photolabile cage is cleaved off and the RNA secondary structure is recovered. The NPE-group is shown on the right.

Initial screening of RNA constructs by NMR

A typical NMR investigation of an RNA sample starts with mapping of the conformational space the RNA adopts in solution. In case of ribozymes, for instance multiple folding topologies or stages of RNA processing can be present. Simple ^1H -1D NMR spectra provide an overview of the topological coherence of the analyzed RNA.

The most relevant resonances in this approach are the ^1H resonances of the imino protons in uracil and guanine nucleobases. In the absence of hydrogen bonds or other means to protect from solvent exchange, the exchange rate with solvent water molecules exceeds the acquisition time regime of an NMR measurement^[29] thus rendering these protons invisible to NMR. Imino protons involved in a stable base pair on the other hand resonate in the region between 10 and 15 ppm and can be detected as their exchange speed is substantially slower.^[30] These imino protons are referred to as protected from exchange. Furthermore, exchange protected imino protons in a Watson-Crick type base pair resonate in the region between 12 and 15 ppm while non-canonically base paired imino protons might exhibit a different chemical shift. These characteristics paired with the comparatively high dispersion of their chemical shifts make imino protons ideal probes for the analysis of base pair stability and conformational homogeneity of the RNA in general. Recent studies show that amino protons can be monitored in comparable ways using double quantum NMR-spectroscopical methods.^[31]

Aromatic protons of all bases resonate between 6 and approximately 8.5 ppm. Since their exchange with the solvent is neglectable (but for the H8 of purine nucleotides), aromatic protons are visible for all bases and typically feature lower linewidth. Therefore these resonances are better suited for quantification than imino proton resonances. Multiple peaks of a single well separated aromatic proton resonance hint to the presence of inhomogeneities in the sample. Therefore, peak integration can yield reliable results as possible differences in exchange are not to be expected. Unfortunately, due to the high number of aromatic protons in a longer RNA and their low chemical shift dispersion, the region is often too crowded to be properly analyzed in a 1D experiment.

The short measurement time of only a few minutes qualifies 1D-NMR for high throughput screening approaches or screenings of a single sample under multiple conditions.

Probing of general base pair stabilities can be accomplished by measuring at different temperatures while monitoring the imino proton region.

Destabilization of base pairs is instantly visible by broadening and subsequent diminishing of imino proton peaks. 1D measurements at different temperatures can serve as a valuable basis to establish experimental conditions for further characterization^[32].

Furthermore, perturbation of chemical shifts can probe the effects of interaction partners such as small molecular weight ligands or proteins on an RNA sample. Upon addition of such an interaction partner, the strength of the chemical shift perturbation gives information about the location of the binding event in case of assigned resonances.^[33] Additionally, the spectral characteristics allow the classification of the interaction into a strong intermediate or weak exchange regime. In a titration process, thermodynamic parameters such as the dissociation constant can be determined from such data.

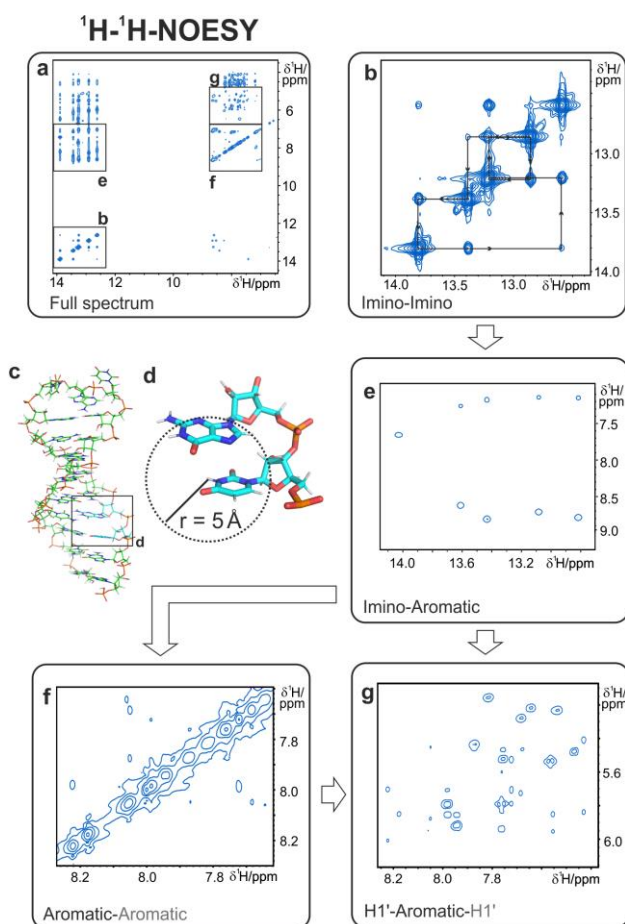
Especially in the process of functional explorations of ribozymes, the interaction with metal ions is a highly investigated topic. As magnesium ions play a catalytic role in numerous ribozymatic reactions, Mg^{2+} binding is extensively studied.^[34–36] Even though magnesium has an NMR active isotope with a natural abundance of 10%, NMR of Mg^{2+} is seldom measured. Alternatively, it is reasonable to spike samples with Mn^{2+} ^[37] or Cd^{2+} ^[38] ions to substitute for Mg^{2+} . As the addition of Mn^{2+} leads to strong paramagnetic relaxation effects, a structural placement of the binding location is possible. Cd^{2+} ions on the other hand lead to prominent chemical shift perturbation in ^{31}P -1D spectra, which will be discussed in more detail later on.

With 1D-NMR techniques, even kinetic traces can be determined in case of reasonably slow reactions while the ongoing process is simultaneously monitored on an atomic level^[26].

The high number of possible applications make 1D-NMR one of the best suited tools for the characterization and screening of RNA and ribozymes in particular, while keeping the experimental effort on an attainable level.

Resonance Assignment

In order to determine secondary structure of a ribozyme, NOESY experiments are very useful.^[39] NOESY experiments – nuclear Overhauser effect and exchange spectroscopy – are 2D experiments, which are based on the nuclear overhauser effect arising between protons of a target ribozyme. To investigate the base pairing pattern of a target ribozyme, exchangeable protons in the imino-region (15-10 ppm) have to be examined. In this case, the observation of proton resonances in the direct and indirect dimension results in a 2D NMR spectrum with diagonal- and cross-signals. Diagonal-peaks arise for every G or U residue. Cross-signals on the other hand display the NOE contacts between neighboring paired Gs and Us. Scheme 3 shows imino cross- and diagonal peaks in the top panel.



Scheme 3: Graphical representation of the assignment of an RNA, utilizing ¹H-¹H-NOESY and a 3D structure calculated from NOE data. In case of inter-base connectivities, the neighboring base is written in grey below the respective spectrum. a) Relevant region of a NOESY of RNA to facilitate assignment of imino, aromatic and H1' protons. Zooms of the regions annotated are shown in b), e), f) and g). c) Three dimensional Structure calculated from NOE data. d) Zoom of two adjacent bases, outlining the range of the NOE. The spectrum of a 21nt RNA was measured at 700 MHz.

In general, a so called *iminowalk* can determine elements of the secondary structure by displaying paired bases, due to spatial proximity. In all remaining cases, paired bases indicated by diagonal signals in this kind of 2D spectrum, do not have any related paired base in their close surrounding (<5 Å). To differentiate between the bases, it is essential to have a look at the positions and intensities of the respective signals. G:U base pairs for instance, are usually located in the region from 12-10 ppm and exhibit an intense NOE crosspeak. With the information on the imino region it is possible to examine NOE contacts to the aromatic H2 protons as well as the amino protons of cytosine residues. As especially the cytosine amino protons lie in close proximity to the aromatic protons and aromatic protons in the stem are typically within a distance below 5 Å to each other, an assignment of these is possible from the NOESY alone. The aromatic assignment can be obtained as well from NOE to the intraresidual H1' protons and their NOE to the aromatic proton of the neighbouring base. Further, NOE contacts from the aromatic spin system to the sugar protons of a target ribozyme enable these assignments in principle as well. Unfortunately, the peak dispersion for sugar protons other than H1' is poor. As a consequence, this assignment becomes more difficult with larger ribozymes, due to more crowded spectra.

Resonance assignment by uniform labeling

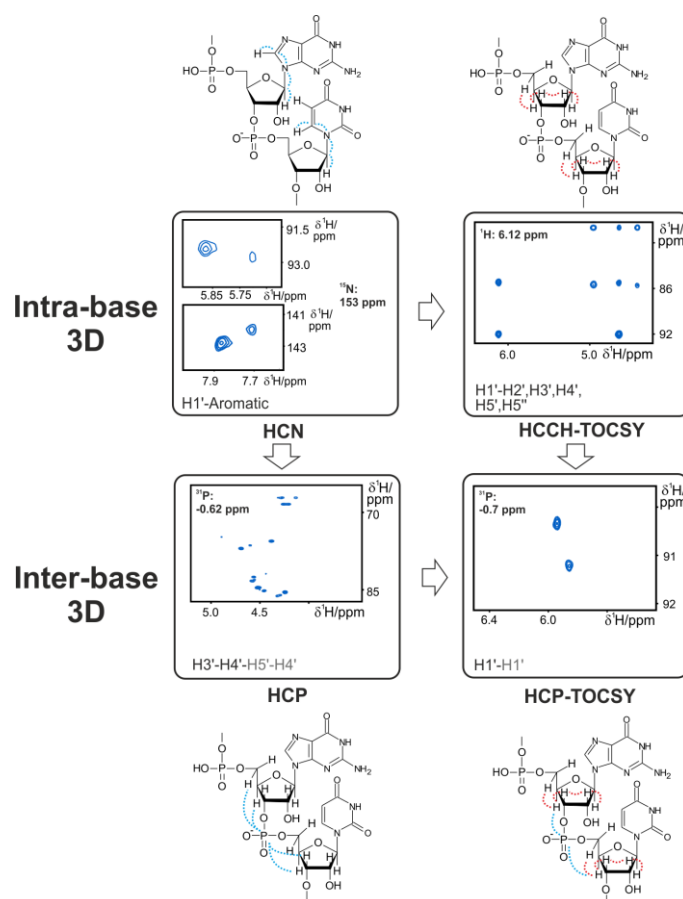
In order to facilitate the proton assignment heteronuclear NMR experiments have to be applied, assuming ^{15}N and/or ^{13}C labeled ribozyme samples (scheme 4).

These isotopically labeled RNA samples allow a multitude of experiments which are inaccessible at natural abundance. Most commonly, for a resonance assignment, HCP^[40], HCP-TOCSY^[41], HCN^[42] and HCCH-TOCSY^[43] are applied to monitor connectivity of H3' and H5' sugar protons of neighboring bases, H1' of neighboring bases, intra-base aromatic to H1' correlation and assignment of sugar protons based on H1' resonances respectively. In contrast to a NOESY assignment, the assignment with the help of 3D experiments is performed utilizing coupling between nuclei, leading to unambiguous assignments.

Resonance assignment by selective labeling

Selective labeling is a method, which allows the detection of one single nucleotide by NMR spectroscopy. Assuming a single $^{13}\text{C}^{15}\text{N}$ labeled residue in a ribozyme, an unambiguous assignment of proton resonances of the labeled residue can be achieved, running ^{13}C or ^{15}N X-edited ^1H -1D experiments. Assignment of the hetero nuclei takes place via ^1HX -HSQC. This allows in principle the full resonance assignment of a ribozyme if a selectively labeled sample for each residue is available. The isotopic enrichment for specific residues can be as low as 1%.^[44]

An alternative approach, which targets specifically the assignment of the sugar region is given by sparse labeling within the sugar. Bases with sugar moieties, in which only specific carbons are labeled can be prepared chemically or enzymatically, while the required NTPs are accessible by in vivo production in *E. coli*.^[45] Common labeling schemes include C3' and C5' labeling or C1', C2' and C4' labeling. While this type of isotopic labeling requires considerable preparative effort, the resulting samples display significantly more resolution in the sugar regions and can simplify their assignment in huge RNA constructs.



Scheme 4: Graphical representation of further assignment of an RNA, utilizing 3D heteronuclear NMR experiments. The magnetization transfer is depicted below the spectra, with INEPT transfers shown in blue and TOCSY transfers shown in red. Arrows depict how the information from one spectrum is conferred to the other spectra. Spectra are measured on a 21 nt RNA at 600 MHz.

NMR-based characterization of particular ribozymes

Hepatitis delta virus – HDV

The human hepatitis delta virus ribozyme is known for its ability to cleave itself, yielding a 5' hydroxyl residue and a 2',3'-cyclic phosphate group.^[46] Some early studies found great similarity between the genomic and antigenomic HDV secondary structures.^[46,47] Kou *et al.* and Perrotta *et al.* showed that it is possible to reduce the genomic HDV ribozyme sequence to a minimum, which is still able to cleave itself.^[48–50] Further investigation of Perrotta *et al.* and Rosenstein *et al.* based on an antigenomic minimal HDV sequence identified a pseudoknot structure as key structural motif required for proper cleavage.^[47,49]

In order to have further insight into the secondary and tertiary structures of HDV ribozymes, NMR can be applied. Lee *et al.* performed NMR experiments on a 74 mer HDV ribozyme.^[51] Melting of the ribozyme was measured via ^1H -1D experiments recorded at temperatures between 12 °C and 77 °C. The proton resonances of non-Watson-Crick bound loop regions - shifted to 9-12 ppm - became less sharp at higher temperatures. This effect was attributed to faster solvent exchange. On the other hand, proton resonances located at 12-15 ppm did not change at different temperatures indicating higher stability for base pairs in the stems. Further, Lee *et al.* were able to observe a signal from an A-U imino proton at 14.28 ppm, which split into two proton resonances at

12-16 °C indicating the existence of two conformations in slow exchange. The spectra are shown in figure 2.

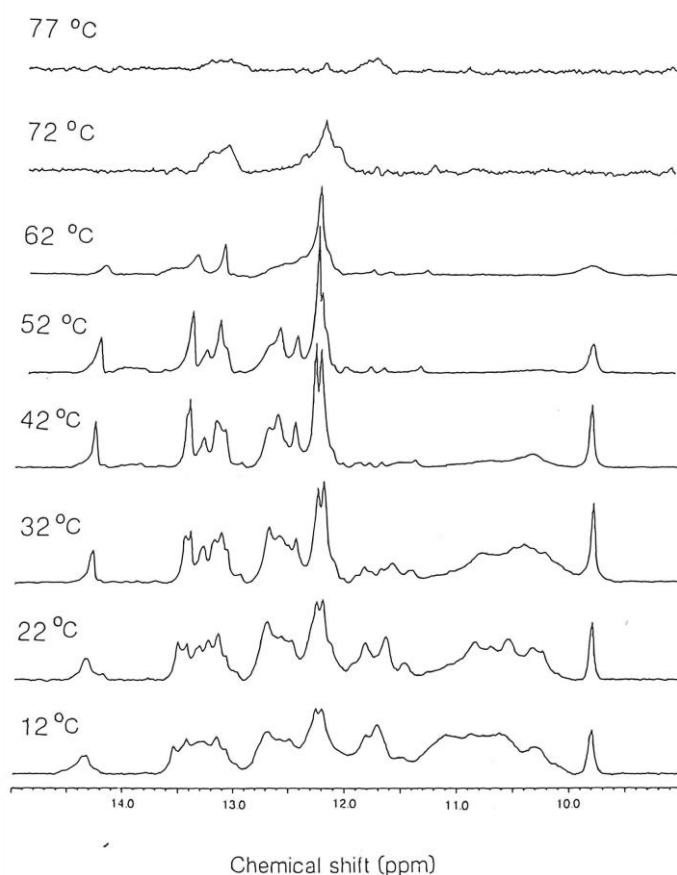


Figure 2 ^1H -1D NMR spectra of the HDV ribozyme measured and published by Lee *et al.*^[51] The spectra were recorded at different temperatures in order to investigate the melting of the ribozyme. The split signal of an A-U imino proton at 14.28 ppm appearing at temperatures between 12 °C and 22 °C points towards two conformations. The original picture was scanned from a written issue and vectorized to improve the quality with permission of the original publisher

For the assignment of the imino proton resonances, ^1H - ^1H -NOESY spectra were analyzed. These 2D spectra exhibit NOE contacts between almost all imino and amino protons. The NOE cross-peaks are located at 6.7-8.2 ppm and are in general only visible for G-C base pairs. Two of the imino signals at 13.47 ppm and 14.28 ppm, however display one NOE contact to the aromatic H2 proton between 7 and 8 ppm referring to the presence of A-U base pairs. Moreover, tetra-loop formation has been observed due to three imino resonances belonging to adjacent bases as suggested by inter-base cross-peaks. In conclusion, Lee *et al.* demonstrated that the 74 nt long HDV ribozyme assumes multiple conformations, while its major conformer is in agreement with the proposed pseudoknot structure.

The role of metal ions (Mg^{2+} and Na^+) in the cleavage reaction of HDV ribozymes was investigated by Ganguly *et al.*^[52] by MD simulations. To independently cross-validate the theoretical data, NMR spectroscopy was applied. In order to show, whether cleavage is preceded by metal-mediated protonation of the 2'-hydroxyl, 3'-adenosine monophosphate was titrated with NaOH and $\text{Ca}(\text{OH})_2$ and ^1H -1D NMR spectra were acquired. Samples were prepared with 10% D_2O , as a 100% D_2O sample tends to feature an N-linked proton pKa shift of about 0.4-0.6 units.^[53,54] The chemical shift of H1' proton resonances, which are far apart from the water resonance, were displayed as a function of the pH. In conclusion, divalent ions seemed to have a greater

effect on the cleavage activity, involving a precluded protonation of the 2'-hydroxyl at the cleavage site.

Other high resolution NMR research was done by Kolk *et al.*^[55] They used an isolated hairpin sequence from the antigenomic HDV ribozyme, to examine the pseudoknot secondary structure element. They chose a sequence containing 7 nt, which are important for cleavage activity. This sequence was extended to a 19 nt long hairpin loop. ¹H-¹H NOESY and ¹H-¹H TOCSY NMR spectra were recorded and most of the proton resonances were assigned. The NMR data were used to calculate a more precise loop conformation by MD simulations.

The crystal structure of HDV ribozyme published by Ferre-D Amare *et al.*^[56], was approved by Tanaka *et al.* by screening HDV sequences with ¹³C¹⁵N labeled Gs, designed in respect of the pseudoknot model.^[57-59] For NMR analysis the ribozyme sequences were split into three separate ribozyme parts, to facilitate the proton assignment. The imino proton resonances from ¹H-¹H-NOESY and ¹H-¹⁵N-HSQC experiments were partially assigned, pointing out the similarity with the crystal structure.

As a starting point the imino proton resonances of a G*U base pair were assigned first, since the resonances are shifted to the range of 10-12 ppm and exhibit a strong NOE cross signal. The other NOE cross signals were assigned via a walk through the imino region of the NOESY spectrum. Further, NMR samples containing a substrate sequence were prepared. NOESY and HSQC spectra of these full-length constructs did not show any major changes, except the imino proton resonances of the substrate. In order to determine temperature and Mg²⁺ dependencies, ¹H-1D spectra were recorded at different conditions. The findings hinted a catalytic effect of Mg²⁺ during cleavage reactions. An other HDV ribozyme based publication from Andrej Luptak *et al.* considers pK_a determination via ¹³C NMR measurements.^[60] They focused on a C suspected to be involved in the cleavage reaction and examined the ¹³C chemical shifts of this carbon atom, which allows the calculation of the pK_a value.

In order to understand the effect of mutational evolution, Schultes *et al.* have analyzed the interconversion of a shortened HDV ribozyme (HDVP) and a synthetic class III ligase (LigP), as both catalyze fundamentally different reactions, do not have any related structural elements and do not share any evolutionary history.^[61] They connected both prototypes by creating a neutral network, which is defined as a set of sequences with various genotypes but a specific phenotype. Starting with one prototype sequence 86 neutral mutations were introduced to generate the other prototype sequence, whereby each mutation step passes on the biologically important activity of the respective prototype. As a result, they show an intersecting sequence (INT) in the middle of this network, a ribozyme featuring catalytic activities of both prototypes.

At this point, NMR spectroscopy can be very helpful to screen through conformational space, to understand the correlation between sequence, structure and function of the neutral network. The screening through conformational space can be achieved by the preparation of NMR samples of particular constructs exhibiting strong change in structure and/or structural dynamics. Hence, we choose five ligase constructs and four HDV constructs from the neutral network, situated between the intersecting sequence and the particular prototype.^[62] Figure 3 shows the imino proton region of the ¹H-1D NMR spectra of the selected constructs.

The ¹H-1D NMR spectra of the constructs INT, Lig1 and HDV1 show very inhomogeneous but similar imino proton resonances, without alteration of

chemical shifts although the catalytic activity increases strongly with the incorporation of mutations into the sequence of INT. This indicates the existence of multiple secondary structures for one sequence and shows that the increase in catalytic activity is caused by the incorporation of mutations but not by a secondary structure rearrangement. Moreover, the homogeneity enhances with the incorporation of mutations converging in the direction of the respective prototype, suggesting an improvement of secondary structure stability. In order to further investigate the structure of INT, it was forced into just one structure by caging two nucleotides with a photolabile group and detecting the refolding process by real-time NMR spectroscopy.^[63]

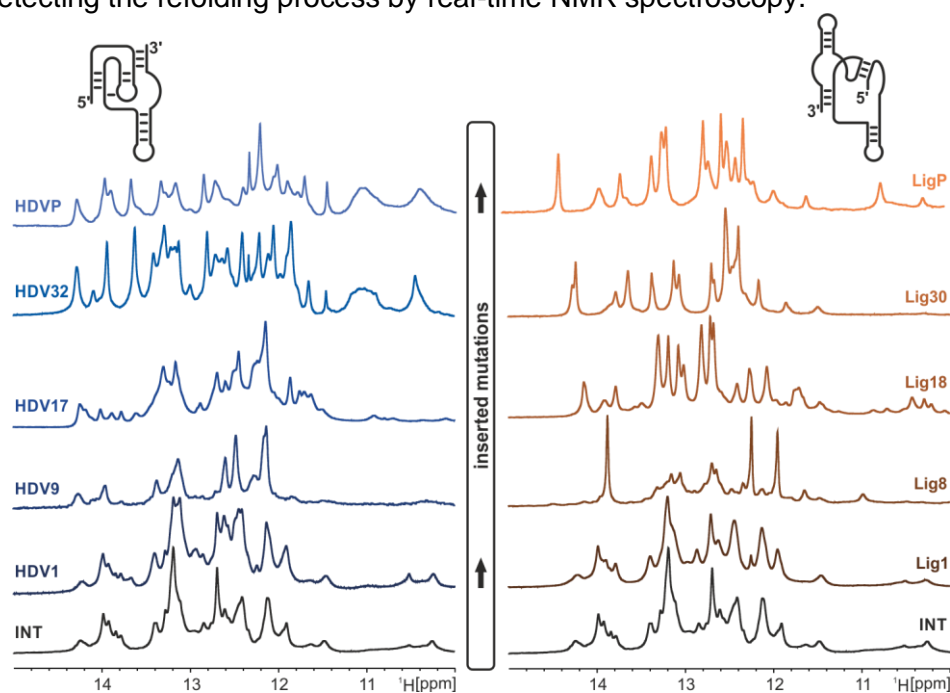


Figure 3 The imino proton regions of ^1H - $^{1\text{D}}$ NMR spectra of different ribozyme sequences are shown.^[62] The sequences were obtained from a neutral network of a HDV ribozyme and a class III ligase.^[61] The spectra were acquired via 'jump-and-return' experiments using an 800 MHz Bruker spectrometer. The samples were set up to a concentration of 100-400 μM in phosphate-buffer. The secondary structures and the direction of mutation insertion are shown.

The investigation of ligase ribozymes was first studied by R. Naylor and P. T. Gilham in 1966.^[64] Further, *in-vitro* evolution based screenings could show the existence of autocatalytic systems of RNA template-directed condensation of two RNA substrates.^[65] An efficient self-replicating ribozyme R3C was found by Natasha Paul and Gerald F. Joyce,^[66,67] leading to the discussion of the origin of life with reference to the 'RNA-world' hypothesis.

Hammerhead ribozyme

The hammerhead ribozyme is a self-cleaving ribozyme with a structural motif consisting of three stems and a highly conserved catalytic core. The ribozyme was first discovered in plant pathogens and viral satellites as a part of a rolling circle replication mechanism.^[68] Catalytic activity is achieved by Mg^{2+} -mediated formation of an in-line conformation around the scissile phosphodiester bond and subsequent cleavage reaction.^[69] In several studies, NMR was employed to investigate structural and functional aspects of the hammerhead ribozyme.

After the determination of the crystal structure^[70,71], additional structural characterization was carried out in solution, utilizing NMR. In depth characterization of the catalytic core took place by ^1H - ^1H -NOESY, ^1H - $^{1\text{D}}$ as well as ^{13}C -HSQC spectroscopy, hinting at structural changes leading to an

active conformation, different to the crystal structure. These studies were carried out in absence and presence of Mg^{2+} ions and structural differences between cleaved and uncleaved construct were assessed.^[72] Further studies were carried out determining the structure in solution and the formation of the active conformation also involving mutation studies, relying mostly on 1H - 1H -NOESY data^[73-75]. The global structure of the hammerhead ribozyme was determined by measurement of residual dipolar coupling (RDC), showing fundamental differences from the crystal structure^[76]. In addition, an NMR structure of a wide-spread loop-loop interaction motif could be determined in 2009 using 1H - 1H -NOESY data and various 3D methods on ^{13}C ^{15}N labeled samples to determine structural parameters.^[77]

NMR-aided functional investigation of the hammerhead ribozyme centers mostly on the characterization of the active state of the ribozyme and the interstate transitions. As the catalytically active state involves binding of a divalent cation, the determination of metal binding sites and the thermodynamics and kinetics of such a binding are of high interest. In 1997, Herschlag *et al.* could show that phosphorothioate modification on the scissile phosphate of the ribozyme had deleterious effects on magnesium binding. The affinity could be restored by spiking with small amounts of Cd^{2+} .^[78] As this modification induces a considerable chemical shift in the ^{31}P signal of the modified phosphate, it enables unambiguous assignment of its resonance and therefore ^{31}P monitoring of ribozyme activity on a specific ^{31}P -1D peak. Subsequent studies allowed the localization of metal binding sites in the catalytic core by either ^{31}P CSP analysis^[79] or exploitation of the deleterious effect of manganese spiking to the bound phosphates signal due to paramagnetic relaxation^[80]. While the phosphorothioate method led to significant results in the characterization of metal binding, possible structural alterations introduced by the modification could not be ruled out^[81]. In following studies the method was often avoided in favor of less invasive strategies. Lilley *et al.* were able to determine a folding model upon metal interaction by ^{19}F modification of uridine residues in and next to the catalytically important CUGA sequence. The resulting ^{19}F resonances were monitored in 1D transient studies and revealed a two-step folding model.^[82] ^{15}N detected metal binding essays for the identification of metal binding sites were introduced, using selectively and site-specifically ^{15}N labeled RNA and exploiting the large CSP introduced by metal binding^[83,84]. Extensive NMR studies in the late 2000s helped the in-depth characterization of the catalytic process of the hammerhead ribozyme. Uncoupling the metal binding event from the cleavage reaction by incorporation of a photocleavable protecting group both processes could be monitored separately in atomic resolution^[85].

Hairpin ribozyme

During the investigation of plant satellite RNA in the late 1980s, the hairpin ribozyme was first identified in negative-strand satellite RNA of the tobacco ringspot virus^[86]. The short ribozyme catalyzes cleavage and joining reactions releasing a substrate with a 5'-hydroxyl group leaving the ribozyme with a 2',3'-cyclic phosphate upon cleavage. It consists of four helical stems and two unpaired loops with the scissile phosphate located in one of the loops. Nucleobases in the loop regions generally termed loop A and B are highly conserved throughout species.^[87]

Structural investigation of the hairpin ribozyme by NMR spectroscopy mainly took place in the late 1990s. For the data collection preceding the calculation of the solution structure of loop A region selective ^{13}C -labeling was employed to facilitate the assignment of aromatic protons and to allow the application of ^{13}C -filtered NOESY experiments. Along with the data collected from

conventional NOESY and ^{13}C -NOESY-HMQC experiments the structure could be determined.^[88] Further structural studies naturally focused on the solution structure of the loop B which was determined few years later. In the according study, 3D-HCCH-TOCSY along with the typical 2D-NOESY and 2D-TOCSY spectra were used for assignment, while the NOE data was gathered from 2D-NOESY and ^{13}C -3D-NOESY-HMQC spectra.^[89]

NMR spectroscopy was further utilized in studies characterizing the function of the hairpin ribozyme. Determination of the metal binding sites took place by chemical shift mapping, paramagnetic line broadening and intermolecular NOE. Corresponding binding sites could only be identified in loop B, though no structural change was indicated upon binding of Mn^{2+} , Mg^{2+} or $\text{Co}(\text{NH}_3)_6^{3+}$, leading to the assumption, that metal ions do not facilitate the catalytic activity but help in maintaining a correct tertiary structural fold.^[37] Monitoring the ^{13}C chemical shift perturbation upon pH change, pK_a values significantly different from the normally observed values could be determined for loop adenine residues. It was concluded, that protonation of adenine plays a role in catalytic activity of the ribozyme.^[90]

The hairpin ribozyme served as a model system in later NMR studies employing atom and site selective labeling of the loop A domain proving that six ^{15}N residues can be labeled and unambiguously distinguished by indirect ^{13}C tagging.^[91] In a later study, the adenine binding mutant of the hairpin ribozyme (ADHR1)^[92] and its catalytic mechanism was investigated. Kinetic experiments and monitoring of ^{15}N -HSQC, ^{13}C -HSQC and ^{31}P -1D spectra under various conditions revealed a significantly higher reaction barrier than in the wild type ribozyme with dependence on the adenine cofactor.^[93]

Neurospora VS ribozyme

In the class of nucleolytic ribozymes also consisting of HDV, hammerhead and hairpin ribozymes, the Neurospora Varkud Satellite ribozyme is the largest with a length around 150 nucleotides. It catalyzes ligation or cleavage by transesterification on a specific loop site of a substrate and can act in cis or in trans. It is composed of six helical structures with two three-way junctions.^[94] As a crystal structure of the ribozyme could only be determined recently^[95], structural investigation was carried out via NMR and started in 2000 with the structure determination of the substrate stem-loop. Distance restraints for the structure calculation were determined from ^1H - ^1H -NOESY while COSY and TOCSY type experiments yielded the assignment.^[96] Subsequent NMR spectroscopic studies of the neurospora VS ribozyme were mostly carried out in the Legault lab and focused on the determination of the global structure. An NMR structure of a mutated stem-loop I, which mimics the active cleavage site of the ribozyme was determined with the internal loop featuring a G-A sheared base pair and a A-G-A base-triplet.^[97] The structure of stem-loop V was solved in presence of Mg^{2+} ions, with the binding sites determined by paramagnetic relaxation enhancement of a Mn^{2+} spiked sample monitored by ^{13}C -HSQC experiments. Four distinct binding sites could be determined in the loop region.^[98] In 2008 kinetics and substrate binding of the stem-loop V were investigated. By combining magnesium dependent ^{31}P and ^{15}N chemical shift perturbation with native PAGE data for several mutants it became apparent that the loop residues tolerate a multitude of mutations retaining binding of the substrate, while deletion of U700 substantially reduces activity.^[99] Further structural studies addressed the A730 loop located in helix VI^[100] which is involved in a loop-loop interaction not unlike the hairpin ribozyme^[89] as well as the substrate recognition site^[101] and the III-IV-V junction. Distance and angular restraints were determined from ^1H - ^1H -NOESY data while magnesium binding sites in the III-IV-V junction structure were determined by Mn^{2+} induced

paramagnetic relaxation enhancement.^[102] Shortly afterwards, the structure of the II-III-VI junction was determined as well, showing remarkable structural diversity in the junction core including an A•G base pair and minor groove base triplets. For the unambiguous resonance assignment in the junction an A-specific H(NC)-TOCSY-(C)H experiment and 3D HCCH-TOCSY as well as HCCH-COSY experiments were utilized. Summarizing all the structural data determined for the separate elements, a model of the full structure of the 150 nucleotide ribozyme could be determined by NMR.^[103]

Leadzyme

Selection experiments opting to find RNA cleavage reactions catalyzed by Pb²⁺ ions yielded several ribozymes first described by Pan *et al.*^[104] Later the same year they derived a 30 nucleotide ribozyme from a suggested alternative secondary structure of one of the cleavage active variants.^[105] It features an asymmetric six nucleotide internal loop and is rapidly cleaved in the presence of Pb²⁺. This specific structure, one of many lead ribozymes, is referred to as leadzyme.

Legault *et al.* used heteronuclear NMR experiments to investigate the leadzyme. They observed a significant upfield shift for the C2 resonance of adenine 25 (A25).^[106] The C2 resonance shifts 8 ppm with a pH range from 4.4 to 8.3 and a pKa of 6.5 was determined. A25 is in close proximity to the site of cleavage.

A solution structure of the leadzyme was determined by Hoogstraten *et al.* and a model of a lead binding bulged guanine base critical for ribozyme function was proposed.^[107] A feature of the structure is a protonated AH⁺-C base pair between the cleavage site and the regular A-Form RNA. Due to the involvement in a base pair it is unlikely that the protonated A25 participates in a "histidine-like" catalytic mechanism. The structure also showed a conformation that requires rearrangement to reach the trigonal bipyrimidal transition state proposed for ribozyme reaction. The chemical shift analyses showed that Pb²⁺, Mg²⁺ or Na⁺ addition does not result in large structural rearrangement. Complementary studies by Katahira *et al.* also found little change in the imino spectrum upon addition of lead ions.^[108] An analysis of the relaxation rate along the B1 field (T1ρ) of the purine C8 showed diverse dynamics for the residues of the internal loop.^[107] The longer T1ρ of A8 compared to helical purine regions indicated a sub-nanosecond fluctuation. In contrast G7 and G24 fluctuate on a micro- to millisecond timescale. These results were also found in structural probing assays using chemical modification. The low solvent accessibility of G24 could be explained through base stacking with G23 and A25.

Ribonuclease P

Ribonuclease P is a large multi turnover ribozyme that cleaves 5' ends of pre-tRNA *in vivo* facilitating its maturation. It consists of a protein domain and an RNA domain which are both essential *in vivo* while the RNA alone is proven to be catalytically active *in vitro* and is therefore considered a ribozyme with a protein cofactor.^[109] The large RNA domain comprises 18 helical stems with the catalytic center located around helix P4.^[110] NMR assisted structural investigation of RNase P began in 1996 with the structure determination of the 31 nucleotide sequence around the conserved substrate binding GGU-motif. To facilitate the resonance assignment prior to the structure calculation, cytidines were partially deuterated in this study, simplifying the spectra especially in the sugar regions.^[111] Structural research of RNase P moved on not much later with the determination of the NMR structure of the RNase P protein from data obtained in ¹³C and ¹⁵N correlated 3D-NOESY experiments.^[112] Investigation of the RNA proceeded with the structure of helix

P4 which is located in the catalytic core next to the binding site of the pre-tRNA. Metal binding sites were determined, by adding $\text{Co}(\text{NH}_3)_6^{3+}$ ions which gave rise to ^1H - ^1H -NOESY cross-peaks to adjacent protons.^[113] Another study could show that the pre-tRNA as well binds Mg^{2+} ions even prior to RNase P binding.^[114] For the highly resolved NMR structure of helix P5.1 from bacterial RNase P type II the structure calculation was carried out with distance restraints from ^1H - ^1H -NOESY. Dihedral restraints from ^1H - ^{31}P -hetero TOCSY, HCP and HMQC-TOCSY as well as residual dipolar coupling (RDC) data were implemented for refinement, resulting in a high resolution structure.^[115]

Further research focused mainly aspects of ribozyme function employing a variety of methods to assess metal binding and dynamics. The highly conserved and catalytically important C and U residues, forming a bulge in the center of the P4 helix, could be identified as a pivot point for structural rearrangement upon Mg^{2+} binding. RDC and ^{15}N relaxation data were measured in presence and absence of Mg^{2+} revealing remarkable flexibility around these specific residues while mapping the global conformation of the helix.^[116] In the protein domain, the RNA binding surface was determined flexible in contrast to the protein being relatively rigid in general, as shown by ^{15}N relaxation data of the backbone.^[117] A three step folding pathway was proposed by monitoring CSP in ^{15}N -HSQC spectra a few years later. Metal binding sites of the protein could be localized using paramagnetic relaxation enhancement by adding $[\text{Cr}(\text{CN})_6]^{3-}$.^[118] Metal binding of the RNA domain was reviewed in 2010 when EXAFS spectroscopy data shed new light on the metal binding sites of the P4 helix, characterizing the binding of a Zn^{2+} ion in the catalytic core. Complexation of a Mg^{2+} ion between two conserved G residues as found by PRE of $\text{Co}(\text{NH}_3)_6^{3+}$ is described in great detail showing a six-coordinated geometry.^[119]

Other studies concerned the stability of the pre-tRNA substrate stem whose opening is considered a rate limiting step in substrate binding. Kinetics of pre-tRNA constructs with several modifications were determined while non invasiveness was checked via ^1H -1D NMR of the imino region.^[120] NMR proves to be particularly useful when hard evidence of structural change is needed, as shown in a binding assay of the whole RNase P complex.^[121]

Group I intron

Group I introns have the ability to self-splice. Therefore, the diverse transcripts known for this ribozyme, include a guanosine binding site.^[122,123] This binding site is essential for the self-splicing process, which is composed of two transesterification steps requiring a bound guanosine.^[124] Due to the large size of the group I intron, the different regions have been studied independently by NMR spectroscopy. A 20 nt long RNA with a UUGC tetraloop was studied by Frederic H.-T. Allain *et al.*^[125] This tetraloop construct has the properties of the P1 helix of the group I intron, which is located at the cleavage site. Using the assignment data of ^1H -1D, ^1H - ^1H -NOESY, ^{13}C correlated 3D-NOESY, ^1H - ^{13}C - ^{31}P HCP and H-COSY-H,C-HMQC spectra, it was possible to simulate a solution structure of P1 by MD simulations. The stability of P1 was investigated by Joon-Hwa Lee *et al.*^[126] NH_3 and HPO_4^{2-} have been used as base catalysts for the base pair opening. The effect of those was observed by ^1H -1D and ^1H - ^1H -NOESY experiments and based on the imino proton exchange rates, the equilibrium constants have been determined. Mayumi Amano investigated a 17 nt long RNA from the loop 2.1 region of group I intron by NMR spectroscopy.^[127] With the help of NOE contacts from ^1H -1D and ^1H - ^1H -NOESY spectra, imino proton, H1' and sugar proton assignments were accomplished. Further, Mayumi Amano studied a 20 nt RNA from the P9.1 region of group I intron.^[128] Therefore, imino proton resonances from ^1H -1D

spectra were assigned. Another investigation of both hairpin structures, the 17 mer and 20 mer, showed low-range interactions between those.^[129] A 25 nt long model RNA from the P4-P6 region of group I intron was analyzed by ¹H-¹H-NOESY, COSY and natural abundance ¹H-¹³C HMQC experiments.^[130] Data from assigned NOE contacts were converted to constraints for MD simulations of the structure. Michael Chastain *et al.* concluded, that a triple-helix is formed in the minor groove of the ribozyme. One year later, Michael Chastain and Ignacio Tinoco, Jr. extended this study.^[131] Other MD simulations of a slightly larger RNA model of P4-P6 were based on the resonance data from ¹H-1D, ¹H-¹H-NOESY, COSY, TOCSY and x-filtered NOESY spectra.^[132] The largest model construct of P4-P6 was investigated by Eric C. Johnson *et al.*^[133] The NOESY and WaterLOGSY spectra were received including a screening through a ligand library. The resonance changes of free ligands due to faster or slower tumbling were recorded, showing which library compound is bound to the ribozyme. One specific part of the P4-P6 domain is the P5 helix. One of the first studies were done by Jeffrey A. McDowell *et al.*^[134] They chose a duplex motive similar to P5 and determined the 3D solution structure of it by MD simulations. The data needed for these calculations were extracted from ¹H-1D, ¹H-¹H-NOESY, DQF-COSY and ¹H-³¹P-HETCOR spectra. Furthermore, the solution structure of a 25 nt long model RNA including a 5 nt sized bulge loop was defined, based on 1D proton spectra, ¹H-¹H-NOESY, COSY, ¹H-¹³C-HMQC, TOCSY and homonuclear 3D-TOCSY-NOESY spectra.^[135] Another NMR based solution structure study incorporated Co(NH₃)₆³⁺ and Mg(H₂O)₆²⁺, in order to test secondary structure dependent metal-ion binding.^[136] For the analysis of the P5abc region of group I intron, this ribozyme domain was truncated, yielding a 56 nt long RNA (tP5abc).^[137] NOE signal assignment was done for ¹H-1D, ¹H-¹H-NOESY spectra. An additional titration with Mg²⁺ did not cause any secondary structure rearrangement. In a different study, Co(NH₃)₆³⁺ was used to determine ion-binding sites.^[138] Therefore ¹H-1D, ¹H-¹H-NOESY, ¹H-¹⁵N-HMQC, HNN-COSY spectra have been recorded and the extracted data was used for MD simulations. Minxue Zheng *et al.* investigated secondary structure changes, caused by tertiary structure formation, applying ¹H-1D, ¹H-¹H-NOESY, DQF-COSY (double quantum filtered), TOCSY, fast HSQC, INEPT, HMQC experiments and restrained MD simulation.^[139] The native structure of tP5abc was analyzed by Hashim M. Al-Hashimis group.^[140] The combination of ¹H-1D, ¹H-¹H-NOESY, ¹H-¹⁵N-HSQC, ZZ-exchange NMR experiments and SHAPE data gave rise to the hypothesis, that the P5c region independently forms a native structure.^[141] Furthermore, Brant Gracia *et al.* point out the great ability of single base pairs, to stabilize secondary structures.^[142] They applied SOFAST-HMQC experiments on a P5abc construct combined with the intron core and later on incorporated a single mutation. Satoru Watanabe *et al.* investigated the guanosine binding site.^[143] The imino proton resonances from ¹H-1D, ¹H-¹H-NOESY spectra for a 31 nt long model RNA of P7, P9.0 and a terminal 3'G residue have been assigned. Additional studies on the P7-P9 domain were done by Aya Kitamura *et al.* including ¹H-¹³C-HSQC, ¹H-¹⁵N-HSQC, ¹H-¹H-NOESY and COSY experiments and subsequent structure calculations, in order to determine the solution structure.^[144,145]

Group II intron

The group II intron is a self-splicing ribozyme differentiated into six domains.^[146] Lan Zhang *et al.* have achieved the crystal structure of the domains 5 and 6.^[147] To further investigate the secondary structure of this ribozyme, the solution structures of both domains have been characterized successfully by NMR spectroscopy. Assignment was achieved by analysis of

¹H,¹H-NOESY with aid of 3D-HCCH-TOCSY, HNN-COSY experiments. The extracted NOE restraints led to a defined structure of domain 6 with an RMSD of 4.2 Å. The high RMSD is attributed to the overall flexibility in the bulge region leading to a global conformational heterogeneity.^[148] The structure of domain 5 showed flexibility in the loop region which was consistent with the broad peaks observed in the spectra. Residual dipolar coupling analysis was therefore only applied for stem residues.^[149]

Closing remarks

This overview points out, that some ribozymes are already well analyzed by NMR spectroscopy. This method provided a good insight into the diverse structural elements of the ribozymes shown in Figure 1. Further, it was possible to characterize kinetic and dynamic aspects of the catalytically active molecules, due to a multitude of experiments and applicable conditions. It should be further noted, that the benefit of NMR spectroscopy, in contrast to crystallography, is the capability of mimicking natural conditions.

- [1] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem* **2003**, *4*, 936–962.
- [2] A. V. Cherepanov, C. Glaubitz, H. Schwalbe, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 4747–4750.
- [3] W. Huang, M. F. Bardaro, G. Varani, G. P. Drobny, *J. Magn. Reson.* **2012**, *223*, 51–54.
- [4] A. Marchanka, T. Carlomagno, *eMagRes* **2014**, *3*, 119–128.
- [5] A. Marchanka, B. Simon, G. Althoff-Ospelt, T. Carlomagno, *Nat. Commun.* **2015**, *6*, 1–7.
- [6] R. T. Batey, J. D. Puglisi, J. R. Williamson, M. Inada, E. Kujawinski, *Nucleic Acids Res.* **2007**, *20*, 4515–4523.
- [7] S. Quant, R. W. Wechselberger, M. A. Wolter, K.-H. Wörner, P. Schell, J. W. Engels, C. Griesinger, H. Schwalbe, *Tetrahedron Lett.* **1994**, *35*, 6649–6651.
- [8] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
- [9] R. Losick, *Annu. Rev. Biochem.* **1972**, *41*, 409–446.
- [10] M. Chamberlin, J. Ring, *J. Biol. Chem.* **1973**, *248*, 2235–2244.
- [11] J. Guillerez, P. J. Lopez, F. Proux, H. Launay, M. Dreyfus, *Proc. Natl. Acad. Sci.* **2005**, *102*, 5958–5963.
- [12] C. Helmling, S. Keyhani, F. Sochor, B. Fürtig, M. Hengesbach, H. Schwalbe, *J. Biomol. NMR* **2015**, *63*, 67–76.
- [13] W. S. Marshall, R. J. Kaiser, *Curr. Opin. Chem. Biol.* **2004**, *8*, 222–229.
- [14] G. C. Walker, O. C. Uhlenbeck, E. Bedows, R. I. Gumpert, *Proc. Natl. Acad. Sci.* **1975**, *72*, 122–126.
- [15] R. Silber, V. G. Malathi, J. Hurwitz, *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 3009–3013.
- [16] Z. Huang, S. Jayaseelan, J. Hebert, H. Seo, L. Niu, *Anal. Biochem.* **2013**, *435*, 35–43.
- [17] M. K. Smalley, S. K. Silverman, *Curr. Protoc. Nucleic Acid Chem.* **2004**, *19*, 11.11.1–11.11.12.
- [18] T. E. England, O. C. Uhlenbeck, *Nature* **1978**, *275*, 560–561.
- [19] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859–1862.
- [20] T. E. England, O. C. Uhlenbeck, *Biochemistry* **1978**, *17*, 2069–2076.
- [21] M. Hengesbach, A. Kobitski, F. Voigts-Hoffmann, C. Frauer, G. U. Nienhaus, M. Helm, in *Curr. Protoc. Nucleic Acid Chem.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2008**, p. 11.12.1–11.12.22.
- [22] S. Keyhani, T. Goldau, A. Blümmler, A. Heckel, H. Schwalbe, *Angew. Chemie Int. Ed.* **2018**, *57*, 12017–12021.
- [23] J. A. Barltrop, P. Schofield, *J. Chem. Soc.* **1965**, 4758.
- [24] J. H. Kaplan, B. Forbush III, J. F. Hoffman, *Biochemistry* **1978**, *17*, 1929–1935.
- [25] A. P. Pelliccioli, J. Wirz, *Photochem. Photobiol. Sci.* **2002**, *1*, 441–458.
- [26] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Angew. Chemie Int. Ed.* **2005**, *44*, 2600–2603.
- [27] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *ChemBioChem* **2006**, *7*, 417–420.
- [28] B. Fürtig, J. Buck, V. Manoharan, W. Bermel, A. Jäschke, P. Wenter, S. Pitsch, H. Schwalbe, *Biopolymers* **2007**, *86*, 360–383.
- [29] K. Snoussi, J.-L. Leroy, *Biochemistry* **2001**, *40*, 8898–8904

- [30] H. S. Steinert, J. Rinnenthal, H. Schwalbe, *Biophys. J.* **2012**, *102*, 2564–2574.
- [31] R. Schnieders, *Unpubl. data* **2019**.
- [32] E. N. Nikolova, H. M. Al-Hashimi, *RNA* **2010**, *16*, 1687–1691.
- [33] J. Ferner, M. Suhartono, S. Breitung, H. R. A. Jonker, M. Hennig, J. Wöhnert, M. Göbel, H. Schwalbe, *ChemBioChem* **2009**, *10*, 1490–1494.
- [34] M. Orita, R. Vinayak, A. Andrus, M. Warashina, A. Chiba, H. Kaniwa, F. Nishikawa, S. Nishikawa, K. Taira, *J. Biol. Chem.* **1996**, *271*, 9447–9454.
- [35] J. H. Cate, R. L. Hanna, J. A. Doudna, *Nat. Struct. Biol.* **1997**, *4*, 553–558.
- [36] A. Hampel, J. A. Cowan, *Chem. & Biol.* **1997**, *4*, 513–517.
- [37] S. E. Butcher, F. H.-T. Allain, J. Feigon, *Biochemistry* **2000**, *39*, 2174–2182.
- [38] E. Bonneau, P. Legault, *Biochemistry* **2014**, *53*, 579–590.
- [39] A. Kumar, G. Wagner, R. R. Ernst, K. Wüthrich, *J. Am. Chem. Soc.* **1981**, *103*, 3654–3658.
- [40] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Am. Chem. Soc.* **1994**, *116*, 6472–6473.
- [41] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Biomol. NMR* **1995**, *5*, 87–92.
- [42] V. Sklenář, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1993**, *3*, 721–727.
- [43] L. E. Kay, G. Y. Xu, A. U. Singer, D. R. Muhandiram, J. D. Formankay, *J. Magn. Reson. Ser. B* **1993**, *101*, 333–337.
- [44] A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
- [45] C. S. Thakur, J. N. Sama, M. E. Jackson, B. Chen, T. K. Dayie, *J. Biomol. NMR* **2010**, *48*, 179–192.
- [46] L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, J. Taylor, *J. Virol.* **1988**, *62*, 2674–2679.
- [47] S. P. Rosenstein, M. D. Been, *Nucleic Acids Res.* **1991**, *19*, 5409–5416.
- [48] A. T. Perrotta, M. D. Been, *Nucleic Acids Res.* **1990**, *18*, 6821–6827.
- [49] A. T. Perrotta, M. D. Been, *Nature* **1991**, *350*, 434–436.
- [50] M. Y. Kuo, L. Sharmeen, G. Dinter-gottlieb, J. Taylor, *Microbiology* **1988**, *62*, 4439–4444.
- [51] B.-S. Lee, H.-N. Wu, T.-H. Huang, *FEBS Lett.* **1993**, *324*, 296–300.
- [52] A. Ganguly, P. Thaplyal, E. Rosta, P. C. Bevilacqua, S. Hammes-Schiffer, *J. Am. Chem. Soc.* **2014**, *136*, 1483–1496.
- [53] S. I. Nakano, D. M. Chadalavada, P. C. Bevilacqua, *Science* **2000**, *287*, 1493–1497.
- [54] K. Barbara Schowen, R. L. Schowen, *Methods Enzymol.* **1982**, *87*, 551–606.
- [55] M. H. Kolk, H. A. Heus, C. W. Hilbers, *EMBO J.* **1997**, *16*, 3685–3692.
- [56] A. R. Ferré-D’Amaré, K. Zhou, J. A. Doudna, *Nature* **1998**, *395*, 567–574.
- [57] Y. Tanaka, T. Hori, M. Katahira, F. Nishikawa, T. Sakamoto, Y. Fukunaga, Y. Kurihara, S. Nishikawa, S. Uesugi, *Nucleic Acids Symp Ser* **1999**, *17*, 221–222.
- [58] Y. Tanaka, T. Hori, M. Tagaya, T. Sakamoto, Y. Kurihara, M. Katahira, S. Uesugi, *Nucleic Acids Res.* **2002**, *30*, 766–774.
- [59] N. K. Tanner, S. Schaff, G. Thill, E. Petit-Koskas, A.-M. Crain-Denoyelle, E. Westhof, *Curr. Biol.* **1994**, *4*, 488–498.
- [60] A. Lupták, A. R. Ferré-D’Amaré, K. Zhou, K. W. Zilm, J. A. Doudna, *J. Am. Chem. Soc.* **2001**, *123*, 8447–8452.
- [61] E. A. Schultes, D. P. Bartel, *Science* **2000**, *289*, 448–52.
- [62] B. Knezic, *Untersuchung Der Sequenzabhängigkeit Der RNA-Struktur*, **2016**.
- [63] S. Keyhani, *Unpubl. data* **2018**.
- [64] R. Naylor, P. T. Gilham, *Biochemistry* **1966**, *5*, 2722–2728.
- [65] W. S. Zielinski, L. E. Orgel, *Nature* **1987**, *327*, 346–347.
- [66] N. Paul, G. F. Joyce, *Proc. Natl. Acad. Sci.* **2002**, *99*, 12733–12740.
- [67] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229–1232.
- [68] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, G. Bruening, *Science* **1986**, *231*, 1577–1580.
- [69] J. B. Murray, D. P. Terwey, L. Maloney, A. Karpeisky, N. Usman, L. Beigelman, W. G. Scott, *Cell* **1998**, *92*, 665–673.
- [70] H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **1994**, *372*, 68–74.
- [71] W. G. Scott, J. T. Finch, A. Klug, *Cell* **1995**, *81*, 991–1002.
- [72] J.-P. Simorre, P. Legault, A. B. Hangar, P. Michiels, A. Pardi, *Biochemistry* **1997**, *36*, 518–525.
- [73] J.-P. Simorre, P. Legault, N. Baidya, O. C. Uhlenbeck, L. Maloney, F. Wincott, N. Usman, L. Beigelman, A. Pardi, *Biochemistry* **1998**, *37*, 4034–4044.
- [74] T. Sakamoto, G. Kawai, M. Katahira, M. H. Kim, Y. Tanaka, Y. Kurihara, T. Kohno, S. Watanabe, S. Yokoyama, K. Watanabe, *et al.*, *J. Biochem.* **1997**, *122*, 556–562.

- [75] R. P. Ojha, M. M. Dhingra, M. H. Sarma, Y. P. Myer, R. F. Setlik, M. Shibata, A. L. Kazim, R. L. Ornstein, R. Rein, C. J. Turner, *et al.*, *J. Biomol. Struct. Dyn.* **1997**, *15*, 185–215.
- [76] K. Bondensgaard, E. T. Mollova, A. Pardi, *Biochemistry* **2002**, *41*, 11532–11542.
- [77] D. Dufour, M. de la Peña, S. Gago, R. Flores, J. Gallego, *Nucleic Acids Res.* **2009**, *37*, 368–381.
- [78] A. Peracchi, L. Beigelman, E. C. Scott, O. C. Uhlenbeck, D. Herschlag, *J. Biol. Chem.* **1997**, *272*, 26822–26826.
- [79] M. Maderia, L. M. Hunsicker, V. J. DeRose, *Biochemistry* **2000**, *39*, 12113–12120.
- [80] M. Hansen, J. Simorre, P. Hanson, V. Mokler, L. Bellon, L. Beigelman, A. Pardi, *RNA* **1999**, *5*, 1099–1104.
- [81] K. Suzumura, Y. Takagi, M. Orita, K. Taira, *J. Am. Chem. Soc.* **2004**, *126*, 15504–15511.
- [82] C. Hammann, D. G. Norman, D. M. J. Lilley, *Proc. Natl. Acad. Sci.* **2001**, *98*, 5503–5508.
- [83] G. Wang, B. L. Gaffney, R. A. Jones, *J. Am. Chem. Soc.* **2004**, *126*, 8908–8909.
- [84] Y. Tanaka, C. Kojima, E. H. Morita, Y. Kasai, K. Yamasaki, A. Ono, M. Kainosho, K. Taira, *J. Am. Chem. Soc.* **2002**, *124*, 4595–4601.
- [85] B. Fürtig, C. Richter, P. Schell, P. Wenter, S. Pitsch, H. Schwalbe, *RNA Biol.* **2008**, *5*, 41–48.
- [86] A. Hampel, R. Tritz, *Biochemistry* **1989**, *28*, 4929–4933.
- [87] M. J. Fedor, *Biochemistry* **1999**, *38*, 11040–11050.
- [88] Z. Cai, I. Tinoco, *Biochemistry* **1996**, *35*, 6026–6036.
- [89] S. E. Butcher, F. H.-T. Allain, J. Feigon, *Nat. Struct. Biol.* **1999**, *6*, 212–216.
- [90] S. Ravindranathan, S. E. Butcher, J. Feigon, *Biochemistry* **2000**, *39*, 16026–16032.
- [91] A. J. Shallop, B. L. Gaffney, R. A. Jones, *Nucleosides, Nucleotides and Nucleic Acids* **2004**, *23*, 273–280.
- [92] M. Meli, J. Vergne, M.-C. Maurel, *J. Biol. Chem.* **2003**, *278*, 9835–9842.
- [93] J. Buck, Y.-L. Li, C. Richter, J. Vergne, M.-C. Maurel, H. Schwalbe, *ChemBioChem* **2009**, *10*, 2100–2110.
- [94] D. M. J. Lilley, *RNA* **2004**, *10*, 151–158.
- [95] N. B. Suslov, S. DasGupta, H. Huang, J. R. Fuller, D. M. J. Lilley, P. A. Rice, J. A. Piccirilli, *Nat. Chem. Biol.* **2015**, *11*, 840–846.
- [96] P. J. A. Michiels, C. H. J. Schouten, C. W. Hilbers, H. A. Heus, *RNA* **2000**, *6*, 1821–1832.
- [97] B. Hoffmann, G. T. Mitchell, P. Gendron, F. Major, A. A. Andersen, R. A. Collins, P. Legault, *Proc. Natl. Acad. Sci.* **2003**, *100*, 7003–7008.
- [98] D. O. Campbell, P. Bouchard, G. Desjardins, P. Legault, *Biochemistry* **2006**, *45*, 10591–10605.
- [99] P. Bouchard, J. Lacroix-Labonte, G. Desjardins, P. Lampron, V. Lisi, S. Lemieux, F. Major, P. Legault, *RNA* **2008**, *14*, 736–748.
- [100] G. Desjardins, E. Bonneau, N. Girard, J. Boisbouvier, P. Legault, *Nucleic Acids Res.* **2011**, *39*, 4427–4437.
- [101] P. Bouchard, P. Legault, *Biochemistry* **2014**, *53*, 258–269.
- [102] E. Bonneau, P. Legault, *Biochemistry* **2014**, *53*, 6264–6275.
- [103] E. Bonneau, N. Girard, S. Lemieux, P. Legault, *RNA* **2015**, *21*, 1621–1632.
- [104] T. Pan, O. C. Uhlenbeck, *Biochemistry* **1992**, *31*, 3887–3895.
- [105] T. Pan, O. C. Uhlenbeck, *Nature* **1992**, *358*, 560–563.
- [106] P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1994**, *116*, 8390–8391.
- [107] P. Legault, C. G. Hoogstraten, E. Metlitzky, A. Pardi, *J. Mol. Biol.* **1998**, *284*, 325–335.
- [108] M. Katahira, T. Sugiyama, M. Kanagawa, M. H. Kim, S. Uesugi, T. Kohno, *Nucleosides and Nucleotides* **1996**, *15*, 489–503.
- [109] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
- [110] S. M. Marquez, J. L. Chen, D. Evans, N. R. Pace, *Mol. Cell* **2006**, *24*, 445–56.
- [111] C. Glemarec, J. Kufel, A. Földesi, T. Maltseva, A. Sandström, L. A. Kirsebom, J. Chattopadhyaya, *Nucleic Acids Res.* **1996**, *24*, 2022–2035.
- [112] C. Spitzfaden, N. Nicholson, J. J. Jones, S. Guth, R. Lehr, C. D. Prescott, L. A. Hegg, D. S. Eggleston, *J. Mol. Biol.* **2000**, *295*, 105–115.
- [113] M. Schmitz, I. Tinoco Jr., *RNA* **2000**, *6*, 1212–1225.
- [114] T. Zuleeg, R. K. Hartmann, R. Kreutzer, S. Limmer, *J. Mol. Biol.* **2001**, *305*, 181–189.
- [115] T. C. Leeper, M. B. Martin, H. Kim, S. Cox, V. Semenchenko, F. J. Schmidt, S. R. Van Doren, *Nat. Struct. Biol.* **2002**, *9*, 397–403.
- [116] M. M. Getz, A. J. Andrews, C. A. Fierke, H. M. Al-Hashimi, *RNA* **2007**, *13*, 251–266.
- [117] C. H. Henkels, Y.-C. Chang, S. I. Chamberlin, T. G. Oas, *Biochemistry* **2007**, *46*, 15062–15075.
- [118] Y.-C. Chang, W. R. Franch, T. G. Oas, *Biochemistry* **2010**, *49*, 9428–9437.

- [119] K. S. Koutmou, A. Casiano-Negroni, M. M. Getz, S. Pazicni, A. J. Andrews, J. E. Penner-Hahn, H. M. Al-Hashimi, C. A. Fierke, *Proc. Natl. Acad. Sci.* **2010**, *107*, 2479–2484.
- [120] S. Cuzic, K. A. Heidemann, J. Wöhnert, R. K. Hartmann, *J. Mol. Biol.* **2008**, *379*, 1–8.
- [121] W.-Y. Chen, Y. Xu, I.-M. Cho, S. V. Oruganti, M. P. Foster, V. Gopalan, *J. Mol. Biol.* **2011**, *411*, 368–383.
- [122] B. Reinhold-Hurek, D. A. Shub, *Nature* **1992**, *357*, 173–176.
- [123] F. Michel, E. Westhof, *J. Mol. Biol.* **1990**, *216*, 585–610.
- [124] T. R. Cech, D. Herschlag, J. A. Piccirilli, A. M. Pyle, *J. Biol. Chem.* **1992**, *267*, 17479–17482.
- [125] F. H.-T. Allain, G. Varani, *J. Mol. Biol.* **1995**, *250*, 333–353.
- [126] J.-H. Lee, A. Pardi, *Nucleic Acids Res.* **2007**, *35*, 2965–2974.
- [127] M. Amano, *Nucleic Acids Symp. Ser.* **2001**, *1*, 169–170.
- [128] M. Amano, *Nucleic Acids Symp. Ser.* **2000**, *44*, 281–282.
- [129] M. Amano, *Nucleic Acids Symp. Ser.* **2012**, *3*, 173–174.
- [130] M. Chastain, I. Tinoco Jr., *Biochemistry* **1992**, *31*, 12733–12741.
- [131] M. Chastain, I. Tinoco Jr., *Biochemistry* **1993**, *32*, 14220–14228.
- [132] J. Nowakowski, I. Tinoco Jr., *Biochemistry* **1996**, *35*, 2577–2585.
- [133] E. C. Johnson, V. A. Feher, J. W. Peng, J. M. Moore, J. R. Williamson, *J. Am. Chem. Soc.* **2003**, *125*, 15724–15725.
- [134] J. A. McDowell, D. H. Turner, *Biochemistry* **1996**, *35*, 14077–14089.
- [135] K. J. Luebke, S. M. Landry, I. Tinoco Jr., *Biochemistry* **1997**, *36*, 10246–10255.
- [136] G. Colmenarejo, I. Tinoco Jr., *J. Mol. Biol.* **1999**, *290*, 119–135.
- [137] S. K. Silverman, M. Zheng, M. Wu, I. Tinoco, T. R. Cech, *RNA* **1999**, *5*, 1665–1674.
- [138] S. Rüdiger, I. Tinoco, *J. Mol. Biol.* **2000**, *295*, 1211–1223.
- [139] M. Zheng, M. Wu, I. Tinoco Jr., *Proc. Natl. Acad. Sci.* **2002**, *98*, 3695–3700.
- [140] Y. Xue, B. Gracia, D. Herschlag, R. Russell, H. M. Al-Hashimi, *Nat. Commun.* **2016**, *7*, ncomms11768.
- [141] E. Koculi, S. S. Cho, R. Desai, D. Thirumalai, S. A. Woodson, *Nucleic Acids Res.* **2012**, *40*, 8011–8020.
- [142] B. Gracia, Y. Xue, N. Bisaria, D. Herschlag, H. M. Al-Hashimi, R. Russell, *J. Mol. Biol.* **2016**, *428*, 3972–3985.
- [143] S. Watanabe, G. Kawai, Y. Muto, K. Watanabe, T. Inoue, S. Yokoyama, *Nucleic Acids Res.* **1996**, *24*, 1337–1344.
- [144] A. Kitamura, Y. Muto, S. Watanabe, I. Kim, T. Ito, Y. Nishiya, T. Ohtsuki, G. Kawai, K. Watanabe, K. Hosono, et al., *Nucleic Acids Symp. Ser.* **2012**, *42*, 191–192.
- [145] A. Kitamura, Y. Muto, S. Watanabe, I. Kim, T. Ito, Y. Nishiya, K. Sakamoto, T. Ohtsuki, G. Kawai, K. Watanabe, et al., *RNA* **2002**, *8*, 440–451.
- [146] K. Lehmann, U. Schmidt, *Crit. Rev. Biochem. Mol. Biol.* **2003**, *38*, 249–303.
- [147] L. Zhang, J. A. Doudna, *Science* **2002**, *295*, 2084–2088.
- [148] R. K. O. Sigel, D. G. Sashital, D. L. Abramovitz, A. G. Palmer III, S. E. Butcher, A. M. Pyle, *Nat. Struct. Mol. Biol.* **2004**, *11*, 187–192.
- [149] M. C. Erat, O. Zerbe, T. Fox, R. K. O. Sigel, *ChemBioChem* **2007**, *8*, 306–314.

15. German Summary – Zusammenfassung in deutscher Sprache

In dieser kumulativen Arbeit wurden strukturelle und funktionale Untersuchungen an Nucleinsäuren durchgeführt, hauptsächlich, aber nicht ausschließlich unter Verwendung von NMR-Spektroskopie (Kernspin Resonanzspektroskopie) als Analyseverfahren. Die untersuchten Biomoleküle umfassten kleinere und größere biologisch relevante RNAs sowie einen artifiziellen DNA G-Quadruplex. Hierbei konnten Ergebnisse im Bereich der Bestimmung der molekularen Struktur, der Aufklärung der biologischen Funktion und der Wirkstoffentwicklung gewonnen werden, die in sechs verschiedenen Publikationen dargelegt sind, an deren Erstellung der Autor maßgeblich oder hauptverantwortlich beteiligt war. Des Weiteren wird in einem mehrgliedrigen Einleitungssegment auf den Stand der aktuellen Forschung in den jeweiligen Teilgebieten eingegangen.

Nucleinsäuren waren im biologischen Kontext lange nahezu ausschließlich durch ihre Funktionen als dauerhafte (DNA) oder transiente (RNA) Informationsspeichernde Moleküle charakterisiert. Neuerlich wurden allerdings, vor allem im Bereich der RNA, eine Vielzahl regulativ interagierender Nucleinsäuren entdeckt, deren Funktionen weit über die vormals angedachten biologischen Rollen hinausgehen. Regulative Funktionen werden hierbei zum Beispiel über Sequenzen in nicht-kodierenden Bereichen der mRNA ausgeübt, zum anderen können kleine RNAs involviert sein. Zudem weisen Nucleinsäuren eine Vielzahl von Strukturen auf, die von einfachen Helices über komplexe Tertiärstruktur motive bis zu den völlig anders aufgebauten G-Quadruplexen reichen. Diese hohe strukturelle Heterogenität erlaubt eine Reihe von Funktionen sowohl in der Natur als auch im technischen Bereich oder der Medikamentenentwicklung. Diese Funktionen wurden in dieser Arbeit mittels NMR-Spektroskopie untersucht.

NMR-Spektroskopie ist eine vielseitige biophysikalische Methode die nicht nur die strukturelle, sondern auch die funktionale Charakterisierung von Nucleinsäuren, aber auch anderen Biomolekülen und kleinen Molekülen erlaubt. Insbesondere erlaubt die differentielle Resonanz jedes einzelnen NMR-aktiven Atoms die Beobachtung biologischer Prozesse in atomarer Auflösung. Mit Hilfe verhältnismäßig einfacher und schneller eindimensionaler NMR-Experimente lassen sich zum Beispiel bereits Bindungsverhalten zu Liganden oder strukturelle Veränderungen kleiner und mittelgroßer Nucleinsäuren beobachten. Diese eignen sich zudem auch zur Betrachtung dynamischer Prozesse. Weiterentwickelte mehrdimensionale Experimente erweitern außerdem die experimentellen Möglichkeiten durch die Erhöhung der spektralen Auflösung. In diesem Fall ist auch die Gewinnung struktureller Parameter, wie interatomarer Abstände und Winkel möglich, womit sich die dreidimensionale Struktur kleiner bis mittelgroßer Nucleinsäuren bestimmen lässt. Die Vielzahl der möglichen Experimente zur Aufklärung biologischer Prozesse und Eigenschaften eignet sich gut für die Verwendung an diversen Nucleinsäuren und ist daher die hauptsächlich verwendete Methode in dieser Arbeit.

Die Regulation der prokaryotischen Genexpression kann von Proteinen aber teilweise auch von RNA völlig ohne Proteininteraktion ausgeübt werden. Die wichtigste Rolle spielen hier sogenannte

Riboswitches (auch: Riboschalter). Diese Sequenzen liegen im 5'-untranslatierten Bereich der bakteriellen mRNA vor und erlauben Genregulation in Abhängigkeit von Zellmetaboliten. Riboswitches sind durch eine, den Liganden bindende, Aptamerdomäne und eine, die Regulation ausübende, Expressionsplattform charakterisiert. Bei Bindung des Liganden erfolgt eine strukturelle Umwandlung und daraus resultierend eine Expressionsmodulation. Die Ligandenbindung zeichnet sich üblicherweise durch eine hohe Selektivität aus, da Zellmetabolite oft ihren biochemischen Vorläufer- oder Nachfolgemolekülen sehr ähnlich sind. Die Regulation durch Riboswitches kann auf zwei Arten erfolgen. Transkriptionelle Kontrolle wird während der Transkription durch die RNA Polymerase ausgeübt. Hierbei bildet der Riboswitch bei An- oder Abwesenheit des Liganden (On- oder Off-Switch) eine Terminatorstruktur aus, die die weitere Transkription behindert. Im Falle von transkriptioneller Kontrolle wird dagegen die ribosomale Bindestelle je nach An- oder Abwesenheit des Liganden maskiert und so die Transkription verhindert oder ermöglicht. In einer der Publikationen aus dieser Arbeit wurde ein ZMP-bindender transkriptioneller On-Switch untersucht, der sich durch zwei miteinander interagierende Unterdomänen (P1 und P3) der Aptamerdomäne in einem sogenannten Pseudoknoten auszeichnet. Die Haarnadelstruktur P3 beinhaltet hierbei Reste, die auch zur Bildung einer Terminatorstruktur beitragen. Die Stabilisierung der P1/P3 Interaktion durch den Liganden behindert hierbei transient die Terminatorbildung und führt zur Transkription. Die Details der transkriptionellen Kontrolle sind durch die Umfaltungskinetik im Zusammenhang mit der Transkriptionsgeschwindigkeit und den unterschiedlichen Stabilitäten der Strukturen komplex. Die Studie nähert sich der Aufklärung dieser Mechaniken durch Untersuchung transkriptioneller Intermediate und ihrer Stabilitäten im Gleichgewichtszustand. Zunächst wurden mit Hilfe von 2D-NMR Methoden die hoch dispergierten Signale der Iminoprotonen zugeordnet und ein Reportersignal für die Interaktion von P1 und P3 identifiziert. Die Bildung des Pseudoknotens ohne Ligandenbindung, welche vor der Studie nicht vollständig evident war konnte mit Hilfe von einer teilweise isotopenmarkierten Probe der beiden Subdomänen direkt beobachtet werden. Zehn Sequenzen, die Intermediate der RNA während der Transkription darstellen wurden hergestellt und mittels 1D-NMR Titration und Isothermer Titrationskalorimetrie untersucht. Hierbei wurde ein sehr scharfer Übergang zwischen bindungskompetenten und nicht bindungskompetenten Transkriptionsintermediaten nachgewiesen (Bei einer Länge von 80 Nukleotiden, gegenüber 81 Nukleotiden). Zudem deutet die beobachtete sehr langsame Umfaltung in den bindungskompetenten Zustand auf eine kinetische Barriere hin, die in der Natur zu einem noch schärferen Übergang der Bindungskompetenz führen würde.

Im Bereich der eukaryotischen Genexpression erfolgt die Regulation hauptsächlich durch Proteine, dennoch spielen Nukleinsäuren hier häufig eine indirekte Rolle. Zum einen konnten schon auf dem DNA Templatstrang sogenannte Enhancer-Elements (Verstärkerelemente) nachgewiesen werden, die die Transkription regulieren. Zum anderen wirken RNAs im Bereich der mRNA Degradation, und damit der Verfügbarkeit zur Translation, modulierend. Sogenannte microRNAs und long-non-coding (lange nicht kodierende) RNAs interagieren mit Bereichen des 3'-untranslatierten Bereichs (3'-UTR) einer mRNA und rekrutieren Degradationsfaktoren. Bestimmte Sequenzen in der 3'-UTR interagieren zudem direkt mit Proteinen. Dies trifft zum

Beispiel auf AU-reiche Elemente (AREs) zu, die durch mindestens neun Nukleotide lange, meist aber zwischen 50 und 150 Nukleotide umfassende, AU-reiche und einzelsträngige Sequenzen charakterisiert sind. Diese werden von bestimmten ARE-bindenden Proteinen erkannt, die bei Bindung regulative Proteinfaktoren rekrutieren. Bekannte Beispiele sind die Proteine AUF1 (AU-rich element binding protein 1), Tristetraprolin und HuR (human antigen R). Eine andere Klasse von regulierenden Elementen stellen CDEs (constitutive decay elements) dar. Diese kurzen Sequenzen bilden Haarnadelstrukturen mit einer charakteristischen Schleife, in der die Basenabfolge Y-R-N (Y: Pyrimidin, R: Purin, N: jede Base) phylogenetisch konserviert ist. Diese werden von dem Protein Roquin erkannt, das einen Degradationskomplex rekrutiert. Die, für ein solches Element notwendige Basenabfolge, erlaubt dass ein CDE gleichzeitig auch ein ARE sein kann. Zwei entsprechende Sequenzen wurde in einem der wissenschaftlichen Artikel strukturell untersucht. Obwohl die Bindung der kurzen haarnadelbildenden RNAs an Roquin biologisch nachgewiesen war existierte vor Beginn der Studie kein struktureller Nachweis. Entsprechend wurden die dreidimensionalen Strukturen der RNAs mittels NMR und die ihrer Roquinkomplexe mittels Röntgenkristallographie bestimmt. Der Autor dieser Arbeit war für die NMR-Strukturbestimmung zuständig. Hierzu wurden zunächst alle Signale von Protonen sowie einiger Kohlenstoff- und Stickstoffatome zugeordnet und mittel 2D-NOESY NMR Spektroskopie die interatomaren Abstände der Protonen bestimmt. Es wurde isotopenmarkierte aber auch unmarkierte RNA verwendet. Zusätzliche Bestimmung interatomarer Winkel erfolgte durch moderne RNA-spezifische 2D- und 3D-NMR Experimente. Die erhaltenen Daten wurden in einer molekulardynamischen Rechnung verwendet, um die dreidimensionalen Strukturen der RNAs zu erhalten. Die Strukturen der RNAs zeigen eine überraschend hohe Dynamik der Schleifen, entsprechend kann hier von einer Einpassung in die Bindestelle des Proteins ausgegangen werden. Vergleiche beider Strukturen zeigen zudem unterschiedliche Schleifengeometrien bei relativ hoher Sequenzidentität und gleicher gebundener Struktur, was die beobachtete hohe Dynamik unterstreicht. Untersuchung der Bindung von AUF1 zeigen zudem kompetitive Bindung zu Roquin.

Im Gegensatz zu anderen RNA Elementen deren Strukturen auf der Bildung von Basenpaaren nach Watson & Crick basieren stellen G-Quadruplexe ein grundlegend andersartiges Strukturmotiv dar. Sie basieren auf der Bildung von Tetraden aus Guaninen die um monovalente Kationen herum gestapelt sind und kommen entsprechend nur in G-reichen Sequenzen vor. Die resultierenden Strukturen sind äußerst stabil und bilden sich aus DNA aber auch RNA, wobei DNA G-Quadruplexe eine im Detail deutlich größere Strukturheterogenität zeigen. In der Natur finden sich G-Quadruplexe zum Beispiel in Überhängen von Telomersträngen der DNA, welche überdurchschnittlich in Krebszellen vorkommen. Auch die an mehreren Stellen beobachtete regulative Aktivität G-reicher RNA Sequenzen wird der Bildung von G-Quadruplexen zugesprochen, wobei Studien gezeigt haben, dass G-Quadruplexe in lebenden Zellen nicht gefaltet vorliegen. Implikationen dieser Studie sind Gegenstand zahlreicher wissenschaftlicher Debatten über die Validität biologischer Studien, die regulierende G-Quadruplexe untersuchen, denn viele dieser biologischen Studien beobachten nur das Vorhandensein G-reicher Sequenzen. Ob diese Sequenzen in der Lage sind zumindest in vitro G-Quadruplexe zu falten wird in einem

der wissenschaftlichen Artikel mit biophysikalischen Methoden untersucht. Hierfür wurden sechs G-reiche Sequenzen aus einer hochrelevanten Studie, die diese Sequenzen mit der Wirkung eines Krebstherapeutikums in Verbindung bringt, biophysikalisch untersucht. Ein dreistufiges Protokoll, gegliedert in Zirkulardichroismus-, 1D-NMR- und 2D-NMR-Experimente wurde verwendet. Damit konnte gezeigt werden, dass von sechs untersuchten Sequenzen nur eine strikt einen G-Quadruplex bildete. Zwei der Sequenzen befanden sich im temperaturabhängigen Austausch mit einer Haarnadelstruktur, während drei Sequenzen nur unspezifische Aggregate bildeten.

G-Quadruplexe sind auch außerhalb der Natur bei der Konstruktion von Nanostrukturen von Interesse. Durch ihre Länge und hohe Stabilität stellt DNA ein beliebtes Bauteil für die Konstruktion komplexer dreidimensionaler Strukturen auf molekularer Ebene dar. G-Quadruplexe dienen aufgrund ihrer hohen Stabilität häufig als Verbindungsstück zwischen DNA-Strängen. Während viele dieser Experimente momentan noch der Grundlagenforschung zuzuordnen sind, zeigen sich schon heute Anwendungsbeispiele in der Medizin bei denen zum Beispiel Wirkstoffe in Nano-Behältern verabreicht werden, die dann über ein externes Signal freigesetzt werden. Man spricht in diesem Fall von Nanomaschinen. Bei dem externen Signal handelt es sich häufig um Zugabe eines bestimmten Stoffes, zum Beispiel eines weiteren DNA-Stranges. Eine weniger invasive Strategie könnte hier die Verwendung von Licht als externem Signal sein. Mit dieser Prämisse wurde ein lichtabhängig schaltbarer G-Quadruplex entwickelt, der in einem der wissenschaftlichen Artikel in dieser Arbeit untersucht wurde. Dieser basiert auf einem Azo-Schalter und insgesamt vier G-Resten. Es wurden zunächst drei chemisch leicht unterschiedliche Sequenzen auf Bildung eines homogenen G-Quadruplexes mit Hilfe von 1D-NMR untersucht und das vielversprechendste Molekül ausgewählt. Mit Hilfe umfangreicher 1D-NMR und Zirkulardichroismusuntersuchungen wurde die Reversibilität der Bildung des G-Quadruplexes unter Einfluss sichtbaren Lichts und der Strukturauflösung unter UV-Licht nachgewiesen. Zudem wurde ähnlich wie im Falle der CDE RNAs die dreidimensionale Struktur des G-Quadruplex bestimmt, wobei auf die Verwendung isopenmarkierter Proben verzichtet werden konnte. Aufgrund der hohen Symmetrie der Struktur war zunächst nicht bestimmbar, ob sich eine bimolekulare oder tetramolekulare Struktur bildete. Dies konnte mit Hilfe spezifischer 2D-NOESY Signale, sowie DOSY-NMR und vergleichender Strukturrechnung aufgeklärt werden. Es handelte sich um einen tetramolekularen G-Quadruplex.

Neuerlich werden Nukleinsäuren in wachsender Anzahl in Studien zur Medikamentenentwicklung als Zielmoleküle eingesetzt. Obwohl in den meisten Fällen Proteine in diesen Screeningstudien eingesetzt werden können, gelten in anderen Fällen die gewünschten Zielproteine als untauglich für die Medikamentenentwicklung („undruggable“). In diesem Fall ist eine Lösungsmöglichkeit die entsprechende mRNA als Zielmolekül einzusetzen. Auch in der Entwicklung neuer Antibiotika, seit langem ein problematischer Sektor im Pharmabereich sind RNA-Screening Projekte ein sinnvoller Ansatz, da in Bakterien besonders viel regulatorisch aktive RNA bekannt ist. In einer wissenschaftlichen Publikation wurde daher ein exemplarisches Screening von einer Stoffbibliothek aus 102 Molekülen mit 14 verschiedenen RNAs durchgeführt. Der Fokus lag hier

besonders auf Riboswitches. Zusätzlich wurden fünf Proteine und fünf DNAs als Zielmoleküle eingesetzt. Es handelte sich hierbei um ein fragmentbasiertes Screening, entsprechend waren die Moleküle in der Datenbank besonders klein und ihre Auswahl so konzipiert, dass aus den Fragmenten leicht Medikamentenvorläufer entwickelt werden können. Zudem enthielten alle Fragmente mindestens ein Fluoratom, was die Verwendung von ^{19}F -NMR als Analysemethode erlaubte. Diese birgt den Vorteil gegenüber ^1H -NMR, dass deutlich weniger Signale im Spektrum zu sehen sind, die außerdem deutlich stärker dispergiert sind. Zur Visualisierung der Bindung wurden T_2 -modulierte ^{19}F -1D-NMR Experimente eingesetzt, die auch die Beobachtung von schwacher Bindung ermöglichen, ein Vorteil im Kontext des fragmentbasierten Screenings. Für nahezu alle Biomoleküle konnten bindende Fragmente, sogenannte Hits, beobachtet werden. An einigen Riboswitch/Hit Kombinationen wurden weiterführende Experimente durchgeführt, die 2D-NMR Untersuchungen, kompetitive Screenings gegen den nativen Liganden und Bestimmung der Affinität mittels ^{19}F -NMR Titration umfassten. Des Weiteren wurde beispielhaft die Entwicklung eines potentiellen Arzneimittelvorläufers durch chemische Verknüpfung eines Fragments mit dem bekannten RNA-bindenden Molekül Acridin demonstriert.

In einer weiteren Publikation, dem Unterkapitel eines Buches über Ribozyme, wird die Verwendung von NMR-Spektroskopie bei der Untersuchung von Ribozymen diskutiert. In diesem Kapitel werden zunächst Herstellungsmethoden für RNA und NMR-Methoden zu deren Untersuchung beschrieben, unter anderem 1D-, 2D- und 3D-Methoden zur Zuordnung, Analyse und Strukturaufklärung. Das Kapitel beschreibt daraufhin den Stand der wissenschaftlichen Entwicklung, betreffend der Untersuchung von Ribozymen mittels NMR-Spektroskopie, sortiert nach einzelnen Ribozymen. Hierbei wurde besonders Augenmerk auf die Betrachtung vieler verschiedener Publikationen gelegt und außerdem das wichtige Hepatitis delta Virus (HDV) Ribozym besonders hervorgehoben.

16. Danksagung

Zunächst möchte ich Prof. Dr. Harald Schwalbe für die Möglichkeit danken, diese Arbeit in seiner Gruppe anfertigen zu können. Ich hatte schon während des Studiums immer großes Interesse an NMR Spektroskopie und die Aufnahme in die Gruppe bedeutete mir sehr viel. Ich danke dir, Harald, außerdem für die vielen interessanten Projekte, die Unterstützung bei allen wissenschaftlichen Fragen und die immer freundliche und offene Betreuung.

Ich danke dem tollen Team aus Postdocs und jungen Arbeitsgruppenleitern, die in Meetings aber auch außerhalb jederzeit ein offenes Ohr für Fragen und Probleme hatten. Insbesondere danke ich hier Dr. Anna Wacker für die vielen Hilfen, Diskussionen und Denkanstöße im Bereich der RNA-Forschung und Dr. Christian Richter, von dem ich viel über NMR-Spektroskopie gelernt habe. Außerdem danke ich Dr. Henry Jonker für die Hilfestellung bei allen Fragen zur Strukturrechnung.

Ein herzliches Danke spreche ich Dr. Irene Bessi aus, die mich während meiner Masterarbeit betreute und mich mit ihrem Wissen und ihrer aufgeschlossenen Art immer beeindruckt hat. Ich hätte mir niemand besseren für den Start in der Gruppe vorstellen können.

Außerdem möchte ich mich bei allen, am organisatorischen Beteiligten, in der Schwalbe Gruppe bedanken. Insbesondere möchte ich hier Kerstin Dathe erwähnen, die das Sekretariat gerockt hat (Busserl). Vielen Dank dass du es so schnell geregelt hast (gilt eigentlich für alles) und die spaßigen Gespräche. Auch Elke Stinal möchte ich für ihre Arbeit an der HPLC und die Organisation des Labors danken.

Ich danke meinen Kooperationspartnern: Dr. Julie Thevarpadam und Prof. Alexander Heckel für die Zusammenarbeit beim DNA G4 Projekt, Dr. Jan-Niklas Tants, Johannes Braun, Stephen Peter, Dr. Julia Weigand und Dr. Andreas Schlundt für die Zusammenarbeit beim CDE Projekt und natürlich auch allen internen Kollegen und Mitarbeitern mit denen ich die Möglichkeit hatte zusammen zu arbeiten. Prof. Alexander Heckel möchte ich außerdem für die Übernahme des Zweitgutachtens danken.

Ein herzliches Danke an die Kollegen aus dem „Beyond Horizon Room“, besonders Albrecht Völklein, Gül Aydin, Lars-Erik Bendel, Tom Landgraf und Diana Müller. War eine super Zeit mit euch, danke für alle lustigen, lehrreiche und spannenden Gespräche und den regen wissenschaftlichen Austausch. An dieser Stelle nochmal ein Danke an Albrecht und Tom, für die gute Zeit beim Essen, auf Konferenzen, im Büro, aber auch für die Zusammenarbeit in den verschiedenen Projekten.

Vielen Dank natürlich auch an den „Boysroom“, also Erhan Cetiner, Fabian Hiller, Daniel Hymon, Sara Keyhani, Bozana Knezic, Laurie Lannes, Martin Hähnke, Marie Hutchison, Robbin Schnieders und Alix Tröster für wissenschaftliche und unwissenschaftliche Diskussionen, den Spaß, die verrückten Witze und alles was ich von euch gelernt habe. Besonders danke ich Martin für den IT Support, das Rechner Upgrade und generell alles was du für die Gruppe tust.

Außerdem danke ich besonders Robbin für das Korrekturlesen des NMR-Teils und die vielen NMR Diskussionen und Beratungen.

Vielen Dank an meine Praktikanten, Bachelor- und Masterstudenten die an einigen Stellen zu dieser Arbeit beigetragen haben, vor allem Tatjana für das Verlängerungsessay.

Schlussendlich danke ich meiner Familie, die mich immer unterstützt hat und die immer für mich da war.

17. Lebenslauf

Zur Person

Name: Oliver Binas
Geburtsdatum: 17.07.1987
Geburtsort: Frankfurt am Main
Staatsangehörigkeit: Deutsch

Ausbildung

11/2015 – Doktorarbeit in Chemie
Goethe Universität Frankfurt am Main
Titel der Abschlussarbeit: Structural and functional investigation of nucleic acids with NMR spectroscopy

09/2012 – 04/2015 Master of Science in Chemie (Abschlussnote: 1.4)
Goethe Universität Frankfurt am Main
Titel der Abschlussarbeit: NMR Untersuchungen an G-Quadruplex bildenden Oligonukleotiden

09/2006 – 08/2012 Bachelor of Science in Chemie (Abschlussnote 2.3)
Goethe Universität Frankfurt am Main
Titel der Abschlussarbeit: Spektroskopische Messungen an Flavinen in unterschiedlichen Redoxzuständen

06/2006 Abitur
Wöhler Gymnasium Frankfurt am Main

Publikationen

Anna Wacker, Julia Weigand, ..., **Oliver Binas**, ..., Harald Schwalbe**, Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy and DMS footprinting analysis, 2020, *Nucleic Acids Research*, *Manuskript in peer-review*

Bozana Knezic*, **Oliver Binas***, Albrecht Völklein, Harald Schwalbe, „NMR on Ribozymes“, 2020, *Buchkapitel in „Ribozymes“, Wiley-VCH*

Oliver Binas*, Vanessa de Jesus*, Tom Landgraf*, Albrecht Eduard Völklein*, Jason Martins, Daniel Hymon, Jasleen K. Bains, Hannes Berg, Thomas Biedenbänder, Boris Fürtig, Santosh L.

Gande, Anna Niesteruk, Andreas Oxenfarth, Nusrat S. Qureshi, Tatjana Schamber, Robbin Schnieders, Alix Tröster, Anna Wacker, Julia Wirmer-Bartoschek, Maria Alexandra Wirtz Martin, Elke Stinal, Kamal Azzaoui, Christian Richter, Sridhar Sreeramulu, Marcel J. J. Blommers, Harald Schwalbe, „¹⁹F-NMR-based fragment screening for 14 different biologically active RNAs and 10 DNA and protein counter-screens“, 2020, *ChemBioChem*, *Manuskript akzeptiert*

Oliver Binas, Tatjana Schamber, Harald Schwalbe, “The conformational landscape of transcription intermediates involved in the regulation of the ZMP-sensing riboswitch from *Thermosinus carboxydivorans*”, 2020, *Nucleic Acids Research*, 48, 6970 – 6979

Oliver Binas*, Jan-Niklas Tants*, Stephen A Peter, Robert Janowski, Elena Davydova, Johannes Braun, Dierk Niessing, Harald Schwalbe, Julia E Weigand, Andreas Schlundt, “Structural basis for the recognition of transiently structured AU-rich elements by Roquin”, 2020, *Nucleic Acids Research*, 48, 7385 – 7403

Oliver Binas, Irene Bessi, Prof. Dr. Harald Schwalbe, “Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome”, 2020, *ChemBioChem*, 21, 1656 – 1663

Julie Thevarpadam*, Irene Bessi*, **Oliver Binas***, Dr. Diana P. N. Gonçalves, Dr. Chavdar Slavov, Dr. Hendrik R. A. Jonker, Dr. Christian Richter, Prof. Dr. Josef Wachtveitl, Prof. Dr. Harald Schwalbe, Prof. Dr. Alexander Heckel, „Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif“, 2016, *Angewandte Chemie International Edition*, 55, 2738 – 2742

*Geteilte Erstautorenschaft

**gekürzte Autorenliste, enthält 62 Autoren

Präsentationen

Oliver Binas, Johannes Braun, Stephen Peter, Jan-Niklas Tants, Andreas Schlundt, Julia E. Weigand, Harald Schwalbe, *Structural investigation of the UCP3 regulatory element from human 3'-UTR by NMR-spectroscopy*, Posterpräsentation, EUROISMAR, Berlin 2019

Oliver Binas, Johannes Braun, Stephen Peter, Jan-Niklas Tants, Andreas Schlundt, Julia E. Weigand, Harald Schwalbe, *Structural investigation of the UCP3 regulatory element from human 3'-UTR by NMR-spectroscopy*, Posterpräsentation und Young Scientist Poster Presentation, 60th Experimental Nuclear Magnetic Resonance Conference, Asilomar Conference Center, Pacific Grove, CA 2019

Oliver Binas, Johannes Braun, Julia Weigand, Andreas Schlundt, Michael Sattler, Harald Schwalbe, *Structural investigation of the ARE2 hairpin motif*, Posterpräsentation, GDCh FGMR 40th Annual Meeting, Leipzig 2018

Christian Richter, Sridhar Sreeramulu, Oliver Binas, Tom Landgraf, Albrecht Völklein, Tatjana Schamber, Julia Wirmer-Bartoschek, Robbin Schnieders, Thomas Biedenbänder,

Vanessa de Jesus, Boris Fürtig, Anna Wacker, Kamal Azzaoui, Marcel Blommers, Harald Schwalbe, *RNA screening using 19F-NMR*, Posterpräsentation, Structural biology approaches for drug development (iNext), Budapest 2018

Oliver Binas, Harald Schwalbe, *NMR spectroscopic investigation of the pfl riboswitch from thermosinus carboxydivorans* und **Oliver Binas** , Johannes Braun, Julia Weigand, Harald Schwalbe, *Structural investigation of the ARE2 hairpin motif*, Posterpräsentation, GDCh FGMR 39th Annual Meeting/ISMAR joint meeting, Bayreuth 2017

Oliver Binas, Julie Thevarpadam, Irene Bessi, Diana P. N. Gonçalves-Schmidt, Hendrik R. A. Jonker, Christian Richter, Harald Schwalbe and Alexander Heckel, *Photoresponsive formation of an inter-molecular minimal G-Quadruplex motif*, Posterpräsentation, GDCh FGMR 38th Annual Meeting, Düsseldorf 2016

Albrecht Völklein, **Oliver Binas**, **Tom Landgraf**, Harald Schwalbe, *NMR-spectroscopic investigation of the molecular mechanisms of secondary messenger-sensing riboswitches*, Posterpräsentation, Sensory and regulatory RNAs in Prokaryotes, München 2016

Oliver Binas, Julie Thevarpadam, Irene Bessi, Diana P. N. Gonçalves-Schmidt, Hendrik R. A. Jonker, Christian Richter, Harald Schwalbe and Alexander Heckel, *Photoresponsive formation of an inter-molecular minimal G-Quadruplex motif*, Posterpräsentation, GDCh FGMR 37th Annual Meeting, Darmstadt 2015

Akademischen Lehrer

Prof. Dr. Wolf Aßmus, Dr. Markus Braun, Prof. Dr. Bernd Brutschy, Prof. Dr. Andreas Dreuw, Prof. Dr. Ernst Egert, Prof. Dr. Joachim Engels, Prof. Dr. Michael Göbel, Prof. Dr. Alexander Heckel, Dr. Rainer Hegger, Prof. Dr. Max Holthausen, Prof. Dr. Bernd Kolbesen, Prof. Dr. Georg Manolikakes Prof. Dr. Thomas Prisner, Prof. Dr. Martin Schmidt, Prof. Dr. Gerhard Stock, Prof. Dr. Harald Schwalbe, Prof. Dr. Andreas Terfort, Prof. Dr. Matthias Wagner, Prof. Dr. Josef Wachtveitl