Structure of an RNA aptamer in complex with the fluorophore tetramethylrhodamine

Elke Duchardt-Ferner[1], Michael Juen[2], Benjamin Bourgeois[3], Tobias Madl [3,4], Christoph Kreutz [2], Oliver Ohlenschläger[5] and Jens Wöhnert[1,*]

[1]Institute for Molecular Biosciences, Goethe University, Frankfurt/M., Max-von-Laue-Str. 9, D-60438 Frankfurt, Germany and Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University, Frankfurt/M., Max-von-Laue-Str. 9, D-60438 Frankfurt, Germany, [2]Institute of Organic Chemistry, Centre for Molecular Biosciences (CMBI), University of Innsbruck, Innrain 80/82, 6020 Innsbruck, Austria, [3]Gottfried Schatz Research Center for Cell Signaling, Metabolism and Aging, Institute of Molecular Biology & Biochemistry, Medical University of Graz, Neue Stiftingtalstrasse 6, 8010 Graz, Austria., [4]BioTechMed-Graz, Graz, Austria and [5]Leibniz-Institute on Aging - Fritz-Lipmann-Institute, Beutenbergstrasse 11, D-07745 Jena, Germany

ABSTRACT

RNA aptamers—artificially created RNAs with high affinity and selectivity for their target ligand generated from random sequence pools—are versatile tools in the fields of biotechnology and medicine. On a more fundamental level, they also further our general understanding of RNA-ligand interactions e. g. in regard to the relationship between structural complexity and ligand affinity and specificity, RNA structure and RNA folding. Detailed structural knowledge on a wide range of aptamer–ligand complexes is required to further our understanding of RNA–ligand interactions. Here, we present the atomic resolution structure of an RNA–aptamer binding to the fluorescent xanthene dye tetramethylrhodamine. The high resolution structure, solved by NMR-spectroscopy in solution, reveals binding features both common and different from the binding mode of other aptamers with affinity for ligands carrying planar aromatic ring systems such as the malachite green aptamer which binds to the tetramethylrhodamine related dye malachite green or the flavin mononucleotide aptamer.

INTRODUCTION

RNA aptamers are artificially created RNA sequences able to bind to a wide variety of target ligand molecules with high affinity and specificity. At a time when only a few RNA structures were known and the idea that RNA could bind specifically to other molecules with widely different chemical characteristics was far from being generally accepted, these synthetic RNA molecules demonstrated the functional and structural potential of RNA. Aptamers are generated in a process first described almost 30 years ago which was termed SELEX (Systematic Evolution of Ligands by Exponential enrichment) (1,2). In this procedure, large pools of random sequences are subjected to several rounds of affinity selection and subsequent amplification. Already the early examples for RNA aptamers showed that RNAs are capable of distinguishing even chemically and structurally closely related ligands and are able to bind their targets with a specificity rivalling those of antibodies. The structures of aptamer–ligand complexes demonstrated for the first time that even small RNAs can fold into highly complex 3D structures. This is illustrated for instance by the structures of the ATP- (3–5), FMN- (6,7) and malachite green aptamer (8,9) bound to their respective ligands. These structures also provided a rationale for the impressive ligand specificity of these artificial molecules. Probably one of the most intriguing examples for ligand discrimination by an RNA is found in the theophylline binding aptamer (10,11). This aptamer is able to discriminate against caffeine, a compound that differs from theophylline only by the presence of an additional methyl group, by ~10 000-fold because the nitrogen-bound extra methyl group in caffeine sterically prevents the formation of an intermolecular hydrogen bond with the RNA (11). Structural and functional investigations of an aptamer for the amino acid citrulline also showed how easy it can be for RNA molecules to gain new functions. Here, three point mutations of nucleotides directly involved in ligand binding were sufficient to create an aptamer with a new specificity for the amino acid arginine (12,13). Similar observations were made for the ATP aptamer, which was evolved into a GTP binding aptamer by only a few point mutations (14). Furthermore, extensive selection experiments for a given ligand revealed that there are often multiple solutions for the specific recognition of this ligand differing widely in sequence and structural fea-

tures. An especially impressive example in this regard is the work of the Szostak group on GTP-binding aptamers (15–17). They identified as many as 11 sequentially and structurally distinct aptamer classes all binding to GTP with a wide range of affinities, different sets of recognition elements and different degrees of structural complexity. Only three classes of these aptamers have been structurally characterized so far (18–20). However, each of these structures revealed unexpected motifs for ligand recognition such as an intermolecular G-quadruplex (19) or the presence of a base quartet containing a highly stable protonated adenine (20) that extended our insights into RNAs structural capabilities. Finally, structural studies on a number of different RNA aptamers showed that ligand binding leads to large structural and dynamical changes in the aptamer RNA. In most cases, the ligand free state of the aptamer was found to be highly dynamic with ligand binding inducing a transition to a well-ordered structure (21).

The properties of the synthetic RNA aptamers described above inspired the search for naturally occurring counterparts where these properties could be potentially harnessed for gene regulatory purposes. These naturally occurring counterparts where found in the early 2000s in the form of the so-called riboswitches which are regulatory elements in the 5′-UTRs of bacterial mRNAs. In analogy to aptamers, riboswitches bind a wide variety of naturally occurring small metabolite molecules with high affinity and specificity (22) and adopt highly intricate three-dimensional structures upon ligand binding (23). Consequently, the ligand binding domains of riboswitches were termed aptamer domains. The often extensive ligand-induced structural and dynamical changes in the riboswitch aptamer domains are exploited to couple ligand binding to changes in gene expression by influencing either transcription or translation of the respective mRNAs. In analogy to previous observations on artificial aptamer families, it has been observed that for certain ligands such as SAM (*S*-adenosyl methionine), the modified nucleobase preQ$_1$ or guanidine multiple structurally and sequentially different riboswitch classes exist (24–26). These different riboswitch classes often vary widely in their ligand binding mode and specificity as well as in their structural complexity. Furthermore, in some cases riboswitch classes sensing structurally similar ligands differ only by a small number of mutations (27–29) in their aptamer domains resembling the earlier observations for e.g. the arginine/citrulline aptamer pair (12,13) and suggesting possible pathways for the evolution of novel gene regulatory elements.

The discovery of natural riboswitches in turn has inspired efforts to develop artificial aptamers for biorthogonal ligands into synthetic riboswitches (30) for the design of artificial gene regulatory circuits in synthetic biology applications. This approach was successful, e.g. for the theophylline aptamer, the tetracycline aptamer, the malachite green aptamer and the neomycin aptamer (30–34). However, no general rules for the design of artificial riboswitches from known aptamers could be established so far. A particularly illuminating example in this regard was the development of the neomycin riboswitch. The pool of neomycin aptamers obtained after only an *in vitro* SELEX is dominated by an RNA motif that binds with high affinity and specificity to neomycin but is not active as a riboswitch (35). Only when the complete pool of aptamers was subjected to an additional *in vivo* screening an aptamer with riboswitch properties emerged (33). The *in vitro* and *in vivo* selected aptamers differ strongly in their structural and dynamic properties and are therefore particularly helpful in order to define determinants for riboswitch activity. Interestingly, the artificial neomycin riboswitch echoes the exceptional ligand discrimination capabilities of the theophylline aptamer since it is triggered by neomycin but not by the closely related paromomycin which differs from neomycin only by the exchange of a single amino group with a hydroxyl group. High resolution structures of the neomycin riboswitch in complex with both aminoglycosides show that the replacement of a single functional group results in the loss of several intramolecular hydrogen bonding and electrostatic interactions followed by a general dynamic destabilization of the non-cognate ligand complex (32). In general, the development of synthetic riboswitches requires aptamer/ligand pairs where the aptamer is structurally rather unstable in its apo form and binds its target with a K_D in the nanomolar range in the presence of limited Mg^{2+}-concentrations and a bio-orthogonal ligand that is not toxic and is either easily taken up by the cell or able to penetrate cell membranes.

In order to extend the number of options of aptamer/ligand pairs for the development of potential RNA-based gene regulatory elements that fulfill these conditions Carothers *et al.* used SELEX in the presence of low Mg^{2+} concentrations to create a series of four aptamers that bind to the xanthene dye tetramethylrhodamine (TMR) (36). TMR (Figure 1A) is a cell permeable red-coloured dye with an absorption maximum at 535 nm which emits fluorescent light in the green-yellow region of the spectrum with an emission maximum at 573 nm. The aptamer TMR3 from the series of TMR-aptamers is reported to be the structurally most complex and information-rich aptamer (36) containing three predicted helical regions (P1–3) connected by three formally single-stranded junctions (Figure 1B). Sixteen of the 20 nucleotides in these predicted junction regions were found to be completely invariant. Three nucleotides vary between two bases and only one nucleotide is completely variable. The TMR3 aptamer binds its ligand TMR with a K_D in the high nM range and efficiently quenches its fluorescence emission (36).

In order to gain insights into the structural basis for the specific molecular recognition of a rigid fluorophore ligand by an RNA we determined the 3D structure of the TMR3 aptamer/TMR complex in solution using high-resolution multidimensional NMR-spectroscopy and characterized the structural changes in the aptamer between the apo and the ligand bound state. The structure of the TMR3/ligand complex reveals that ligand binding induces the formation of a stable three-way junction stabilized by Watson-Crick as well as by non-Watson-Crick pyrimidine-pyrimidine and purine-purine mismatch base pairs. The ligand stacks between a base triple closing helix P1 and a platform formed by consecutive G:A and G:G mismatches extending helix P3 and forms specific hydrogen bonds and polar interactions in the binding pocket. Some features of the TMR binding pocket in our structure are reminiscent of the binding sites for planar aromatic ligands observed in other
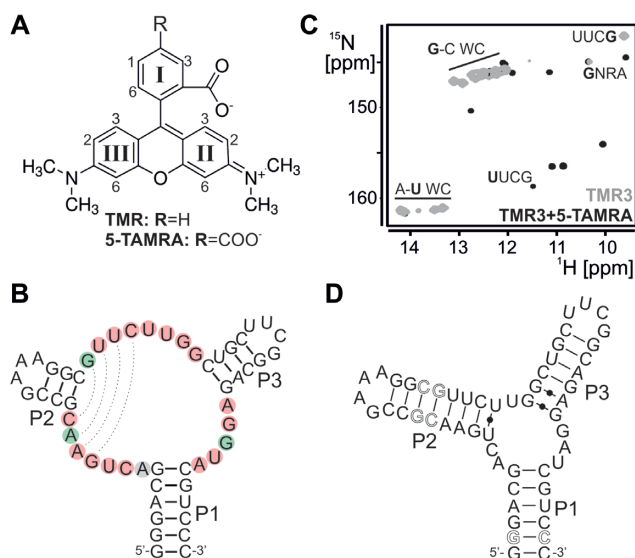
Figure 1. The tetramethylrhodamine binding aptamer 3 (TMR3). (**A**) Chemical structure of tetramethylrhodamine (TMR) and 5-carboxy-tetramethylrhodamine (5-TAMRA). Atom numbering for the sites carrying hydrogen atoms is given. (**B**) Sequence and secondary structure of TMR3 as proposed by Carothers and co-workers (36). The conservation pattern of residues in the predicted open internal bulge reported previously (36) is given. Completely conserved residues are shown in red, those which can be either one of two nucleotides in green. Residues, which are not conserved, are colored grey. The additional Watson–Crick base pairs identified in the initial NMR-spectra of free TMR3 RNA are indicated by dashed lines. (**C**) Imino group ^{1}H,^{15}N HSQC spectra of free TMR3 wt (gray) and TMR3 wt in complex with 5-TAMRA (black). The spectral regions characteristic for WC A–U and G–C base pairs are indicated. Additional signals characteristic for the resonances in the canonical UUCG and GNRA tetraloop are assigned with the residue giving rise to the NMR signal highlighted in bold. (**D**) Secondary structure of TMR3 RNA bound to 5-TAMRA using Leontis–Westhof notation (62). The three G–C base pairs deleted in the minimized construct TMR3_48nt are shown in open letters.

aptamer/ligand complexes. Interestingly, while ligand binding to the wild-type TMR3 aptamer leads to fluorescence quenching, structure-based mutations in the ligand binding pocket significantly reduce the quenching of the TMR fluorescence without affecting ligand affinity. Furthermore, the extensive structural changes induced by ligand binding suggest that TMR3 might indeed by a good candidate for the development of artificial riboswitches with fluorescent properties.

MATERIALS AND METHODS

Sample preparation

TMR3 RNAs were prepared by *in vitro* transcription with T7 RNA polymerase using linearized plasmid DNA as templates in general as described (37). The templates contained a 3′ self-cleaving hepatitis delta virus ribozyme sequence in order to ensure the presence of a homogenous 3′-terminus (38). An unlabeled and a uniformly ^{15}N-labeled sample of the 54 nt TMR3 wt RNA (5′-GGGACGACUGAACGCCGAAAGGCG UUCUUGGCUGCUUCGGCAGAGGUACGUCCC-3′) was prepared. Of the minimized TMR3_48nt RNA

(5′-GGACGACUGAACCGAAAGGUUCUUGGCU GCUUCGGCAGAGGUACGUCC-3′) an unlabeled, a uniformly ^{15}N-labeled ([U-^{15}N]) and an uniformly ^{13}C,^{15}N-labeled ([^{13}C,^{15}N]) sample and samples selectively ^{13}C,^{15}N-labeled in the guanosine and uridine nucleotides ([^{13}C,^{15}N]-Gua, Ura), in the adenosine and cytosine nucleotides ([^{13}C,^{15}N]-Ade, Cyt) and in the adenosine nucleotides alone ([^{13}C,^{15}N]-Ade) were produced by *in vitro* transcription for the structure determination of the TMR3_48nt RNA/ligand complex. Unlabeled nucleoside triphosphates were purchased from Sigma, [^{15}N]- and [^{13}C,^{15}N]-labeled nucleoside triphosphates were obtained from Silantes. All samples had a final concentration of 1.2 mM RNA in 130 mM glutamate buffer, pH 7.5 and 15 mM NaCl (NMR buffer) and contained 5% D_2O and 95% H_2O. This buffer corresponds to the buffer conditions used in the original SELEX-experiments (36). They were titrated with a slight excess of 5-TAMRA (1.4 mM), which was purchased from Sigma-Aldrich. In addition, unlabeled TMR3 wt and the TMR3_48nt RNA as well as the TMR3_48nt mutants ΔU24, ΔA43 and ΔU8-U24 (ΔP2) were produced for fluorescence titrations and 1D ^{1}H NMR experiments. Unlabeled TMR3_48nt RNA was also used in the SAXS measurements.

SAXS measurements

SAXS data for solutions of free TMR3_48nt RNA and in presence of 1.2 stoichiometric equivalents of 5-TAMRA, respectively, were recorded in NMR buffer on a laboratory SAXS instrument (SAXSpace, Anton Paar, Graz, Austria) equipped with a Kratky camera, a sealed X-ray tube source and a 2D Princeton Instruments PI•SCX:4300 (Roper Scientific) CCD detector. The scattering patterns were measured with 90-min exposure time (540 frames, each 10 s) at a solute concentration of 3.0 mg/ml. Radiation damage was excluded based on a comparison of individual frames of the 90-min exposures, where no changes were detected. A range of momentum transfer of $0.012 < s < 0.63$ Å^{-1} was covered ($s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 1.542$ Å is the X-ray wavelength). All SAXS data were analyzed with ATSAS (version 2.5). The data were processed with the SAXSQuant software (version 3.9), and desmeared using the programs GNOM and GIFT (39,40). The forward scattering, $I(0)$, the radius of gyration, R_g, the maximum dimension, D_{max} and the inter-atomic distance distribution functions ($P(R)$) were computed with the programs GNOM and GIFT. The masses of the solutes were evaluated by comparison of the forward scattering intensity using Porods law. For *ab initio* model generation, a total of 50 models were calculated using the program DAMMIF (41) and aligned and averaged using the program DAMCLUST. The generated models were aligned with the NMR structure of the TMR3_48nt/5-TAMRA complex.

Fluorescence quenching measurements

Fluorescence quenching titrations were carried out in triplicates at 25°C using a Fluorolog 3 fluorescence spectrometer (Horiba Yobin Ivon). 20 nM 5-TAMRA in NMR buffer was titrated with the different TMR3 RNA variants up to

RNA concentrations of 3–6 μM in a 2 ml cuvette. The samples were excited at 548 nm and emission spectra were collected between 550 and 650 nm. The normalized intensity at the excitation maximum at 575 nm was used to follow RNA binding. The resulting curves were fitted to a one-site binding event according to equation (1).

$$\frac{F_{[RNA]}}{F_0} = 1 - \frac{\left(1 - \frac{F_{min}}{F_0}\right)[RNA]}{K_D + [RNA]} \quad (1)$$

In equation (1) F_0, F_{min} and $F_{[RNA]}$ are the emission intensities of 5-TAMRA in the absence of RNA, the minimal fluorescence and the fluorescence in the presence of the concentration [RNA] of the titrated TMR3 variant, respectively; K_D is the dissociation constant.

NMR measurements

NMR measurements were carried out on 600, 900 and 950 MHz Bruker Avance III HD NMR spectrometers equipped with triple resonance H,C,N cryogenic probes. The unlabeled samples of free and ligand saturated TMR3 wt and TMR3_48nt were used to acquire ^{1}H,^{1}H NOESY spectra at 950 MHz at 10°C employing a jump-and-return water suppression scheme. The mixing time of these experiments as well as that of all the subsequent NOESY spectra was set to 150 ms. A 2D CPMG ^{15}N-HSQC-NOESY experiment (42) was obtained on the [^{15}N]-sample at 10°C on the 900 MHz spectrometer. All remaining experiments were recorded at 25°C. To establish base pairing patterns, the [^{15}N]-sample was used to obtain a standard 2D HNN-COSY spectrum as well as a selective long range (sellr)-HNN-COSY spectrum starting from the adenine H2 spins with a subsequent selective long-range transfer to the adenine N1 (43). In addition, a 2D HNN-COSY spectrum starting on the amino groups served to confirm the identity of the base pairing partners in the G–C base pairs. All HNN-COSY spectra were recorded at 900 MHz. To derive the geometry of the U8–U23 base pair from the carbonyl chemical shifts, a 2D H(N)CO spectrum was recorded at 600 MHz (44,45). The N-CO transfer delay was set to 8.3 ms in this experiment. 3D ^{1}H,^{13}C NOESY-HSQC spectra were obtained for the aromatic C-H moieties on the [^{13}C,^{15}N]- and the [^{13}C,^{15}N]-Gua, Ura labeled samples at 25°C and 950 MHz. A 2D F_1,F_2-filtered NOESY was acquired on the [^{13}C,^{15}N]-sample at 600 MHz to obtain intraligand NOEs. In addition, a 3D ^{1}H,^{13}C NOESY-HSQC for the aliphatic C-H moieties was recorded on this as well as on the [^{13}C,^{15}N]-Ade sample employing gradient coherence selection and sensitivity enhancement at 950 MHz mainly to collect and resolve NOEs between the ribose moieties and the pyrimidine H5 atoms and the imino and amino hydrogens. The [^{13}C,^{15}N]- and the selectively labeled [^{13}C,^{15}N]-Ade, Cyt and [^{13}C,^{15}N]-Gua, Ura samples were subsequently exchanged into 100% D$_2$O by repeated lyophilization and dissolution in D$_2$O. Subsequently, aliphatic 3D ^{1}H,^{13}C NOESY-HSQC spectra were acquired for the two selectively labelled samples at the 950 MHz spectrometer and for the [^{13}C,^{15}N] sample at 900 MHz. On the [^{13}C,^{15}N]-sample a 3D F1-^{13}C,^{15}N-filtered ^{1}H,^{13}C NOESY-HSQC was recorded on the 900 MHz spectrometer to obtain intermolecular NOEs between

the RNA and the unlabeled ligand. In addition, {^{1}H},^{13}C hetNOE spectra for the aromatic and anomeric C–H moieties were recorded on the uniformly labeled sample at 600 MHz. These experiments were run in duplicates in an interleaved manner with 5 s of hydrogen saturation and a recycling delay of 3 s. NMR data were recorded and processed using Bruker TOPSPIN software. Cara was used for spectra analysis (46).

Input restraints and structure calculation

Resonance assignment for the TMR3/5-TAMRA complex was achieved by standard procedures as described in detail previously (47). Based on this resonance assignment, NOE peak intensities were extracted from the various NOESY spectra and translated into five groups of upper distance restraints (<2.8, 3.5, 4.3, 5.8 and 6.5 Å). Distance calibration was performed in reference to the averaged intensity of the H5–H6 cross peaks of pyrimidine residues in stable base pairs, which were set to 2.4 Å in the 3D ^{1}H,^{13}C aromatic NOESY-HSQC spectrum recorded on the [^{13}C,^{15}N]-sample. This calibration was transferred to all other NOESY spectra. In case of the ^{1}H,^{1}H NOESY, a sixth class of even longer upper limit restraints of <7.5 Å due to the very high sensitivity of this experiment was introduced. Hydrogen bonding patterns obtained from the HNN-COSY spectra were incorporated by inclusion of two upper and two lower limit restraints, one between the donor hydrogen and the acceptor heteronucleus (2.0 and 1.8 Å) and one between the two heteronuclei (3.0 A and 2.8 A). Consequently, confirmed standard Watson–Crick G–C base pairs were constrained by three such constraint pairs, standard Watson–Crick A–U base pairs by two. The Watson–Crick G26–A39 base pair was constraint by distance restraints between the G imino group and the N1 of the adenine as well as the G O6 and the A N6 amino group. Hydrogen bonds that were consistently present based on the angles and distances of the donor and acceptor groups for the majority of the 100 calculated structures in the course of the calculations were also introduced as additional distance constraints in the final round of structure calculations. Accordingly, hydrogen bonds between the G25 imino and amino groups as well as the G40 amino group and the ligand I$_5$ carboxylate group as well as the G25 C6 carbonyl and the G40 amino group were identified and introduced. The geometry of the non-canonical base pair between U8 and U23 was established from the ^{13}C chemical shifts of the C2 and C4 carbons in the 2D H(N)CO experiment (44). It was restrained by two hydrogen bonds between the U8 C4 carbonyl and the U23 N3 imino group and the U8 imino and the U23 C2 carbonyl group. The GAAA tetraloop topping stem P2 (G14 to A17) folds into a canonical GNRA tetraloop. Therefore, the hydrogen bonds typical for this fold were introduced between the A17 N6 amino group and the G14 N3, the G14 amino group and the A17 N7 and between the G14 amino and the A16 phosphate moiety. Similarly, the UUCG tetraloop in P3 (U31 to G34) adopts a canonical UNCG tetraloop fold and was restrained by hydrogen bonds between the U31 C2 carbonyl group and both the G34 imino and amino group. Additionally, hydrogen bonding restraints were introduced between

U31 2′OH and the G34 C6 carbonyl group and the C33 N4 amino group and the U32 phosphate group as well as the C33 2′OH group and its own C2 carbonyl group. In addition, backbone as well as glycosidic torsion angles of the non-flexible loop residues G14 and U31, C33 and G34 were restricted to values extracted from published high resolution tetraloop structures (48,49). The ribose puckers of all residues in canonical base pairs with the exception of G41, U42 and C44 were restricted to C3′-endo. Canonical coordinates derived for residues for which the aliphatic carbons were resolved, suggest C2′-endo ribose conformations for U24, G41, U42, A43, while C44 shows a can_1 value in agreement with an averaged ribose pucker (50). Accordingly, U24, G41 and U42 were restricted to C2′-endo and the ribose pucker of C44 was left unrestricted. In the case of A43 NOE data did not agree with one single conformation and two different structures were calculated for this residue (A43$_{WC}$ and A43$_{HS}$, see below). The ribose pucker of A43 was restricted to C2′-endo in the A43$_{WC}$ conformation, while it was set to C3′-endo in the A43$_{HS}$ conformation. The glycosidic torsion angle of this residue appeared consistently as *syn* in the structure calculation of both conformers and was restricted accordingly in the final rounds of structure calculations. The backbone torsion angles as well as the glycosidic torsion angles of all canonical A-form residues were set to canonical values with a variance of $\pm 20°$.

Structure calculation was carried out in CYANA. A total of 100 structures was calculated with 32 000 simulated annealing steps per conformer comprising 12 000 high temperature steps of torsion angle molecular dynamics followed by 20 000 steps of slow cooling and 8000 steps of conjugate gradient minimization.

RESULTS AND DISCUSSION

TMR3 construct optimization

The tetramethylrhodamine aptamer 3 (TMR3) identified in a SELEX effort by Carothers *et al.* is an RNA of 54 nucleotides (nt). In this selection experiment, Carothers and co-workers explicitly aimed at obtaining aptamers capable of ligand binding at low Mg^{2+} concentrations. Accordingly, dissociation constants for tetramethylrhodamine (TMR, Figure 1A) of 370 nM at 5 mM Mg^{2+} and 400 nM at 1 mM Mg^{2+} concentration were originally reported that showed already a negligible effect of Mg^{2+} concentration on ligand affinity. We measured the affinity of TMR3 for TMR in the absence of Mg^{2+} and in the presence of 1 mM Mg^{2+} (Supplementary Figure S1) and found it to be independent of the presence of Mg^{2+} ions and even slightly higher than reported previously ($K_D = 353$ nM at 1 mM Mg^{2+} and 233 nM in the absence of Mg^{2+}, respectively). Consequently, all subsequent NMR-experiments were carried out in the buffer described by Carothers *et al.* (see Material and Methods) in the absence of Mg^{2+}.

The TMR3 aptamer was predicted to fold into an open three-way junction (Figure 1B) comprising a closing stem (P1) and two short hairpins (P2 and P3) connected by stretches of 5–8 formally single stranded residues. The short 3 base pair containing hairpin P2 is closed by a canoni-

cal GNRA tetraloop with a GAAA sequence and a C–G closing base pair. P3 is topped by a canonical cUUCGg tetraloop. The originally predicted secondary structure of free TMR3 suggested the presence of 13 Watson–Crick base pairs (11 G–C and 2 A–U base pairs). However, initial spectra of the free wt-RNA immediately showed that already in the absence of the ligand more Watson–Crick base pairs than originally predicted are present. Imino group signals corresponding to 11 Watson–Crick G–C and 4 WC A–U base pairs as well as the typical imino group signals for the G residues in the GAAA and UUCG tetraloops were observable in the $^1H,^{15}N$-HSQC spectrum of the free wt-RNA (Figure 1C). $^1H,^1H$ NOESY data (data not shown) suggests that already in the free TMR3 RNA P2 is actually extended by two G–C and two A–U base pairs compared to the original prediction (Figure 1B). Upon addition of 5-carboxy tetramethylrhodamine (5-TAMRA, Figure 1A), the commercially available TMR analogue used in the original SELEX procedure, most of the resonances of free TMR3 are retained, in particular those of the two canonical tetraloops (Figure 1C). In addition, a number of new resonances appear in spectral regions characteristic for both canonical and non-canonical base pairs. Based on distance information from $^1H,^1H$ NOESY data for imino protons stems P1, P2 and P3 contain at least 6, 7 and 4 consecutive Watson–Crick base pairs, respectively, in the RNA–ligand complex (Figure 1D). The presence of extended stem regions not directly involved in ligand binding prompted us to devise a minimized TMR3 variant of only 48 nt length, which lacks one G–C base pair in P1 and two G–C base pairs in P2 (Figure 1D). This minimized aptamer (TMR3_48nt) shows a $^1H,^{15}N$-HSQC spectrum similar to the wt-RNA both for the free RNA and when bound to 5-TAMRA (Figure 2A). Importantly, in the spectrum of the TMR3_48nt/5-TAMRA-complex the imino resonances corresponding to non-canonical hydrogen bonding interactions associated with 5-TAMRA binding to the wt aptamer are retained and only resonances for three canonical G–C base pairs are missing, indicating that TMR3_48nt adopts a structure very similar to the wt RNA. In addition, fluorescence titrations show that the ligand affinity of TMR3_48nt is only reduced slightly ($K_D = 516 \pm 16$ nM) compared to the wt (233 ± 6 nM, Figure 2B). Initial structural data obtained from $^1H,^{15}N$ HSQC spectra together with $^1H,^1H$ NOESY distance information verifies the presence of three stems coming together to form a closed three-way junction which forms the ligand binding site (Figure 2C and D). In order to profit from improved spectral resolution of the smaller RNA–ligand complex, we proceeded with the structure determination of the TMR3–ligand complex by NMR using the minimized TMR3_48nt aptamer.

Overall structure of the TMR3/5-TAMRA complex

The structure of the TMR3_48nt/5-TAMRA complex was derived from a vast network of uniformly distributed NOE-distance restraints with an average of 31 restraints per residue (Supplementary Figure S2A). The ligand 5-TAMRA is positioned in the core of the three-way junction by ∼140 interresidual NOE distance restraints to adjacent residues of all three helical stems (Supplementary Fig-
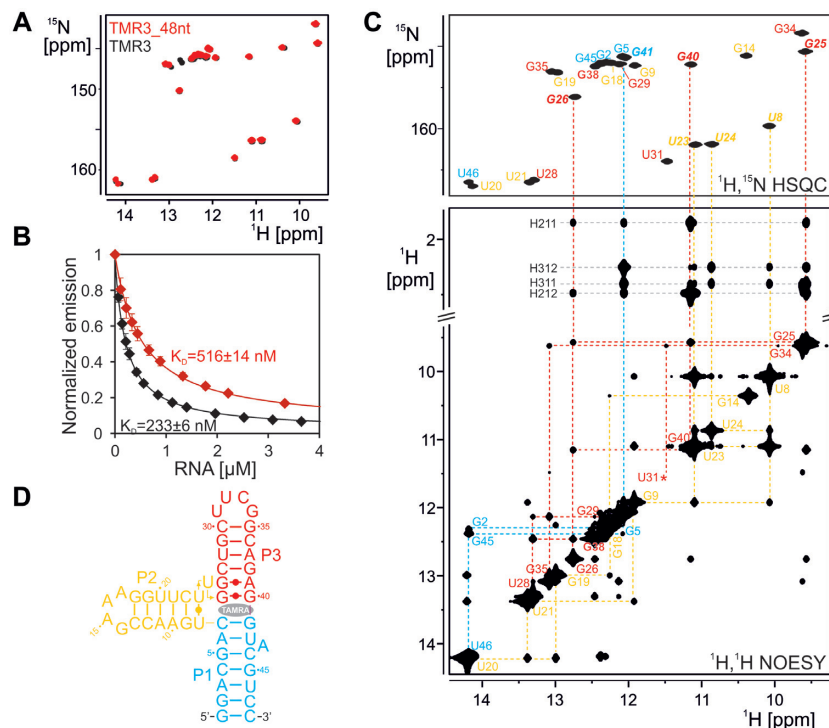
Figure 2. TMR3 construct optimization and initial structure characterization. (**A**) Imino group ^{1}H,^{15}N HSQC spectra of TMR3 wt (black) and its minimized variant TMR3_48nt (red) in the presence of 5-TAMRA. (**B**) Fluorescence titrations of 5-TAMRA with TMR3 wt (black) or TMR3_48nt RNA (red). Single binding site fits to the titration data are also shown (solid lines) together with the resulting dissociation constants. (**C**) Imino group ^{1}H,^{15}N HSQC spectrum (top) and imino as well as ligand methyl group regions of a ^{1}H,^{1}H NOESY spectrum of TMR3_48nt in complex with 5-TAMRA (bottom). Imino-imino proton NOE connectivities used to obtain the assignment of signal in stems P1 to P3 are given in different colors according to the color scheme in D. Imino resonance and 5-TAMRA methyl group assignments are given. (**D**) Secondary structure of the TMR3_48nt/5-TAMRA complex in Leontis–Westhof notation. Colors of the different stems P1–P3 as in C. The ligand is shown as a gray ellipse.

Table 1. NMR structural statistics for the two conformers of the TMR3_48nt/5-TAMRA complex. In cases where the numbers of input restraints or the RMSD values differ for the two A43 sub-structures, separate values are given (A43$_{WC}$/A43$_{HS}$)

	A43$_{WC}$ / A43$_{HS}$
Total distance restraints	**1615/1616**
TMR3_48nt RNA	1469/1470
Intra-residue	*632*
Sequential	490
Long-range	*231/232*
Hydrogen bonds	*116*
TMR3_48nt/5-TAMRA	146/147
NOE distance restraints	*140/141*
Hydrogen bonds	*6*
Total dihedral restraints	**356**
Ribose pucker	*88*
Backbone	*223*
Glycosidic torsion	*45*
Structural statistics	
Average RMSD from mean structure (Å)	
G2–G14, G18–U31, C33–C47, 5-TAMRA	**0.74 ± 0.45/0.70 ± 0.40**

ure S2B). The NMR-structural bundle of the 10 lowest energy structures of the TMR3_48nt/5-TAMRA complex is shown in Figure 3A. The corresponding secondary structure is depicted in Figure 3B. NMR-structural statistics are summarized in Table 1. The structural bundle is well converged with a heavy atom RMSD from the mean structure of all non-flexible residues of ∼0.7 Å.

In contrast to the prediction, stem P1 does not form only a five base pair canonical helix but a helix with seven continuously stacked Watson–Crick base pairs. A43 is extruded from P1 and thereby enables uninterrupted stacking between the G5:C44 and the A6:U42 base pair. However, the A6:U42 base pair is less stable compared to other A:U base pairs in the molecule due to its vicinity to the bulged out A43 as judged from its imino signal line width and temperature dependence (Supplementary Figure S3). The low intensity of this resonance prevented the characterization of the base pairing of this residue by standard experiments. However, the Watson–Crick geometry of the A6:U42 base pair was unambiguously established in a selective long-range (sellr) HNN-COSY experiment starting on the non-exchangeable H2 of A6 where the signature correlation across the N - - H–N type hydrogen bond between the A6 N1 and the U42 N3 was detected (Supplementary Figure S3).

P2 and P3 form the expected stem-loop structures topped by a canonical GAAA and a canonical UUCG tetraloop, respectively, but are extended by non-Watson-Crick base pairs at the opposite ends. P3 is extended towards the three-way junction by an additional *trans* Watson–Crick/Watson–Crick G26:A39 and a *trans* Watson–Crick/Watson–Crick G25:G40 base pair (Figure 3C, right). The geometry of the Watson–Crick G26:A39 base pair was also verified in the sellr HNN-COSY experiment (Supplementary Figure S3). The G25:G40 base pair
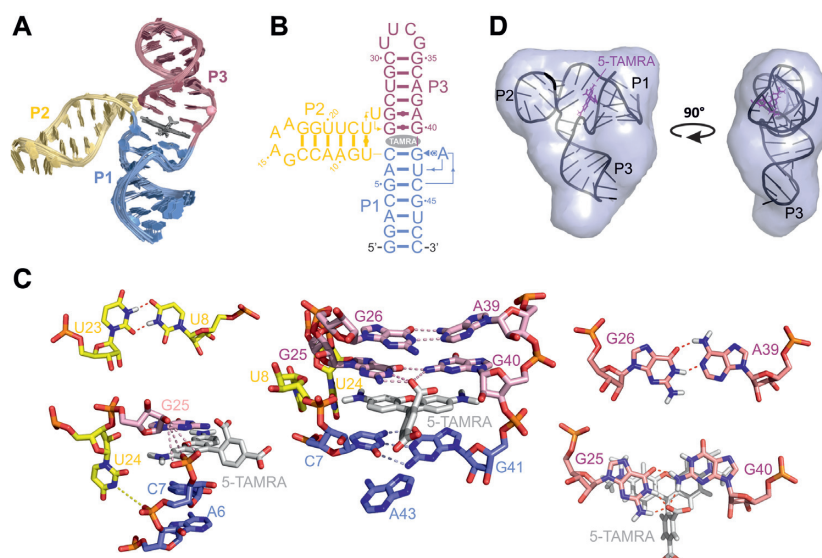
Figure 3. NMR structure of the TMR3_48nt/5-TAMRA complex. (**A**) Bundle of the 10 lowest energy structures of the TMR3_48nt/5-TAMRA complex obtained from the structure calculation. All rigid residues as well as the ligand are superimposed (G2–G14, G18–U31, C33–C47, 5-TAMRA). The RNA is shown in cartoon representation; the ligand is shown as gray sticks. The three different stems are colored corresponding to b. (**B**) Secondary structure of the complex in Leontis–Westhof notation. The variable interaction of A43 with G41 is shown as a combination of Hoogsteen and Watson–Crick base pairing. (**C**) Structural details of the ligand binding pocket (middle). Individual base pairs and bases of P2 (left) and P3 (right) lining the binding pocket are shown. Potential hydrogen bonds are indicated by dashed lines, residues are labeled. RNA and ligand are shown in stick representation. RNA carbon atoms of the residues belonging to the different stems are colored according to the color scheme in B. Oxygen and nitrogen atoms are depicted in red and blue, respectively. 5-TAMRA is shown in gray. (**D**) SAXS envelope of the TMR3_48nt/5-TAMRA complex in two orthogonal orientations with the lowest energy NMR structure fitted into the envelope.

is stabilized by one intra-RNA hydrogen bond between the O6 of G25 and the amino group of G40. The presence of this two non-canonical base pairs at the end of P3 pointing towards the three-way junction leads to a gradual increase of the P3 helix diameter (Supplementary Figure S4). P2 is extended by a *trans* Watson–Crick/Watson–Crick U8:U23 base pair stabilized by two direct hydrogen bonds. While U24 remains formally unpaired it is stably stacked on top of the U8:U23 base pair (Figure 3C, left). Its imino proton is protected from exchange with the solvent and it shows an extended network of NOEs to the ligand as well as to the U8:U23 base pair (see Supplementary Figure S2B). The two individual stem–loops are very well converged. This is reflected in an average heavy atom RMSD from the mean structure for P2 (residues 9–24) of 0.28 Å and a heavy atom RMSD for P3 (residues 25–40) of only 0.20 Å (Supplementary Figure S5).

Due to the extended helices of P1, P2 and P3 in the ligand bound RNA the junctions between the three helical stems are very short. In each case they consist of only one phosphodiester group. Consequently, the orientation of the three helical stems in the three-way junction with respect to each other is structurally very well defined. The ligand 5-TAMRA mediates continuous helical stacking between P1 and P3. P2 is oriented perpendicular relative to the coaxially stacked P1 and P3 stems and to the extended three-ring aromatic system of the ligand. The very dense network of intramolecular and intermolecular NOE distance constraints spanning the junction region is sufficient to precisely define the conformation of the three-way junction even without the inclusion of additional long range structural restraints such as RDCs into the structure calculation. However, we

performed additional SAXS experiments on TMR3_48nt and the TMR3_48nt/5-TAMRA complex (Supplementary Figure S6). The SAXS data demonstrate that TMR3_48nt undergoes only a slight global structural rearrangement including an overall compaction upon 5-TAMRA binding as indicated by the similar Kratky plots, pairwise distance distributions, maximal dimensions (60/55 Å) and radii of gyration (19/19 Å). Importantly, the general triangular shape of the TMR3/5-TAMRA complex and the resulting relative orientation of the three stems with regard to each other as obtained from the NMR NOE data was confirmed by the SAXS measurements (Figure 3D).

The ligand binding pocket

The ligand binding pocket of TMR3 is lined by the C7:G41 Watson–Crick base pair closing stem P1, the *trans* WC/WC G25:G40 base pair of P3 and the unpaired U24 closing stem P2. 5-TAMRA intercalates with its planar xanthene ring system between the C7:G41 base pair of P1 and the G25:G40 base pair of P3 (see Figure 3C). In particular, the large three-ring xanthene moiety of the ligand is involved in extensive stacking interactions with the G25:G40 base pair. Ring I of 5-TAMRA is tilted by an angle of ~113° relative to its xanthene moiety and protrudes into the minor groove of the aptamer RNA. The carboxylate group at position I_5 (see Figure 1A), which is also part of the cognate ligand TMR, is oriented in the direction of the G25:G40 base pair such that it can form hydrogen bonding interactions with both residues. Both the amino and the imino group of G25 as well as the amino group of G40 are in the appropriate distance and orientation to form hydrogen bonds to the I_5

carboxylate group (Figure 3C, right). 5-TAMRA intercalation into the RNA between the two consecutive residues G40 (P3) and G41 (P1) is facilitated by the C2′-endo conformation of the G41 ribose moiety. This increases the inter-residue distance between G40 and G41 (C1′–C1′ distance of 7.72 ± 0.01 Å) which leaves enough room for the xanthene moiety of the ligand to intercalate. Furthermore, one of the positively charged dimethyl amino groups of the ligand is oriented towards the phosphate backbone group connecting G40 and G41 suggesting a favorable electrostatic interaction between the RNA and the ligand at this position.

Ligand stacking with the C7:G41 base pair at the terminus of the P1 stem is further extended by the formation of a dynamic base triple involving this base pair and the looped-out A43. ${^{13}C},^{1}$H-hetNOE data of residues in the TMR binding site demonstrate that for A43 both the nucleobase and the ribose moiety show increased flexibility while all other binding site residues are rigid (Figure 4A). However, although A43 appears to be conformationally dynamic, apparently its presence contributes to ligand binding since its deletion results in a ∼2.5-fold decrease of the binding affinity (Figure 4B). A detailed analysis of the NOE data observed for H2 and H8 of A43 reveals distance restraints which cannot be fulfilled simultaneously in one single conformation (Figure 4C and Supplementary Figure S7). An illustrative example for this is the simultaneous presence of NOE cross peaks for A43 H2 to both U42 H5′ and C44 H4′. To solve this conformational ambiguity, H2 and H8 NOEs of A43 were manually sorted to agree to two different conformations of A43. Starting from the most severe mutually exclusive NOEs, NOE derived upper limit restraints agreeing more with the one or the other conformation were incorporated into two separate structure calculation procedures in a stepwise fashion. A number of distance restraints fulfilled in both resulting sub-structures were included in both calculations, while others with significant distance differences in the two sub-structures were only introduced into the structure with the shorter inter-proton distance (Figure 4C). This procedure resulted in two sub-structures which together explain all observed NOE cross peaks. In both sub-structures the A43 nucleobase is oriented in a potential minor groove base triplet with the WC C7:G41 base pair (Figure 4D). In one conformation (A43$_{WC}$, Figure 4D left), A43 is oriented with its Watson–Crick edge towards C7:G41 such that A43 N1 can form a hydrogen bond with G41 N2H22 and A43 N6H61 can form a hydrogen bond with C7 O2. This conformation of the A43 nucleobase is in good agreement with a C2′-endo sugar pucker and the nucleobase in a *syn* conformation. In the other conformation (A43$_{HS}$; Figure 4D, right) the N7 of the Hoogsteen edge of A43 can interact with the G41 N2H22 and the A43 amino group forms a hydrogen bond to the O2 of C7. A C3′-endo ribose pucker is preferably adopted in this conformation together with a glycosidic bond in a *syn* conformation. The proposed difference in ribose pucker of the two conformations agrees well both with the observed elevated ${^{13}C},^{1}$H-hetNOE value for the A43 C1′ and the averaged canonical coordinates (*can1* = –6.34) for this ribose moiety (50). The difference between the A43 nucleobase and ribose conformation in the two sub-structures results in a local difference in the backbone conformation between residues G41 and

C44 (Figure 4E). However, the global structures are similar with an RMSD between the lowest energy structures of the A43$_{WC}$ and the A43$_{HS}$ conformer of 0.95 Å. Our detailed structural analysis of the two A43 conformers suggests that in both cases two hydrogen bonds are formed to the C7:G41 base pair resulting in roughly the same contribution to the stabilization of the TMR3/5-TAMRA complex and rationalizing the presence of these two A43 conformations in the structural ensemble of the complex. For both conformers, the A43 nucleobase forms a T-shaped stacking interaction with the aromatic ring I of 5-TAMRA. Together, hydrogen bonding and stacking interactions provide an explanation for the decrease in binding affinity upon A43 deletion (Figure 4B).

The P1/P2 junction is formed by the phosphodiester backbone between C7 and U8. The U8 phosphate group and its ribose moiety thus get in close proximity to the ribose moiety of G25 at the base of P3 (see Figure 3C, left). In principle, a hydrogen bond could be present between the 2′OH of G25 and the U8 phosphate group to stabilize the local arrangement. However, the heavy atom distance between the hydroxyl group and either the 5′-bridging or either of the two non-bridging phosphate oxygens in the NMR structures is ∼4.1 and 4.2 Å, respectively, and thus too long by ∼1 Å for a hydrogen bond. Hydroxyl group protons protected against fast exchange with the solvent by hydrogen bonding interactions should be observable in long-range ^{1}H,^{13}C-HSQC spectra due to correlations with the ribose C2′ and C3′ carbons. However, no such correlation was observed in this experiment for the G25 2′OH group. In addition, no correlation was observed between any exchangeable hydrogen and a phosphate resonance in ^{1}H,^{31}P-HMQC spectra with long transfer delays optimized to detect hydrogen bonds (51,52).

The base of P2 is formed by an unpaired uridine residue, U24, which stacks between the U8:U23 base pair and a ligand dimethyl amino group (Figure 3C, left). The imino proton resonance of this residue is comparatively narrow suggesting protection from exchange with the solvent by either hydrogen bonding or steric exclusion. In the NMR structure the only hydrogen bonding acceptor group in the vicinity of the U24 H3N3 group is the non-bridging OP2 phosphate oxygen of C7 in a distance of ∼4.5 Å. The imino hydrogen chemical shift of 10.8 ppm for U24 is in good agreement with a phosphate group as hydrogen bonding acceptor (51) but also for an imino proton protected from exchange due to burial in a hydrophobic environment (53), e.g. due to the spatial proximity to the hydrophobic ligand. In agreement with the latter possibility, no ^{1}H,^{31}P-correlation across a potential hydrogen bond was detected. Notably, the C2 carbonyl group of U24 and the phosphate group of U8 are located within a distance of ∼4 Å of the positively charged nitrogen of the dimethyl amino group from the ligand, suggesting that electrostatic interactions between these moieties further contributes to ligand recognition.

TMR3 – a minimized xanthene binding site

The TMR3 aptamer binds 5-TAMRA with an affinity in the high nanomolar range due to the intercalation of the planar xanthene moiety of the ligand between a large purine:purine
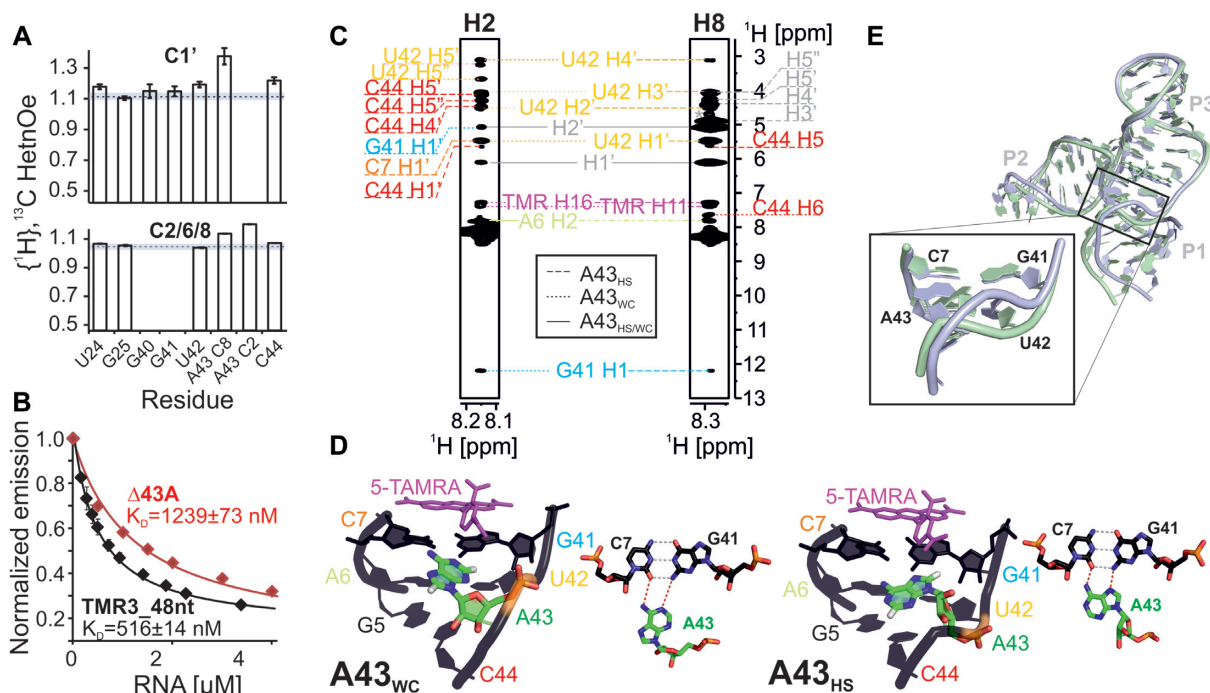
Figure 4. Conformational equilibrium of A43. (**A**) {^{1}H},^{13}C HetNOE of C1′ (top) and the aromatic carbons C2/C6 or C8 (bottom) of residues forming the 5-TAMRA binding site. (**B**) Fluorescence titrations of 5-TAMRA with TMR3_48nt (black) and TMR3_48nt_Δ43A (red). (**C**) ^{1}H,^{1}H slices from 3D ^{1}H,^{13}C NOESY-HSQC spectra at the resonance positions of A43 C2 (left) and A43 C8 (right). Resonance assignments of interresidual NOE cross peaks are given in colors corresponding to the residue color scheme in d. Cross peaks, for which the upper limit restraints were incorporated into the structure calculation of the A43$_{HS}$ subpopulation, are connected to their assignments via dashed lines. Those, which were incorporated into the structure calculation of the A43$_{WC}$ calculation, are connected via dotted lines. Cross peaks, for which the upper limit restraints were incorporated in both calculations are connected to their assignments via solid lines. (**D**) A43$_{WC}$ (left) and A43$_{HS}$ (right) substructure. The upper part of P1 together with 5-TAMRA is shown in stick representation together with a top view of the minor grove base triple formed by A43 and the WC base pair C7–G41 with the hydrogen bonds indicated by dashed lines. The ligand is shown in magenta, the RNA in black with the exception of A43, which is colored according to atom type with carbon atoms in green, nitrogen atoms in blue, oxygens in red and hydrogens in white. Residues are assigned and assignments are colored as in C. (**E**) Global superposition of the A43$_{WC}$ (gray) and the A43$_{HS}$ (green) substructure. The RNA is shown as cartoon, the ligand in stick representation. Local differences in the backbone conformation of P1 are shown in an inset.

base pair on one side and a dynamic base triple on the other side. Specific hydrogen bonding interactions are formed between the RNA and the I$_5$ carboxylate group with possible additional contributions from electrostatic interactions between the positively charged dimethyl amino groups of the ligand and the phosphate backbone between G40 and G41 as well as with the U8 phosphate and the U24 C2 carbonyl group. The binding site of TMR3 shares similarities with other known aptamers. For instance, the FMN aptamer accommodates the extended three-ring aromatic xanthene moiety of the ligand in its binding site by extending the diameter of a helix with a Watson–Crick G–A base pair followed by a G–G base pair of the same geometry as the G26:A38 and G25:G40 base pairs in TMR3, suggesting a common stacking motif for extended aromatic ring systems (6). Another aptamer which is structurally related to TMR3 is the malachite green aptamer (MG-aptamer). While originally selected to bind to the fluorescent triarylmethyl dye malachite green (MG, Figure 5A, top), this aptamer also forms a tight complex with MG's xanthene derivative *N,N′*-tetramethyl-rosamine (TMS; Figure 5A, bottom). In fact, due to its rigid scaffold, TMS surpasses the affinity of MG to the MG-aptamer by a factor of ∼20 (K_D ∼40 nM versus 800 nM) (8). It thus binds ∼10-fold more tightly to the MG-

aptamer than 5-TAMRA to the TMR3-aptamer although compared to 5-TAMRA it lacks the carboxylate group on ring I as potential interaction partner with the RNA. Interestingly, despite its structural similarity to 5-TAMRA, MG does not show any affinity to TMR3 (Figure 5B).

The structure of the MG-aptamer in complex with TMS (Figure 5C–E) solved by X-ray crystallography reveals some interesting similarities to the TMR3/5-TAMRA complex regarding the intercalation of the ligand between the closing stem P1 and a coaxially stacking stem–loop P3 (8). Furthermore, one side of the TMS binding site core is formed by two consecutive residues where one of the dimethyl amino groups of the ligand interacts electrostatically with the phosphate group connecting these two residues (Figure 5D). This is structurally very similar to the intercalation of 5-TAMRA between residues G40 and G41 in the TMR3 complex. On the opposite end the binding site is closed by a uridine in a U-turn structure, which is oriented perpendicular to the aromatic plane of the ligand. The U-turn uridine closes the binding pocket by an interaction with a backbone phosphate group (Figure 5D). This resembles the orientation of U24 in our TMR3 complex structure towards the backbone phosphate group of C7 and to the dimethyl amino group of the ligand. Therefore, the U-
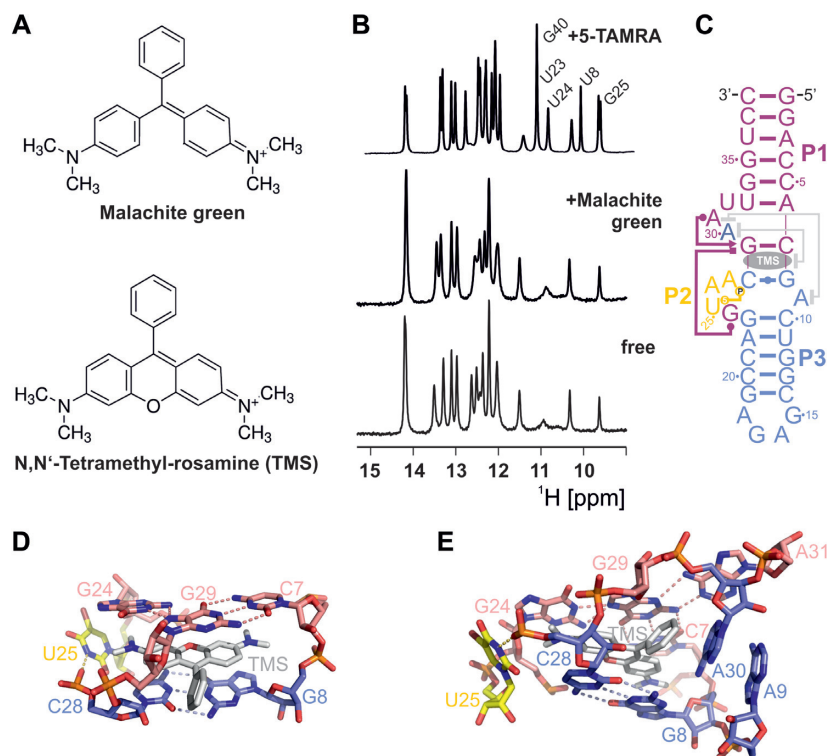
Figure 5. Comparison to the structure of the malachite green aptamer. (**A**) Chemical structure of malachite green (top) and *N*,*N*'-tetramethyl-rosamine (TMS, bottom). (**B**) Imino regions of 1D ^{1}H spectra of free TMR3_48nt RNA (bottom) and titrated with malachite green (middle) or 5-TAMRA (top). Resonances which appear upon binding to 5-TAMRA are assigned in the spectrum of the TMR3_48nt/5-TAMRA complex. None of these resonances appear upon titration with malachite green in agreement with an absence of malachite green binding to TMR3. (**C**) Secondary and tertiary structure of the malachite green aptamer in complex with TMS in Leontis–Westhof notation. The three stems are colored according to the color scheme of the TMR3/5-TAMRA complex in Figure 3B. (**D**) Close-up of the immediate binding site for TMS in the malachite green aptamer (pdb entry 1F1T) shown in stick representation. Carbon atoms are colored according to the secondary structure diagram shown in C. Oxygens are shown in red, nitrogens in blue. Hydrogen bonds are given as dashed lines. Residue assignments are also given. (**E**) Environment of the ligand binding pocket of the malachite green aptamer/TMS complex highlighting the ligand induced formation of a base triple between A31 and the C7:G29 Watson–Crick base pair and the extended stacking interactions between the hydrophobic phenyl ring of the ligand, A30 and A9. Representation as described in D.

turn of the MG-aptamer could be considered as an equivalent of the P2 stem in the TMR3 complex structure. Interestingly, while the floor of the TMS binding site is formed by a G:C WC base pair as in the TMR3-aptamer, its top consists of a G–G–C base triplet (Figure 5E) compared to the extended G:G base pair in TMR3. Remarkably, one of the G-residues (G29) in the base triplet occupies the position of the I$_5$ carboxylate group of 5-TAMRA. In turn, ring I of TMS is involved in a base stacking sandwich with an adenine base moiety and a ribose moiety (Figure 5E). Despite the higher binding affinity compared to the TMR3/5-TAMRA complex no specific hydrogen bonding interactions between ligand functional groups and the RNA can be found stabilizing the MG-aptamer/TMS complex. Instead, the higher affinity of the MG-aptamer/TMS interaction apparently originates from additional ligand induced RNA–RNA interactions (Figure 5E). Thus, the extended binding site reveals additional base pairing and stacking interactions, such as a ligand induced long range A:A stacking interaction which is extended by stacking on ring I of TMS (Figure 5E). Extensive ligand induced RNA–RNA interactions are also a feature of other RNA–ligand complexes resulting in high ligand affinity despite a sometimes limited number of direct specific RNA–ligand interactions.

One very striking example in this regard are the aptamer domains of the purine sensing riboswitches, where ligand binding stabilizes long-range loop–loop kissing interactions distant to the binding site (54). The network of ligand induced RNA-RNA interactions, however, appears to be more limited in the case of the TMR3 aptamer and might therefore explain the lower affinity to its ligand in comparison to the MG-aptamer. TMR3 can be thus seen as a 'minimal' binding motif for TMR.

Potential of TMR3/5-TAMRA as a synthetic riboswitch

Comparison of the ^{15}N-HSQC spectra of the free and 5-TAMRA bound TMR3_48nt aptamer (Supplementary Figure S8A) shows, that in the spectrum of the free RNA a number of imino proton signals is missing compared to the spectrum of the ligand bound RNA. This indicates an increased exchange with the bulk solvent and less stable or absent base pairing in the free RNA. Missing for instance are the imino proton signals for U8 and U23 which form a U:U base pair and for U24 in P2 as well as for G25 of the G25:A40 base pair and for G26 and G41 of the G25:G41 base pair extending P3. Also unobservable are the imino protons for G41 of the C7:G41 base pair and U42 of the
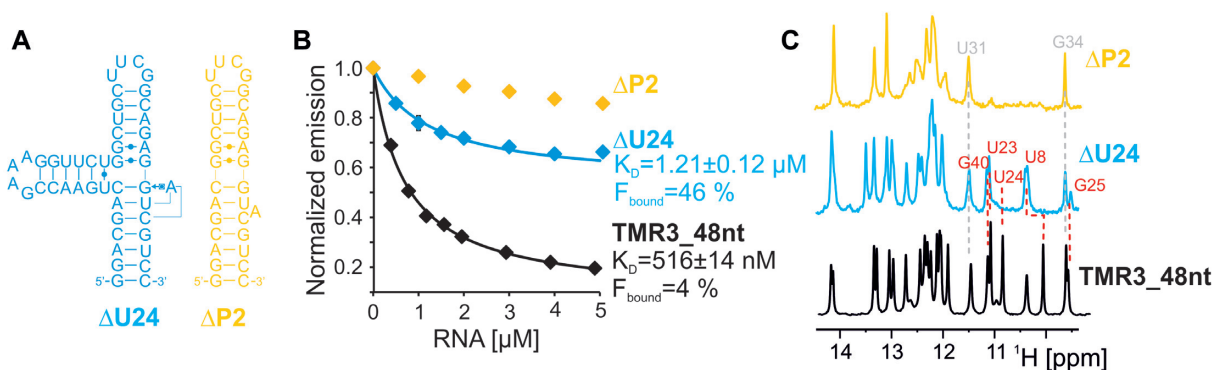
Figure 6. Variation of P2 influences fluorescence properties of TMR3. (**A**) Secondary structures of the TMR3 mutants TMR3_48nt_Δ24U (blue, left) and TMR3_48nt_ΔP2 (yellow, right) based on the secondary structure of TMR3_48nt_wt. (**B**) Fluorescence titrations of 5-TAMRA with TMR3_48nt (black) and the deletion mutants TMR3_48nt_Δ24U (blue) and TMR3_48nt_ΔP2 (yellow). TMR3_48nt_ΔP2 apparently does not bind to 5-TAMRA. For TMR3_48nt and ΔU24, curve fitting to a single-site binding event was performed. The resulting dissociation constants together with the fluorescence intensity of 5-TAMRA in the RNA complex relative to the free 5-TAMRA (F_{bound}) are given. (**C**) Imino group region of ^{1}H 1D spectra of TMR3_48nt (black), ΔU24 (blue) and ΔP2 (yellow) in complex with 5-TAMRA. Resonance assignments of the imino groups of residues close to the 5-TAMRA binding site which appear upon 5-TAMRA binding are assigned and followed from the wild-type to the ΔU24 spectrum by red dashed lines. Resonances characteristic for the presence of the UUCG tetraloop are assigned in gray in the spectrum of the ΔP2 mutant.

A6:U42 base pair in P1. All these nucleotides are in the direct vicinity of the ligand binding site suggesting that the core of the 3-way junction is conformationally unstable in the free RNA (Supplementary Figure S8). Thus, 5-TAMRA binding induces an extension of all three helices by stabilizing canonical and non-canonical base pairings enabling uninterrupted ligand-mediated stacking between P1 and P3 and thereby a stable structure of the otherwise dynamic and open three-way junction of the free TMR3 aptamer with all residues except A43 in a stable structural arrangement in the ligand bound state. Ligand binding thus brings the three otherwise structurally independent stem regions together and couples their stability. In this ligand mediated stability coupling, the TMR3/5-TAMRA system displays the functional characteristics observed in other natural as well as synthetic riboswitches (55). The ligand dependent structural changes accompanied by the formation of a continuously stabilized co-axially stacking P1-P3 stem of 13 base pairs suggests that TMR3 might be a suitable candidate for the generation of a synthetic riboswitch with functional characteristics analogous to the neomycin riboswitch in yeast. This synthetic riboswitch acts on the translational level by forming a ligand-stabilized road block for the scanning small ribosomal subunit when bound to neomycin (33) whereas the unstable structure of the free form of the synthetic riboswitch can be dissolved by the small ribosomal subunit. Systematic studies on the synthetic neomycin riboswitch have shown that the minimum stem length, which cannot be dissolved by the scanning small ribosomal subunit is ~9 consecutive base pairs (33). Free TMR3 features 4–5 base pairs in its three unconnected stems and 13 bp of the continuous P1-P3 stem in its ligand bound state and therefore seems to be well suited as a synthetic riboswitch in this respect. The three-way junctional architecture of TMR3 with ligand binding in the core of the junction also leaves the size and composition of the individual stems variable for optimization with regard to their relative stabilities.

Stem P2 is a determinant for TMR3 fluorescence quenching efficiency

5-TAMRA is a strongly fluorescent dye used frequently as a fluorescence label. Fluorescent RNA aptamer/ligand complexes have attracted significant interest in the recent past, as they harbor great potential for *in vivo* RNA visualization efforts (56–58). In general, fluorogenic RNA aptamers such as Spinach (59) or Mango (60) induce the fluorescence of otherwise non-fluorescent ligands by rigidifying their conformation in the RNA-bound state. Interestingly, the TMR3 aptamer quenches the fluorescence of 5-TAMRA almost completely in the bound state (96 ± 0.7%). In contrast, the MG-aptamer has the opposite effect on the conformationally flexible MG, which is virtually non-fluorescent in the free form and lights up upon binding to the MG-aptamer showing a ~2000-fold increase in fluorescence brightness upon binding (61). This dramatic gain in fluorescence intensity has been attributed to the loss in flexibility upon intercalation of MG into the binding site, leaving fluorescence as the only pathway for energy dissipation after excitation. Considering the partial similarity of the MG- and the TMR3–aptamer binding sites, their different effect on ligand fluorescence is surprising. In order to test determinants for fluorescence quenching, we generated several binding site mutants (Figure 6A). Deletion of A43 in the base triple below the bound ligand has only a moderate effect on affinity (K_D of 1.24 ± 0.07 μM versus 516 ± 14 nM of the wt, see Figure 4B) and almost no effect on fluorescence quenching efficiency (91 ± 3%). In contrast, deletion of U24 at the base of P2, has a similarly moderate effect on affinity (K_D of 1.17 ± 0.21 μM), while fluorescence quenching is significantly reduced to only 46 ± 1% (Figure 6B). Moreover, the imino proton NMR spectrum shows that the relevant binding features are retained for the ΔU24 mutant, while in the case of a ΔP2 mutant with a completely deleted P2 stem (Figure 6A), only correct folding by virtue of the present UUCG tetraloop resonances can be verified but no indication of ligand binding is found

(Figure 6C, Supplementary Figure S9). Thus, the interaction of the base at the P2 stem with the dimethyl amino group of 5-TAMRA appears to constitute a major determinant for fluorescence quenching by the RNA. This first step towards a RNA/5-TAMRA complex which retains the fluorescence of the ligand suggests that a fluorescence optimized variant of TMR3 could form a light-up RNA/ligand complex with a 5-carboxyl analogue of MG.

DATA AVAILABILITY

Atomic coordinates for the 10 lowest energy structures of the reported NMR structural bundles have been deposited with the Protein Data bank under accession number 6GZK (A43$_{WC}$) and 6GZR (A43$_{HS}$). NMR resonance assignments are available from the Biological Magnetic Resonance Databank (BMRB) under accession number 26842.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
2. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage-T4 DNA-polymerase. *Science*, **249**, 505–510.
3. Dieckmann,T., Suzuki,E., Nakamura,G.K. and Feigon,J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.
4. Jiang,F., Kumar,R.A., Jones,R.A. and Patel,D.J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183–186.
5. Sassanfar,M. and Szostak,J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550–553.
6. Fan,P., Suri,A.K., Fiala,R., Live,D. and Patel,D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.*, **258**, 480–500.
7. Burgstaller,P. and Famulok,M. (1994) Isolation of RNA aptamers for biological cofactors by in-vitro selection, *Angew. Chem. Int. Ed. Engl.*, **33**, 1084–1087.
8. Baugh,C., Grate,D. and Wilson,C. (2000) 2.8 angstrom crystal structure of the malachite green aptamer. *J. Mol. Biol.*, **301**, 117–128.
9. Grate,D. and Wilson,C. (1999) Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6131–6136.
10. Zimmermann,G.R., Jenison,R.D., Wick,C.L., Simorre,J.P. and Pardi,A. (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat. Struct. Biol.*, **4**, 644–649.
11. Jenison,R.D., Gill,S.C., Pardi,A. and Polisky,B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–1429.
12. Famulok,M. (1994) Molecular recognition of amino-acids by RNA-aptamers - an citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.*, **116**, 1698–1706.
13. Yang,Y.S., Kochoyan,M., Burgstaller,P., Westhof,E. and Famulok,M. (1996) Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science*, **272**, 1343–1347.
14. Huang,Z. and Szostak,J.W. (2003) Evolution of aptamers with new specificity and new secondary structures from an ATP aptamer. *RNA*, **9**, 1456–1463.
15. Davis,J.H. and Szostak,J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11616–11621.
16. Carothers,J.M., Oestreich,S.C., Davis,J.H. and Szostak,J.W. (2004) Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, **126**, 5130–5137.
17. Carothers,J.M., Oestreich,S.C. and Szostak,J.W. (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J. Am. Chem. Soc.*, **128**, 7929–7937.
18. Carothers,J.M., Davis,J.H., Chou,J.J. and Szostak,J.W. (2006) Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *RNA*, **12**, 567–579.
19. Nasiri,A.H., Wurm,J.P., Immer,C., Weickhmann,A.K. and Wöhnert,J. (2016) An intermolecular G-quadruplex as the basis for GTP recognition in the class V-GTP aptamer. *RNA*, **22**, 1750–1759.
20. Wolter,A.C., Weickhmann,A.K., Nasiri,A.H., Hantke,K., Ohlenschläger,O., Wunderlich,C.H., Kreutz,C., Duchardt-Ferner,E. and Wöhnert,J. (2017) A stably protonated adenine nucleotide with a highly shifted pKa value stabilizes the tertiary structure of a GTP-binding RNA aptamer. *Angew. Chem. Int. Ed. Engl.*, **56**, 401–404.
21. Hermann,T. and Patel,D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
22. Roth,A. and Breaker,R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, **78**, 305–334.
23. Montange,R.K. and Batey,R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biochem.*, **37**, 117–133.
24. Wang,J.X. and Breaker,R.R. (2008) Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. Cell Biol.*, **86**, 157–168.
25. McCown,P.J., Liang,J.J., Weinberg,Z. and Breaker,R.R. (2014) Structural, functional, and taxonomic diversity of three preQ$_1$ riboswitch classes. *Chem. Biol.*, **21**, 880–889.
26. Battaglia,R.A. and Ke,A. (2018) Guanidine-sensing riboswitches: how do they work and what do they regulate? *Wiley Interdiscip. Rev. RNA*, **9**, e1482.
27. Mandal,M. and Breaker,R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.*, **11**, 29–35.
28. 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.
29. Sherlock,M.E., Sudarsan,N. and Breaker,R.R. (2018) Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 6052–6057.
30. Werstuck,G. and Green,M.R. (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science*, **282**, 296–298.
31. Grate,D. and Wilson,C. (2001) Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg. Med. Chem.*, **9**, 2565–2570.

32. Suess,B., Fink,B., Berens,C., Stentz,R. and Hillen,W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.*, **32**, 1610–1614.

33. Weigand,J.E., Sanchez,M., Gunnesch,E.-B., Zeiher,S., Schroeder,R. and Suess,B. (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, **14**, 89–97.

34. Kim,D.S., Gusti,V., Pillai,S.G. and Gaur,R.K. (2005) An artificial riboswitch for controlling pre-mRNA splicing. *RNA*, **11**, 1667–1677.

35. Wallis,M.G., von Ahsen,U., Schroeder,R. and Famulok,M. (1995) A novel RNA motif for neomycin recognition. *Chem. Biol.*, **2**, 543–552.

36. Carothers,J.M., Goler,J.A., Kapoor,Y., Lara,L. and Keasling,J.D. (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic Acids Res.*, **38**, 2736–2747.

37. Duchardt-Ferner,E., Weigand,J.E., Ohlenschläger,O., Schmidtke,S.R., Suess,B. and Wöhnert,J. (2010) Highly modular structure and ligand binding by conformational capture in a minimalistic riboswitch. *Angew. Chem. Int. Ed. Engl.*, **49**, 6216–6219.

38. Ferré-D'Amaré,A.R. and Doudna,J.A. (1996) Use of cis- and trans-ribozymes to remove 5′ and 3′ heterogeneities from milligrams of in vitro transcribed RNA. *Nucleic Acids Res.*, **24**, 977–978.

39. Svergun,D. (1992) Determination of the regularization parameter in indirect - transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.

40. Bergmann,A., Fritz,G. and Glatter,O. (2000) Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). *J. Appl. Crystallogr.*, **33**, 1212–1216.

41. Franke,D. and Svergun,D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.

42. Mueller,L., Legault,P. and Pardi,A. (1995) Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino protons. *J. Am. Chem. Soc.*, **117**, 11043–11048.

43. Dallmann,A., Simon,B., Duszczyk,M.M., Kooshapur,H., Pardi,A., Bermel,W. and Sattler,M. (2013) Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences. *Angew. Chem. Int. Ed. Engl.*, **52**, 10487–10490.

44. Ohlenschläger,O., Wöhnert,J., Bucci,E., Seitz,S., Häfner,S., Ramachandran,R., Zell,R. and Görlach,M. (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure*, **12**, 237–248.

45. Theimer,C.A., Finger,L.D., Trantirek,L. and Feigon,J. (2004) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 449–454.

46. Keller,R. (2004) Cara. The computer aided resonance assignment tutorial. Cantina Verlag, Goldau.

47. Duchardt-Ferner,E., Juen,M., Kreutz,C. and Wöhnert,J. (2017) NMR resonance assignments for the tetramethylrhodamine binding RNA aptamer 3 in complex with the ligand 5-carboxy-tetramethylrhodamine. *Biomol. NMR Assign.*, **11**, 29–34.

48. Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.

49. Nozinovic,S., Fürtig,B., Jonker,H.R.A., 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.

50. Rossi,P. and Harbison,G.S. (2001) Calculation of ^{13}C chemical shifts in RNA nucleosides: structure-^{13}C chemical shift relationships. *J. Magn. Reson.*, **151**, 1–8.

51. Duchardt-Ferner,E., Ferner,J. and Wöhnert,J. (2011) Rapid identification of noncanonical RNA structure elements by direct detection of OH···O = P, NH···O = P, and NH$_2$···O = P hydrogen bonds in solution NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **50**, 7927–7930.

52. Duchardt-Ferner,E. and Wöhnert,J. (2017) NMR experiments for the rapid identification of P = O···H-X type hydrogen bonds in nucleic acids. *J. Biomol. NMR*, **69**, 101–110.

53. Steinert,H., Rinnenthal,J. and Schwalbe,H. (2012) Individual basepair stability of DNA and RNA studied by NMR-detected solvent exchange. *Biophys. J.*, **102**, 2564–2574.

54. Noeske,J., Schwalbe,H. and Wöhnert,J. (2007) Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res.*, **35**, 5262–5273.

55. Duchardt-Ferner,E., Gottstein-Schmidtke,S.R., Weigand,J.E., Ohlenschläger,O., Wurm,J.P., Hammann,C., Suess,B. and Wöhnert,J. (2016) What a Difference an OH Makes: Conformational dynamics as the basis for the ligand specificity of the neomycin-sensing riboswitch. *Angew. Chem. Int. Ed. Engl.*, **55**, 1527–1530.

56. Dolgosheina,E.V. and Unrau,P.J. (2016) Fluorophore-binding RNA aptamers and their applications. *Wiley Interdiscip. Rev. RNA*, **7**, 843–851.

57. Strack,R.L. and Jaffrey,S.R. (2013) New approaches for sensing metabolites and proteins in live cells using RNA. *Curr. Opin. Chem. Biol.*, **17**, 651–655.

58. Truong,L. and Ferré-D'Amaré,A.R. (2019) From fluorescent proteins to fluorogenic RNAs: Tools for imaging cellular macromolecules. *Protein Sci.*, **28**, 1374–1386.

59. Paige,J.S., Wu,K.Y. and Jaffrey,S.R. (2011) RNA mimics of green fluorescent protein. *Science*, **333**, 642–646.

60. Dolgosheina,E.V., Jeng,S.C., Panchapakesan,S.S., Cojocaru,R., Chen,P.S., Wilson,P.D., Hawkins,N., Wiggins,P.A. and Unrau,P.J. (2014) RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.*, **9**, 2412–2420.

61. Babendure,J.R., Adams,S.R. and Tsien,R.Y. (2003) Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc.*, **125**, 14716–14717.

62. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.