Structural investigation of functional nucleic acids

Dissertation
for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program Biomolecules: Structure, Function, Dynamics of the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences

Submitted by
Mateusz Mieczkowski
born in Leżajsk, Poland

Göttingen, 2021

Thesis Committee

Prof. Dr. Vladimir Pena, Division of Structural Biology, The Institute of Cancer Research, London

Prof. Dr. Kai Tittmann, Molecular Enzymology, Georg August University, Göttingen

Prof. Dr. Claudia Höbartner, Organic and Biomolecular Chemistry, University of Würzburg

Members of the Examination Board

Referee: Prof. Dr. Vladimir Pena, Division of Structural Biology, The Institute of Cancer Research, London

2nd Referee: Prof. Dr. Kai Tittmann, Molecular Enzymology, Georg August University, Göttingen

Further members of the Examination Board

Prof. Dr. Claudia Höbartner, Organic and biomolecular chemistry, University of Würzburg

Prof. Dr. Markus Bohnsack, Molecular Biology, University Medical Center Göttingen.

Prof. Dr. Henning Urlaub, Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen

Dr. Alex Faesen, Biochemistry of Signal Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 26.03.2021

Affidavit

I hereby declare that this dissertation has been written independently and with no other sources and aids than quoted

……..

Mateusz Mieczkowski

Göttingen 2021

TABLE OF CONTENT

SUMMARY

DNA enzymes, also known as deoxyribozymes, are synthetic single-stranded DNA molecules able to catalyze chemical reactions. There are two main reasons for studying deoxyribozymes: their practical value in various applications, and the understanding of basic properties - such as folding and catalysis - of a biopolymer that is of central importance for life. Compared to ribozymes, the DNA enzymes have a potential value as tools for industrial or therapeutic applications, owing to more cost-effective synthesis and higher stability. The first crystal structure of a deoxyribozyme demonstrated that DNA possesses the intrinsic ability to adopt complex tertiary folds that support catalysis and unveiled the active site of a DNA enzyme in the post-catalytic state (Ponce-Salvatierra, Wawrzyniak-Turek et al. 2016). The second reported crystal structure of the RNA-cleaving deoxyribozyme complements observations about the folds and catalysis of DNA enzymes although the structure was derived with DNA as a substrate mimic of RNA (Liu, Yu et al. 2017). These crystal structures represent a breakthrough in the field, but they are still insufficient to derive a clear mechanistic picture of the specific features of different RNA ligating and RNA cleaving deoxyribozymes. Therefore, ongoing efforts are devoted to structurally investigating additional deoxyribozymes. The new DNA enzymes were evolved to discriminate modified and unmodified RNA substrates and provide attractive tools for studying the natural epitranscriptomic RNA modification N6-methyladenosine (Sednev, Mykhailiuk et al. 2018). In the present study, the goal is to elucidate the structural basis for recognition of the methylated nucleobase by solving the crystal structure of the m6A sensitive RNA-cleaving deoxyribozyme in complex with an uncleavable analog of the RNA substrate, containing either methylated and unmethylated adenosine. Surprisingly, the RNA substrate dissociated from the deoxyribozyme during the crystallization process. Two structures for unmethylated and one of the methylated RNA substrate analog were solved. The next goal is to elucidate the crystal structure of the RNA-ligating deoxyribozyme in the pre-catalytic state of reaction. The previously reported crystal structure of the 9DB1 in the post-catalytic state of reaction could not explain the role of magnesium cations as cofactors for accelerating RNA ligation and properly describe the ligation mechanism. The structural investigation of the 9DB1 in the pre-catalytic state resulted in the ligation of the two RNA substrates during the crystallization process. In the future, other strategies are necessary to solve the questions on substrate recognition and catalytic mechanism of the RNA-cleaving and RNA-ligating deoxyribozymes investigated in this study.

The second project deals with synthetic RNA aptamers that were identified by in vitro selection to mimic fluorescent proteins for RNA imaging and the development of biosensors. Several examples

of fluorogen-activating RNA aptamers are known, and for some, the crystal structures have recently been solved e.g. of the Spinach, Mango, and Corn aptamers, that bind synthetic analogs of the GFP chromophore (Neubacher and Hennig 2019). The Chili is a new fluorogenic-RNA aptamer that mimics large Stokes shift (LSS) fluorescent proteins (FPs) by inducing highly Stokes-shifted emission from several new green and red HBI (4-hydroxybenzylidene imidazolinone) derivatives that are non-fluorescent when free in solution (Steinmetzger, Palanisamy et al. 2019). The new fluorophores are the first variants of fluorogenic aptamer ligands with permanently cationic sidechains that are bound by the RNA in their protonated phenol form, while emission occurs from the phenolate intermediate after excited-state proton transfer. The Chili–DMHBO$^+$ complex is the longest wavelength-emitting (592 nm) and tightest binding (K_D=12 nM) RNA fluorophore currently known in the growing family of HBI-binding aptamers. By employing X-ray crystallography, I have elucidated the three-dimensional structure of the Chili fluorophore binding site and revealed the structural basis for the large apparent Stokes shift and the promiscuity of the Chili aptamer to activate red and green-emitting chromophores.

1 INTRODUCTION

Nucleic acids are versatile molecules. DNA is primarily known as a carrier of the blueprint of life and for its double helical structure, while RNA is mostly known as an intermediate in the protein translation process. In the 1950s James Watson and Francis Crick proposed a famous model of the DNA double helix (Watson and Crick 1953). From this time, the view on the properties of DNA and RNA has changed. Nucleic acids are not just storage place of the genetic information. DNA and RNA can fold into conformations supporting catalysis (ribozymes and deoxyribozymes) and small molecule ligand binding (aptamers and riboswitches). In nature, there are examples of RNA with important cellular functions (e.g. riboswitches) (Roth and Breaker 2009), however, many functional and catalytic nucleic acids have been identified using in vitro selection (so-called SELEX), with the functions and activities not found in nature (Jäschke and Seelig 2000).

The first examples of RNA catalysis were discovered in the early '80s (Kruger, Grabowski et al. 1982, Guerrier-Takada, Gardiner et al. 1983). This discovery allowed the identification of an extensive variety of self-cleaving ribozymes (Lilley 2011, Wilson, Liu et al. 2016) as well as RNA catalyst generated by in vitro selection techniques (Chen, Li et al. 2007, Martin, Unrau et al. 2015, Muller 2015).

For a long time, the role of the DNA was limited to the long-term storage of genetic information. This point of view was changed when the first RNA cleavage deoxyribozyme (so-called DNAzyme) was discovered using the in vitro selection (Breaker and Joyce 1994). Since this discovery, the potential of the catalytic DNA has been explored, and many DNAzymes were identified by the in vitro selection to catalyze the range of diverse chemical reactions (Silverman 2009, Hollenstein 2015, Silverman 2016). In 2016, the first crystal structure of the RNA-ligating 9DB1 DNAzyme was solved (Ponce-Salvatierra, Wawrzyniak-Turek et al. 2016) proving that the DNA adopts three-dimensional conformation supporting catalysis.

Besides of the catalytic role of the nucleic acid, DNA and RNA can fold into complex three-dimensional structures with specific ligand binding properties, so-called aptamers (Famulok and Mayer 2014, Pfeiffer and Mayer 2016), or with allosteric conformational properties called riboswitches (Breaker 2012, Serganov and Nudler 2013, Peselis and Serganov 2014). Nucleic acid aptamers are oligonucleotides that bind specific ligands, varying from small molecules to proteins. An aptamer for the specific ligand is commonly identified in vitro using SELEX. Riboswitches are found in bacteria, fungi, plants and they are used to sense ions and metabolites to regulate gene expression. To date, almost 40 different classes of riboswitches with their ligands have been

discovered and experimentally validated (McCown, Corbino et al. 2017, Pavlova, Kaloudas et al. 2019).

1.1 NUCLEOTIDES - BUILDING BLOCKS OF NUCLEIC ACIDS

Nucleic acids are biopolymers made of nucleotides. Nucleotides are composed of the nitrogenous base, the phosphate group, and five-carbon sugar (Figure 1). Chemically and structurally speaking, RNA and DNA are similar. Nucleotide polymer contains four building blocks – A, G, C, and T/U (T in DNA and U in RNA) and different pentose sugars – 2'-deoxyribose in DNA and ribose in RNA. Two following nucleotides are linked by the phosphodiester bond between the O5' of the first nucleotide with the O3' of the following nucleotide.

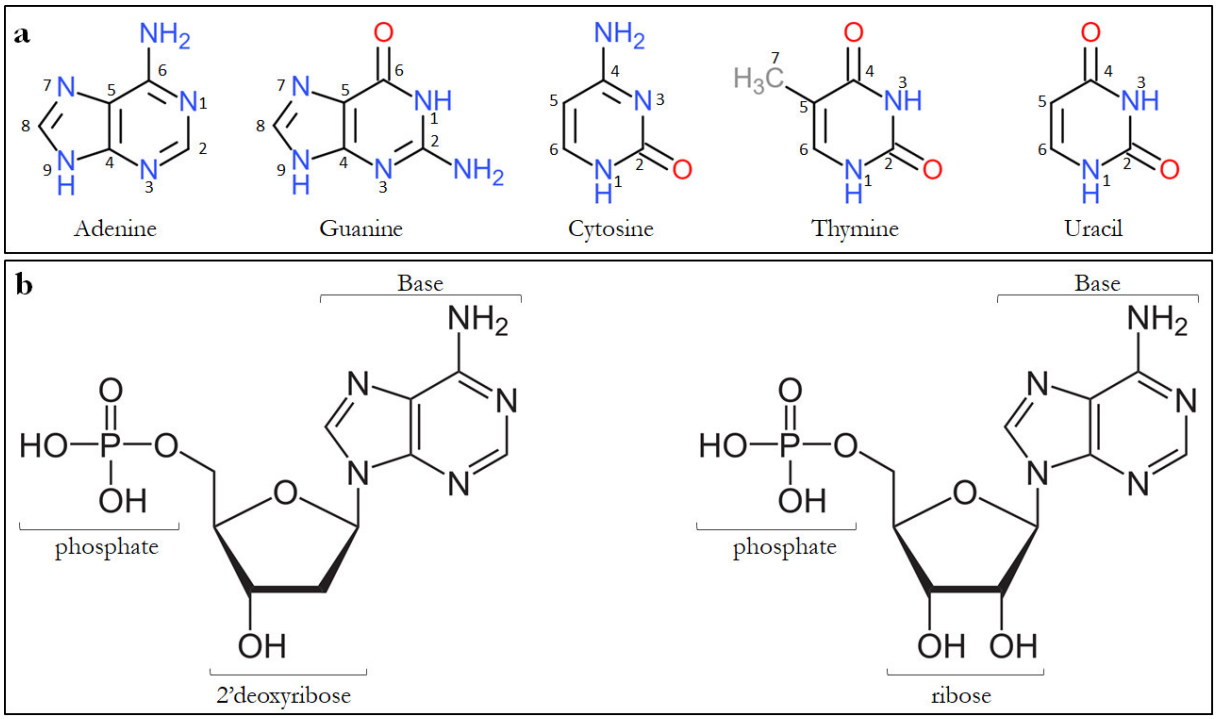


Figure 1. DNA and RNA building blocks.

(a) Nitrogenous bases used in DNA (A, G, C, T) and RNA (A, G, C, U) with corresponding atom numbers (b) Comparison of deoxyadenosine (left) and adenosine (right).

The pentose sugar is non-planar and differing conformations of the ribose have been named concerning those ring atoms which puckers out of the plane given by the other ring atoms. The most notable conformations are C2-*endo* for B-form DNA helices and C3'-endo for A-form RNA helices. The purine or pyrimidine base is connected to the ribose by the N-glycosidic bond. The torsion angle (chi), characterizes the relative base/sugar orientation and can be described as in Figure 2a. The torsion angle χ is correlated with the sugar pucker conformations. Tin the *anti*-conformation usually adopt a C3'-*endo* sugar pucker conformation, while nucleotides in *syn*

conformation prefer to adopt a C2'-*endo* pucker. The nucleotide pentose ring can be used to extract the phase angle P, using the given formula (Figure 2c). The phase angle P illustrates the lowest energy point and is defined by the orientation and location of substituents. If θ_2 is negative one should add 180° to the calculated value (Altona and Sundaralingam 1972).

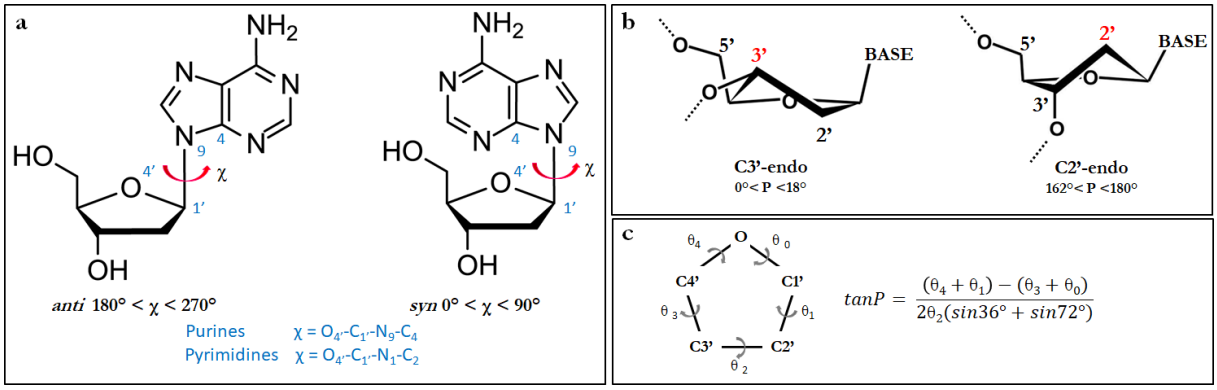


Figure 2. Geometric description of the glycosidic torsion angle and possible sugar puckers conformations.

(a) *anti* and *syn* conformations of the nucleotide and the description of the torsion angle χ (b) C3'-*endo* and C2'-*endo* sugar puckers. The C3'-*endo* conformation is typically observed in the RNA structures, while the C2'-*endo* is preferred in the DNA structures. (c) Relationship between the pseudorotation parameters and the phase angle of pseudorotation P.

1.2 CANONICAL BASE PAIRS

The B-form described by Watson and Crick is perhaps the most iconic of all biological molecules and is the most common form in the human genome. In 1980, the first crystal structure of the B-DNA was solved (Wing, Drew et al. 1980). In the presented model, the two strands are held together by the hydrogen bonds between purines and pyrimidines - guanine pairs with the cytidine and adenosine pairs with thymidine (in DNA) or with uridine (in RNA). This hydrogen bonding pattern is known as a canonical base pairing (so-called Watson-Crick base pairing) and it is shown in Figure 3.

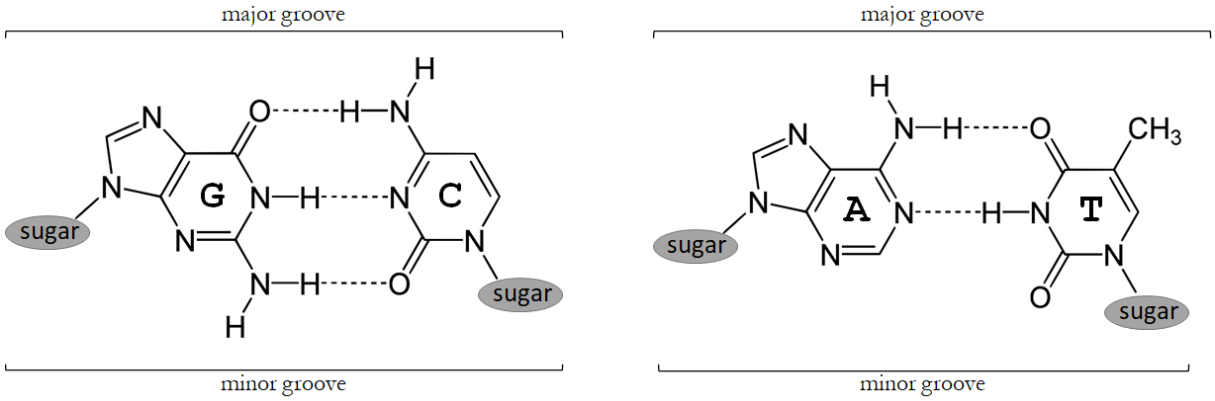


Figure 3. Canonical Watson-Crick base pairing pattern and definition of major and minor grooves.

1.3 NON-CANONICAL BASE PAIRS

Except for canonical base pairs, various interactions between nucleobases are possible. In the canonical base pairs, the Watson-Crick (WC) edge of the purine is forming hydrogen bonds with the Watson-Crick edge of the pyrimidine (Figure 3). The pyrimidine WC edge can form hydrogen bonds with the so-called Hoogsteen edge of purine forming Hoogsteen base pair (Figure 4a,b). Base pairs can be also formed via interactions between sugar edges (Figure 4e) The Hoogsteen base pairs, provide additional interaction patterns, enabling the formation of base triplets (Figure 4c,d) and quartets (Figure 4f). Besides the three interaction edges, base pairs can be also formed in various cis/trans forms. Three hydrogen bonding interaction edges together with cis/trans conformations form twelve types of base pairing geometries observed in the RNA structures (Leontis and Westhof 2001, Leontis, Stombaugh et al. 2002).

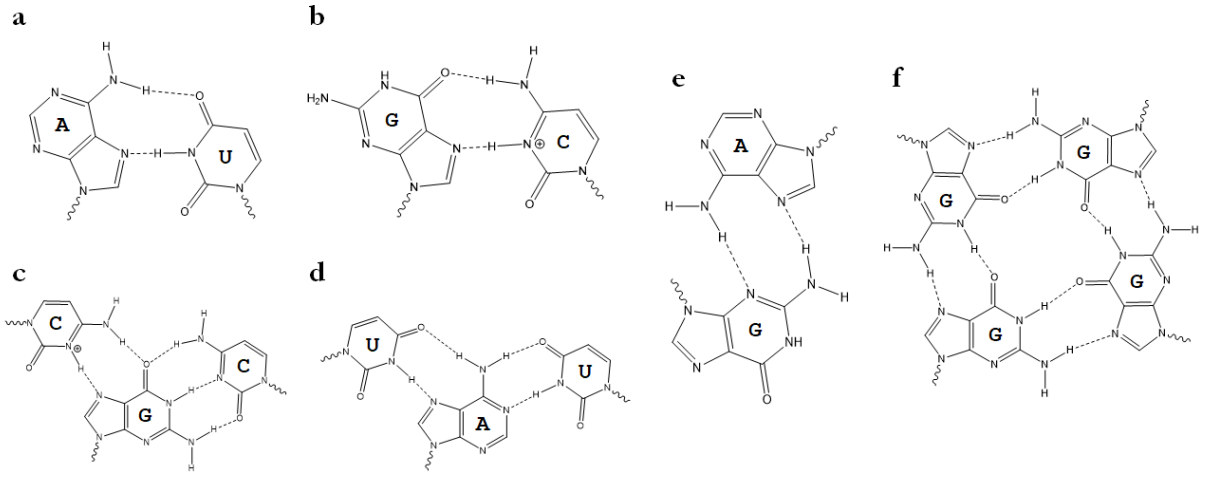


Figure 4. Selected examples of non-canonical base-pairing interactions.

(a,b) Hoogsteen base pairs (c,d) Base triplets (e) Trans Hoogsteen/Sugar edge hydrogen bond interactions (f) The G-quartet.

H-DNA (also called triple-stranded DNA) is an example of a molecule containing nucleotide triplets (Figure 5). In the H-DNA, the DNA oligonucleotide binds to the Hoogsteen edge of adenosine and guanosine in the B-form DNA duplex. These interactions allow the formation of two types of base triplets – T:A:T and C:G:C^+. The WC edge of thymidine and protonated cytidine are forming hydrogen bonds with the Hoogsteen edge of adenine and guanosine respectively (Figure 5). The C:G:C^+ triplet is formed at low pH and dissociate at high pH since the N3 atom of the cytosine has to be protonated (Medhi 2002).

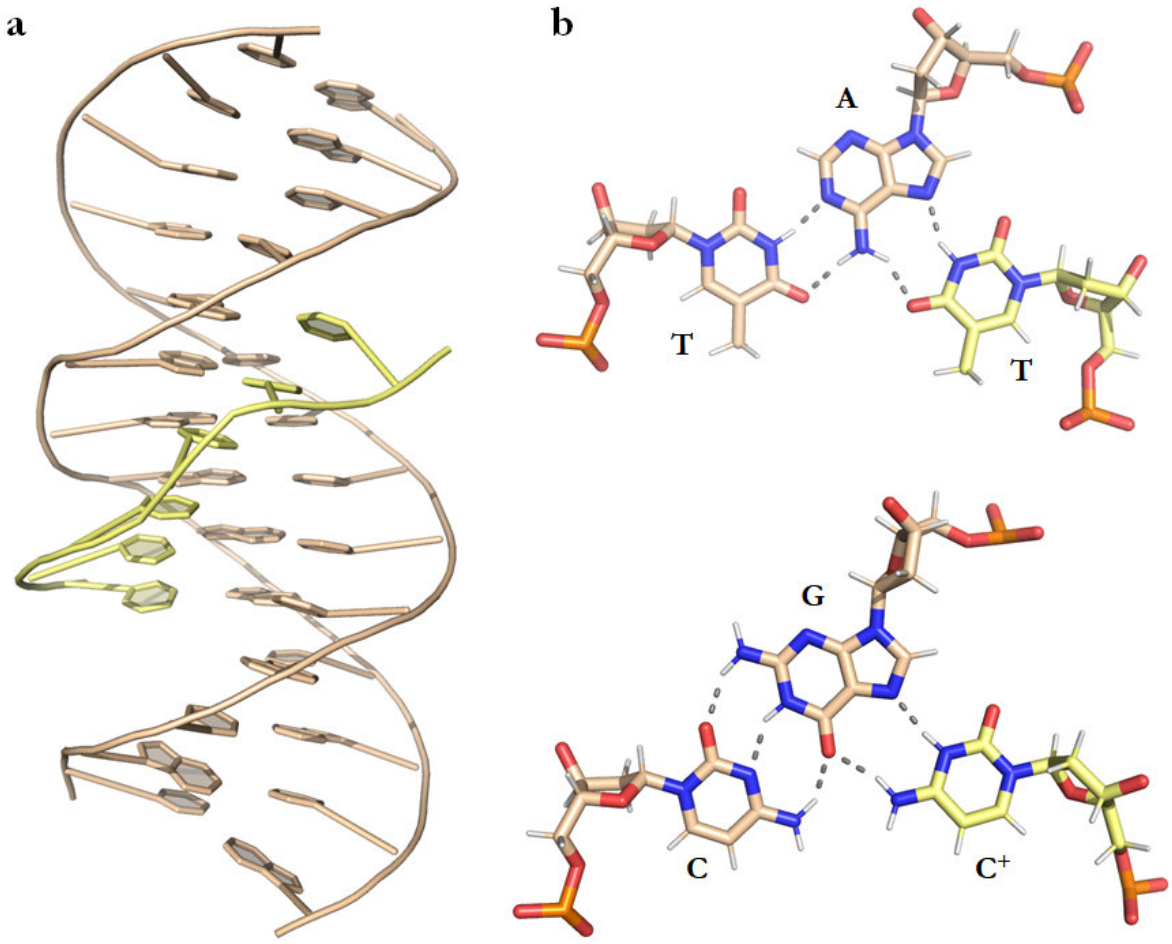


Figure 5. Example of the H-DNA (PDB: 1BWG).

(a) Overall structure of the triple-stranded DNA. B-form duplex is colored in beige and the invading DNA strand is colored in yellow. (b) DNA triplets in the H-DNA. The invading strand binds to the Hoogsteen edge of the B-DNA. Polar contacts are represented as dashed grey lines.

1.3.1 G-quadruplex structure

Multiplets are not limited to triplet formation. In 1910 it was observed that guanine alone in millimolar concentrations can form a gel in an aqueous solution (Bang 1910). This was not proven until 1962 when the involved structure was determined. The G-rich sequences may form a G-quartet, composed of four guanines that are forming hydrogen bonds in a circle via Hoogsteen base pairing (Gellert, Lipsett et al. 1962). In the presence of the monovalent cations, G-quartets can stack on top of each other forming a G-quadruplex (Figure 6) (Largy, Mergny et al. 2016). The phosphate backbones run along the outside four edges of the structure and metal cations occupying the central axis, stabilize G-quadruplex.

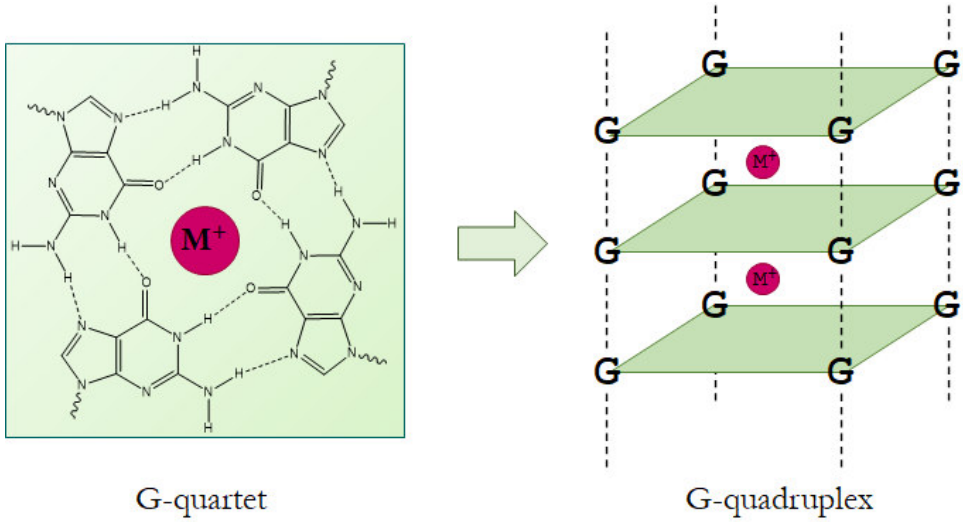


Figure 6. G-quartets and G-quadruplex.

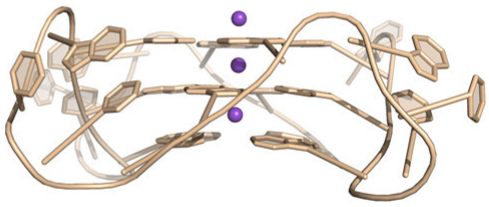
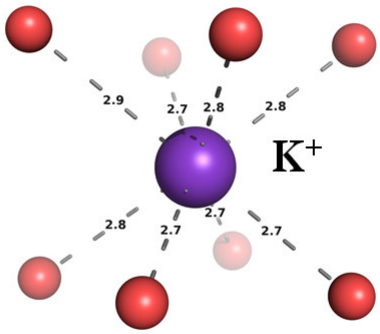
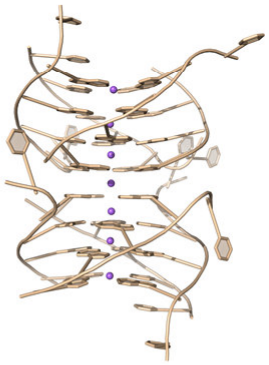
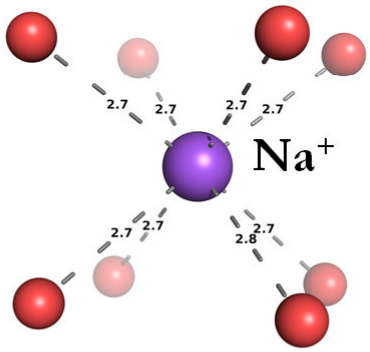
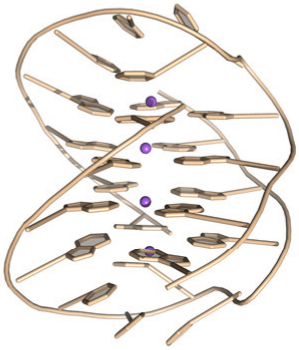
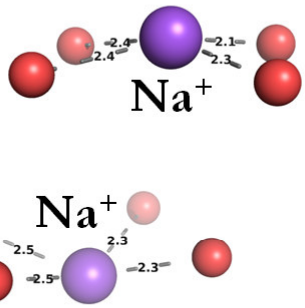
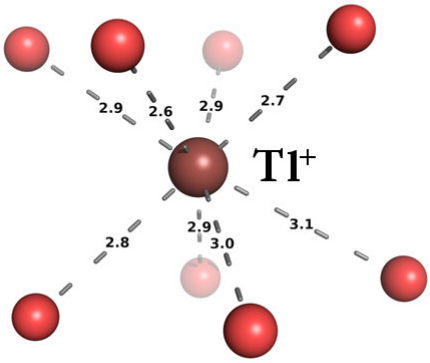
G-quartets (left) in the presence of monovalent cations can stack on top of each other to form a G-quadruplex structure (right) composed of four independent strands.

1.3.1.1 Metal coordination by the G-quadruplex

Stacking of the G-quartets on the top of each other forms a channel possessing a negative charge caused by the partial charge of the O6 of the guanine (Sponer, Leszczynski et al. 2001). Positively charged ions interact with the guanosine O6 of and stabilize the negatively charged channel. The different monovalent cations stabilize G-quadruplex with different capabilities. The following affinity of monovalent cations for a stacking G-quartet structure formed by guanosine 5'-monophosphate was indicated - $K^+ > NH_4^+ > Rb^+ > Na^+ > Li^+$ (Wong and Wu 2003). The Tl^+ cation was also shown to bind to the G-quadruplex (Gill, Strobel et al. 2005, Gill, Strobel et al. 2006). Among the divalent cations, Ba^{2+} (Zhang, Huang et al. 2014) Sr^{2+}(Fyfe, Dunten et al. 2015), and Pb^{2+} (Liu, Zheng et al. 2012) were found to stabilize G-quadruplex. The Ca^{2+} cations, together with Na^+ were also observed to stabilize Q-quadruplex (Lee, Parkinson et al. 2007, Kwan, Wong

et al. 2008). The other divalent metal ions as Mg^{2+}, Co^{2+}, Mn^{2+}, Cd^{2+} Ni^{2+}, and Zn^{2+} are presumed to induce G-quadruplex dissociation by entering transiently opened G-quartet hydrogen bonds then binding to N7 and O6 of the guanine, thus preventing the formation of Watson-Crick/Hoogsteen base-pairing interaction (Hardin, Perry et al. 2000). According to this study, K^+, Na^+, $NH4^+$, $Sr^{2+,}$ and Pb^{2+} are the most effective stabilizers. The selected G-quadruplex structures crystallized with different cations are shown in Figure 7. Na^+ can be coordinated in the plane of the G-quartet or between two consecutive tetrads with square antiprismatic geometry. Because of the size, other metals (K^+, Tl^+, Ca^{2+}, Sr^{2+}, Ba^{2+}, etc.) can be coordinated only in the antiprismatic geometry.

a

PDB	Side view	Cation coordination site
1KF1		
352D		
1JB7		
2HBN		

b

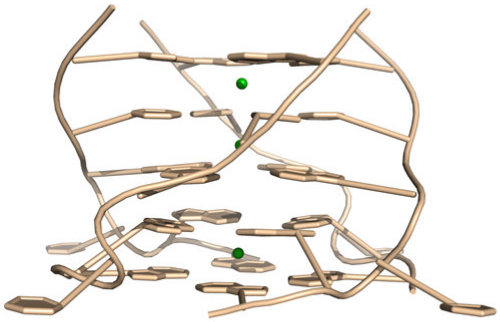
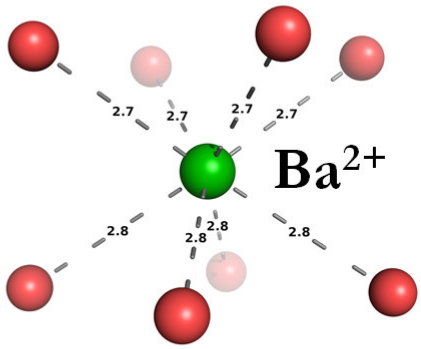
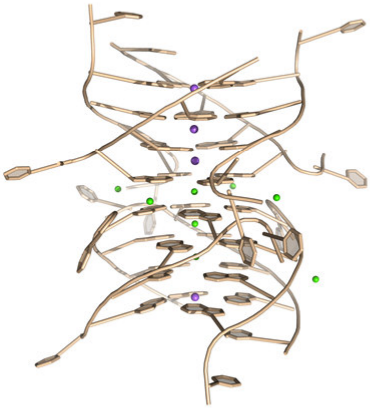
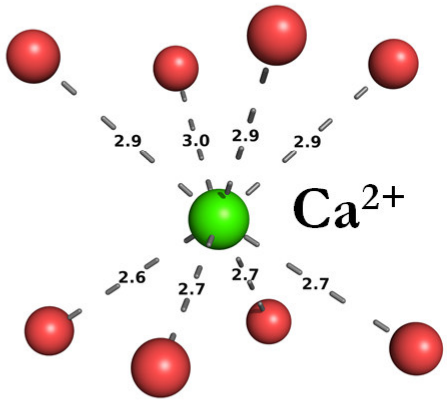
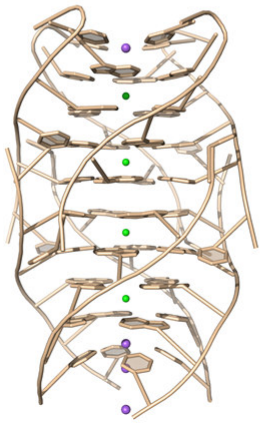
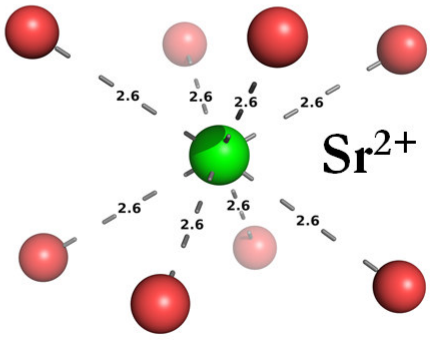
PDB	Side view	Cation coordination site
4U92		
2GW0		
4RKV		

Figure 7. Crystal structures of G-quadruplexes stabilized by different (a) monovalent and (b) divalent cations.

K^+, Na^+, and Tl^+ cations are shown as purple, violet, and brown respectively. Ba^{2+}, Ca^{2+}, and Sr^{2+}cations are colored in different shades of green. The guanine O6 atoms are colored as red spheres. Coordination of the monovalent and divalent metal cations is represented as dashed grey lines.

1.3.1.2 Stacking of the G-quartets

The consecutive G-quartets can possess different stacking geometries. G-quartets polarity is defined by the direction of the Hoogsteen hydrogen bond interactions. G-quartets stacking is described by the opposite and same polarity stacking (Figure 8). The consecutive stacking can be also described according to guanine stacking within the G-quadruplex core. The most common stacking geometries are shown in Figure 9 (Lech, Heddi et al. 2013).

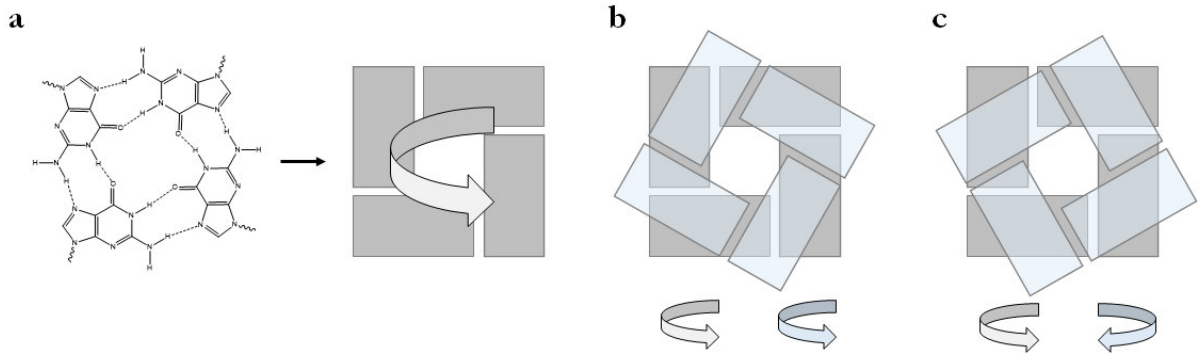


Figure 8. G-quartet stacking.

(a) G-quartet polarity is determined by the hydrogen bond donor/acceptor pattern. Same-polarity (b) and opposite-polarity (c) pattern of the G-tetrads.

The guanine base stacking in the G-quadruplex is intrinsically related to the glycosidic bond conformations: partial 5/6-ring stacking (Figure 9a), where the same-polarity consecutive guanine residues are in *'anti/anti'* conformation and result in partial overlap of the 5-carbon ring of the guanine with the 6-carbon ring of another, 5-ring stacking (Figure 9b), with the *'syn/anti'* step results in opposite-polarity stacking with an overlap of the 5-carbon rings of stacked guanines, and partial 6-ring stacking (Figure 9c), where *'anti/syn'* step forms opposite-polarity stacked G-quartets and results in a partial overlap of the 6-carbon rings of stacked guanines (Lech, Heddi et al. 2013).

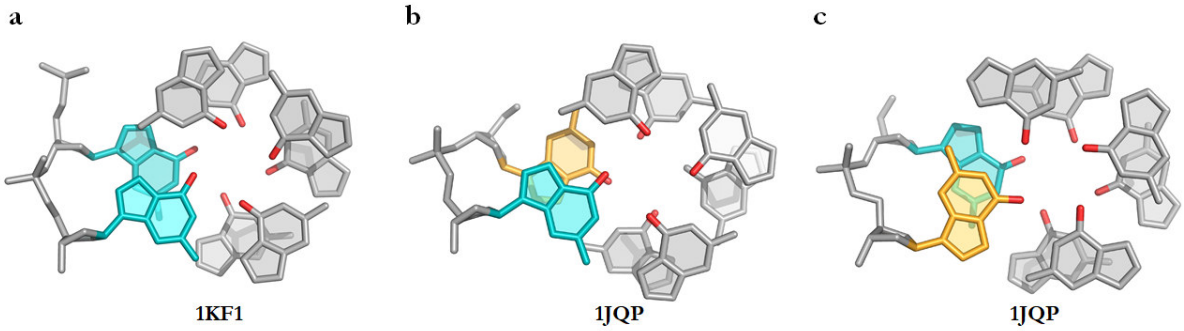


Figure 9. Examples of most common G-quartets stacking geometries with corresponding PDB codes.

(a) Partial 5/6-ring stacking (b) 5-ring stacking (c) Partial 6-ring stacking. The *anti* and *syn* conformations forming particular steps are colored in teal and orange respectively.

1.3.1.3 Polymorphism of the G-quadruplex

G-quadruplexes can be formed intramolecularly between one strand or intermolecularly between two or four strands (tetramolecular) or two strands of DNA or RNA (Figure 10a). They can form various combinations of strand directions as well as length and loop composition (Burge, Parkinson et al. 2006). The strands in the G-quadruplex can have the same orientation (all parallel), two strands in one orientation and one in opposed direction (2+2 anti-parallel) or mixed, with three strands in one orientation and one in opposing direction (3+1 hybrid) (Figure 10b). G-quadruplexes with at least one of the four strands opposite to the others is called *anti*-parallel G-quadruplex. In the all parallel G-quadruplexes, the guanines are in an *anti*-conformation, while in the *anti*-parallel quadruplexes, both *syn* and *anti* conformations are observed, arranged specifically for a given topology and strand orientations.

The loops connecting guanines within the G-quadruplex can adopt four different geometries: propeller type loops (also called strand reversal loops), lateral type loops (also called edgewise loops), diagonal loops (Cang, Šponer et al. 2011), and V-shaped loops (Kuryavyi and Patel 2010). The loops are indicated in different colors on the selected G-quadruplexes schemes (Figure 10b).

The propeller type loop connects two guanine bases belonging to bottom and top G-quartets in adjacent, parallel strands. The lateral type loop connects guanosine residues hydrogen bonding with each other and belonging to antiparallel strands in the same G-quartet, while the diagonal loop connects opposite guanosines within the same G-quartet, belonging to non-adjacent, anti-parallel strands. The last type of loop is called V-shape and it was observed for the first time in 2001

(Zhang, Gorin et al. 2001). The V-shape loop connects two corners of quartets within the same G-quadruplex lacking one supporting column.

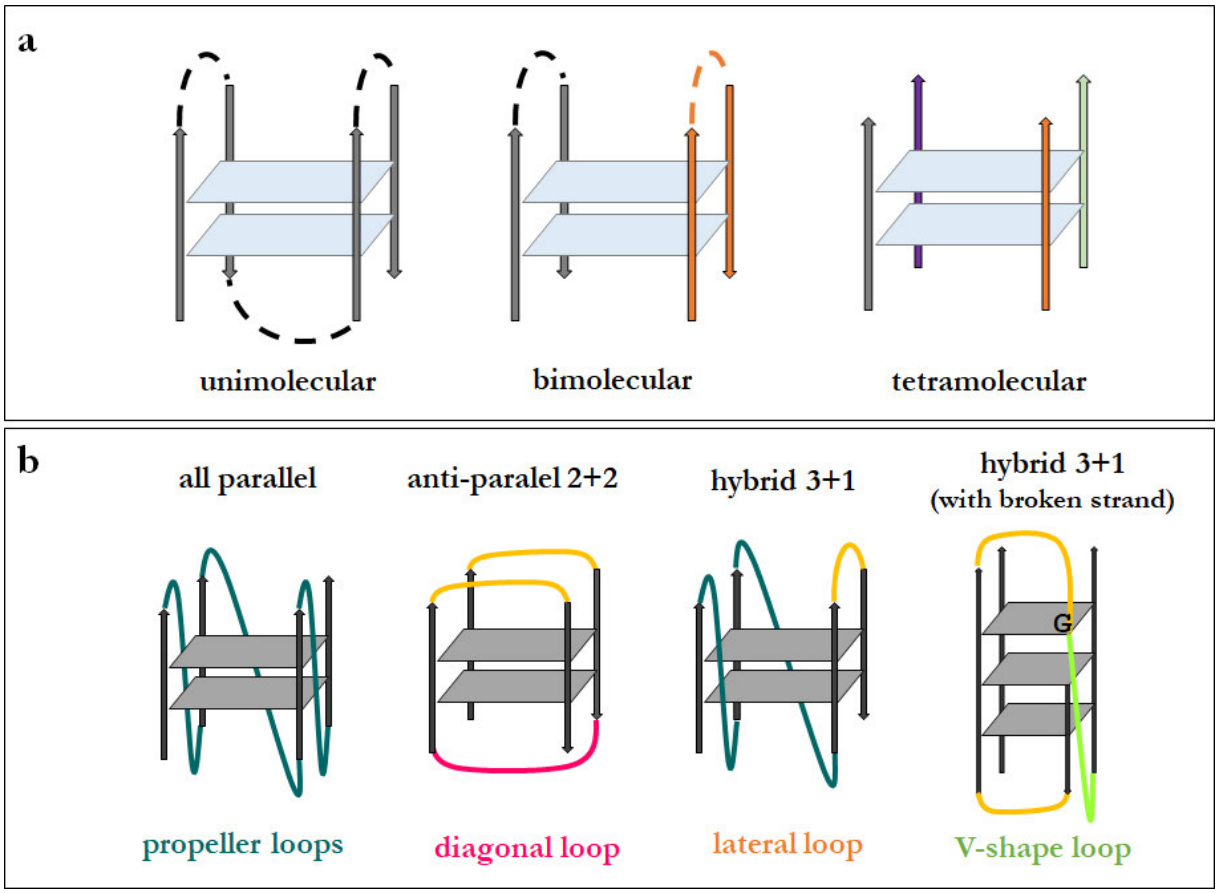


Figure 10. Structural elements defining G-quadruplex polymorphism.

a) Intra- and inter-molecular forms of the G-quadruplexes. b) Strands orientation and loops geometry - a propeller loop is highlighted in teal, a diagonal loop is highlighted in pink, a lateral loop is highlighted in orange and a V-shaped loop is highlighted in green.

1.4 SYSTEMATIC EVOLUTION OF LIGANDS BY EXPONENTIAL ENRICHMENT (SELEX)

In 1990, Craig and Gold have selected RNA binding to the bacteriophage T4 DNA polymerase (Tuerk and Gold 1990). They named this process SELEX (Systematic Evolution of Ligands by EXponential enrichment) and the resulting molecules were named aptamers (Ellington and Szostak 1990). The SELEX is either called in vitro selection or in vitro evolution process. Except for the aptamer selection, SELEX allows identifying novel ribozymes, deoxyribozymes, or aptazymes (Breaker and Joyce 2014).

The in vitro selection incorporates multiple rounds of selection and enriching target molecules in each round. The general overview of the aptamer SELEX process is shown in Figure 11. For a

selection, a large pool of synthetic oligonucleotides is used (~10^{15}). The pool is flanked by the fixed-length primer sequences and the random region of 30 – 80 nucleotides. For a random region composed of N number of nucleotides, the number of possible sequences in the library is 4 to the power of N (4^N; N positions with four possibilities – A, G, C, T for each position). During the first step of SELEX (Figure 11), the DNA or RNA random pool is incubated with the immobilized target. Next, the non/low-affinity binders are washed away from oligonucleotides binding to the target. The binders are recovered by elution, amplified by a polymerase chain reaction (PCR), and introduced to the new selection round. If the RNA pool was used for the selection, then the recovered pool is transcribed into the DNA. The in vitro selection process is carried until the bound sequences stop enriching in the library. Several in vitro selection methods were developed to obtain efficient enrichment of the binding sequences, including target immobilized magnetic beads (Stoltenburg, Reinemann et al. 2005), capillary electrophoresis (Mendonsa and Bowser 2004), affinity chromatography (Wilson and Szostak 1995), or gel-based (Liu, Zhang et al. 2020). After around eight to thirteen selection round, sequences with the best binding properties are cloned and sequenced. It can be possible that a particular sequence is dominating if the final pool, however, several different sequences with similar fold are usually obtained (Kato, Takemura et al. 2000, Bing, Yang et al. 2010). After sequencing, aptamers can be further characterized and shortened to get a minimal sequence required for the target binding.

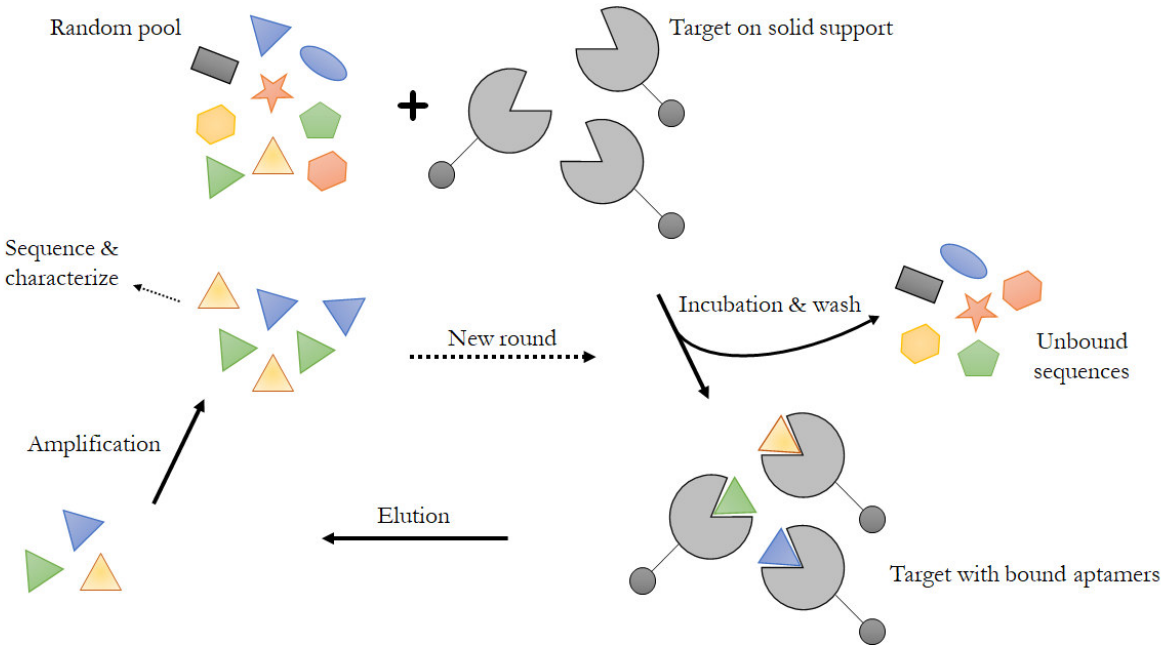


Figure 11. Conventional SELEX.

A library of random DNA or RNA fragments are incubated with the immobilized target. The unbound sequences are separated and the bound sequences are recovered and amplified by the PCR (in case of the RNA - reverse transcription before amplification by the PCR). The selection and amplification process can

be repeated multiple times to obtain better binders. After the selection process, DNA/RNA molecules are sequenced and further characterized.

Aside from aptamers, catalytic nucleic acids like ribozymes and deoxyribozymes were evolved (Jaschke 2001). There is the main difference between the selection of catalytic nucleic acid and aptamers. During catalytic nucleic acid selection, instead of finding molecules that bind with the target, the molecules catalyzing the reaction can be found (Joyce 1994, Wilson and Szostak 1999).

1.5 FUNCTIONAL NUCLEIC ACIDS

As mentioned in previous chapters, functional nucleic acids are described as aptamers and catalytic DNA/RNA. DNA enzymes have a potential value as tools in the industry (Liu and Lu 2004, Wei, Li et al. 2008, Gong, Zhao et al. 2015) or in the therapeutic applications (Isaka 2007, Fokina, Stetsenko et al. 2015), owing to more cost-effective synthesis and the higher stability. DNAzymes can catalyze a wide range of chemical reactions like RNA cleavage (Breaker and Joyce 1994, Geyer and Sen 1997, Santoro and Joyce 1997), RNA 3'-5' ligation (Hoadley, Purtha et al. 2005, Purtha, Coppins et al. 2005); RNA ligation with a branch (Wang and Silverman 2003, Wang and Silverman 2003, Coppins and Silverman 2004, Coppins and Silverman 2005) and lariat formation (Pratico, Wang et al. 2005, Wang and Silverman 2005), DNA phosphorylation (Wang, Billen et al. 2002), adenylation (Li, Liu et al. 2000), DNA ligation (Cuenoud and Szostak 1995, Sreedhara, Li et al. 2004) and many others (Li and Sen 1996, Burmeister, von Kiedrowski et al. 1997, Sheppard, Ordoukhanian et al. 2000, Chinnapen and Sen 2004, Höbartner, Pradeepkumar et al. 2007, Chandra and Silverman 2008). In the present thesis, the main focus is set on fluorogen-activating RNA aptamers and RNA-cleaving and RNA-ligating DNAzymes.

1.5.1 Catalytic nucleic acids

1.5.1.1 RNA-cleaving deoxyribozymes

The most characteristic and well-studied RNA cleaving DNAzymes are the 10-23 and 8-17 (Santoro and Joyce 1997). Both deoxyribozymes catalyze multiple turnover RNA-cleavage by transesterification.

1.5.1.1.1 10-23 DNAzyme

The 10-23 deoxyribozyme cleaves the RNA substrate between a purine and a pyrimidine residue (Figure 12) in the presence of divalent metal cations, such as Mg^{2+}, $Ca^{2+,}$ or Mn^{2+}. The divalent metal cations act as a general base or as an acid either to stabilize the transition state of reaction and/or support folding into the active conformation (Santoro and Joyce 1998). The catalytic core

of 10-23 deoxyribozyme consists of 15 nucleotides flanked with sequence-specific binding arms. The binding arms of the 10-23, flanking the catalytic core region, can be designed to cleave any RNA, as long as the R|Y junction requirement is fulfilled (Figure 12a). 10-23 deoxyribozyme cleaves RNA between the unpaired purine (adenosine/guanosine) and paired pyrimidine (cytidine/uridine) and forms two products: one with the 2',3'-cyclic phosphate at the 3'-end and the second with the hydroxyl group at the 5'-end (Figure 12b). In 1999, the crystal structure of the 10-23 with the uncleavable substrate was solved. However, the structure revealed catalytically irrelevant conformation of a DNAzyme, that reassembled a Holiday junction structure (Nowakowski, Shim et al. 1999). To date, no crystal structure of the 10-23 deoxyribozyme is an active conformation is known.

As a result, of the ability to cleave RNA efficiently in a sequence-specific manner, 10-23 DNAzyme has grabbed attention as a potential therapeutic agent. The 10-23 acts as an antiviral agent against HIV (Dash, Harikrishnan et al. 1998, Goila and Banerjea 1998, Zhang, Xu et al. 1999, Basu, Sriram et al. 2000), hepatitis type B and C (Goila and Banerjea 2001, Hou, Wo et al. 2005, Wo, Wu et al. 2005, Hou, Ni et al. 2006), influenza type A and B (Toyoda, Imamura et al. 2000, Kumar, Khanna et al. 2012, Kumar, Kumar et al. 2013), and many others (Zhang, Dass et al. 2004, Xie, Zhao et al. 2006, Appaiahgari and Vrati 2007, Wu, Xu et al. 2007). 10-23 also acts as a potential antibacterial agent (Chen, Li et al. 2004, Li, Zhu et al. 2005, Hou, Meng et al. 2007, Hou, Meng et al. 2007).

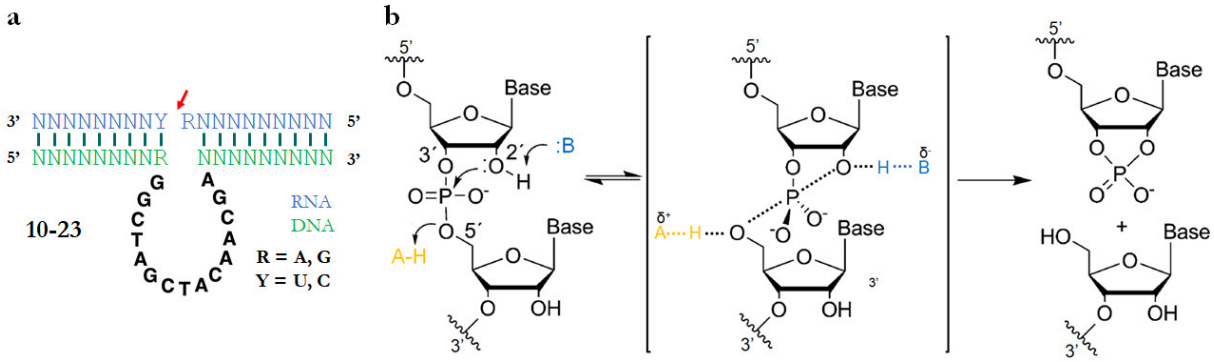


Figure 12. The RNA-cleaving 10-23 DNAzyme.

(a) Secondary structure of the 10-23 deoxyribozyme. The 10-23 DNAzyme binds to its target RNA through Watson–Crick base pairing (depicted as teal lines) and catalyzes cleavage between an unpaired purine (R) and a paired pyrimidine (Y). The red arrow indicates the cleavage site. (b) Mechanism of general acid-base cleavage of an RNA phosphodiester bond. The cleavage is initiated by an in-line nucleophilic attack of the neighboring 2'-oxygen on the adjacent 3'-phosphodiester bond. The general base (in blue) attracts a proton from the 2'-hydroxyl group. The general acid (in orange) is donating a proton to the 5'-oxyanion leaving group. The products of the reaction are the cyclic 2',3'-phosphate, and 5'-hydroxyl leaving group.

1.5.1.1.2 8-17 DNAzyme

One of the most characterized RNA-cleaving DNAzyme is 8-17 (Figure 13a) (Santoro and Joyce 1997). Variants of 8-17 motif have been selected independently many times (Schlosser and Li 2009). The catalytic core of 8-17 DNAzyme consists of 13 nucleotides in length and it is flanked with RNA substrate binding arms. The 8-17 DNAzymes require divalent metal ions for the activity, such as Mg^{2+}, Mn^{2+}, $Zn^{2+,}$ or Pb^{2+} (Faulhammer and Famulok 1997, Li, Zheng et al. 2000, Cruz, Withers et al. 2004). RNA cleavage catalyzed by 8-17 deoxyribozyme in the presence of Mg^{2+} and Zn^{2+} occurs via DNA folding followed by catalysis, while in the presence of Pb^{2+} reactions occurs without folding step, describing a fast rate of the Pb^{2+} assisted cleavage rate (Kim, Rasnik et al. 2007).

In 2017, the first crystal structure of the RNA-cleaving deoxyribozyme was revealed (Liu, Yu et al. 2017). The 8-17 DNAzyme in the complex with non-cleavable substrate analog was co-crystallized with African swine fever virus DNA polymerase X (Figure 13b). The protein-induced crystallization allowed to obtain high-quality crystals, forming only marginal interactions with DNAzyme.

The structure was solved in the presence of partially occupied Pb^{2+} cation (occupancy 0.4), which is responsible for the highest activity of the 8-17. The ion is placed in a cavity between the cleavage site and the catalytic core. The Pb^{2+} is in only one direct contact with O6 of G6 of the DNAzyme (Figure 13c). The longer distances to the bridging phosphate atoms (around 5 Å) may suggest that bound Pb^{2+} may be largely hydrated (Figure 13c). The crystal structure revealed that RNA is cleaved via a general acid-base mechanism, similarly to naturally evolved ribozyme counterparts (Liu, Yu et al. 2017).

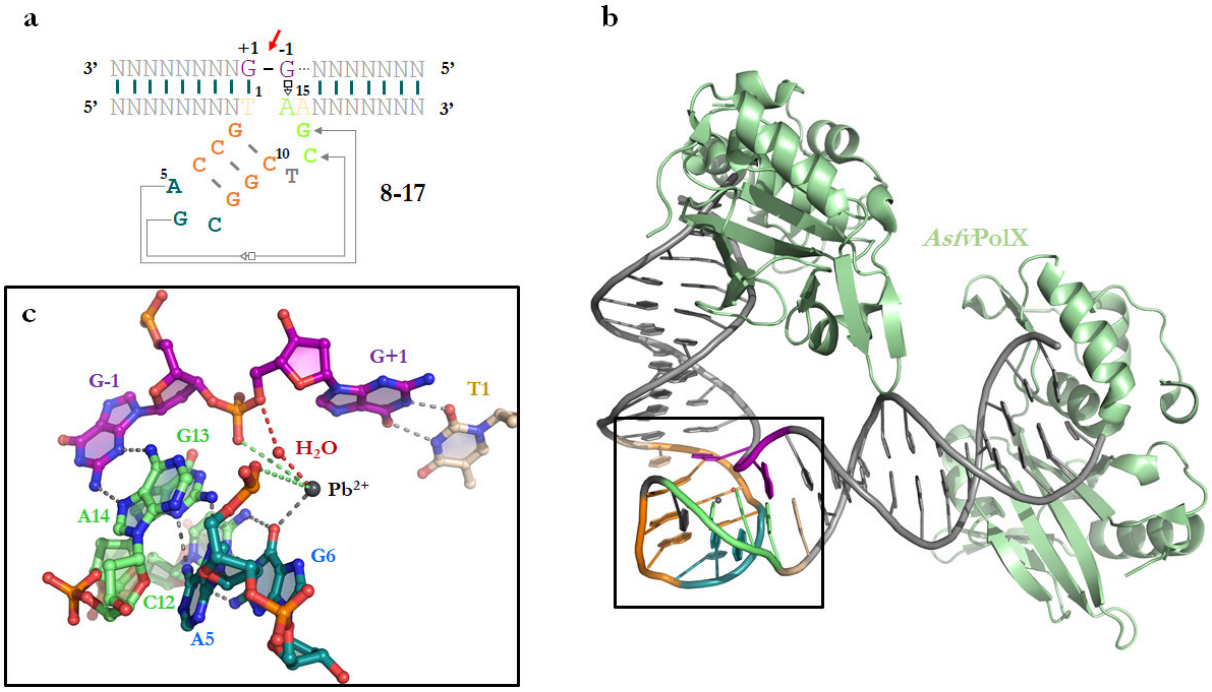


Figure 13. The 8-17 DNAzyme.

(a) Overall 2D structure of the 8-17 deoxyribozyme. (b) Overall crystal structure of the 8-17 DNAzyme in complex with Pb^{2+}. (c) Close up view for the active site. GG kink is colored purple. In the proposed mechanism, G13 deprotonates the 2'-hydroxyl group (missing in structure) of the attacking G−1 residue. The G6 coordinates the Pb^{2+}, which will activate the catalytic water molecule (general acid) to provide a proton to the O5' atom of the leaving residue G+1 (Liu, Yu et al. 2017). The Pb^{2+} and water molecules are shown as black and red spheres respectively. The red arrow indicates the cleavage site.

1.5.1.1.3 m6A sensitive RNA-cleaving deoxyribozymes

RNA-cleaving DNA enzymes were also evolved to recognize chemical modification on the RNA substrate. The first study involving RNA- cleaving DNAzyme to detect chemical modification was to detect 2'-O-ribose methylations in the rRNA of the *Saccharomyces cerevisiae*. The 10-23 type DNAzyme cleaves at position A(973) in the 18S rRNA or position G(1450) in the 25S rRNA from S. cerevisiae, only if the 2'-O-Me modifications at those positions were not present (Buchhaupt, Peifer et al. 2007).

The first RNA-cleaving DNAzymes that distinct modified and unmodified RNA substrate were evolved in 2018 (Sednev, Mykhailiuk et al. 2018). These deoxyribozymes are sensitive to the presence of N6-methyladenosine in RNA near the cleavage site (Figure 14). VMA8 and VMA15 DNAzymes exhibit faster cleavage of methylated RNA, while the VMC10 DNAzyme efficiently cleaves unmethylated RNA while the methylated RNA substrate remains intact.

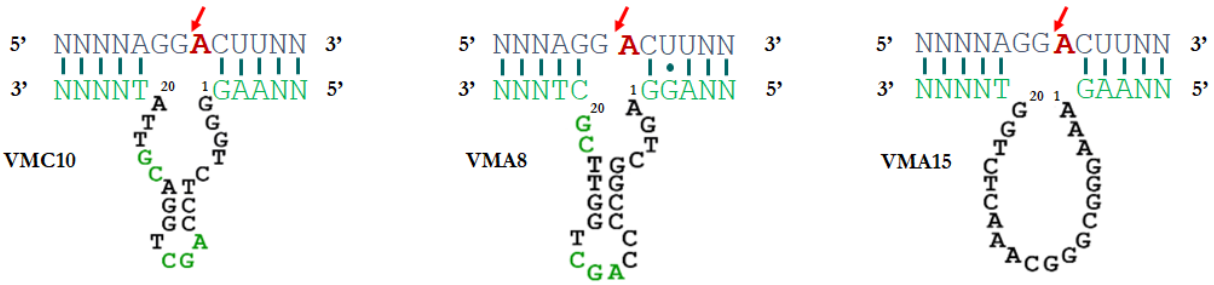


Figure 14. N^6-methylademosine sensitive RNA-cleaving deoxyribozymes.

VMA8 and VMA15 cleave m6A modified substrate faster than unmodified, while VMC10 is inhibited by m6A. The red arrow indicates the cleavage site.

The VMA and VMC cleaves nucleotides within the GG(m6A/A)CU motif which is the most frequent representative of the DRACH (D – A or G or T; R – A or G; H – A or C or T) motif in mRNAs and lncRNAs in higher eucaryotes (Liu, Parisien et al. 2013, Nachtergaele and He 2018). Such deoxyribozymes can be used to validate predicted m6A sites in isolated cellular RNA (Bujnowska, Zhang et al. 2020). To date, it is still unknown how VMA and VMC deoxyribozymes catalytic region folds and/or use magnesium cations to discriminate methylated from unmodified RNA.

1.5.1.2 9DB1 - RNA-ligating deoxyribozyme

Over the years, most attention has been paid to the RNA-cleaving deoxyribozymes. The 9DB1 deoxyribozyme catalyzes the formation of a native 3'-5'-phosphodiester bond between two RNA substrates in the presence of the Mg^{2+} or Mn^{2+}(Purtha, Coppins et al. 2005, Wachowius, Javadi-Zarnaghi et al. 2010). The catalytic core contains 40 nucleotides (parent) or 31 nucleotides (miniaturized) in length and binds their RNA substrates employing Watson–Crick base pairing. In 2016, the crystal structure of 9DB1 was reported (Ponce-Salvatierra, Wawrzyniak-Turek et al. 2016) as the first three-dimensional structure of a deoxyribozyme in an active conformation (Figure 15a,b). The structure mimics the post-catalytic state of ligation. The crystal structure was solved using 44 nucleotide long 9DB1 deoxyribozyme (of which 31 nucleotides from the miniaturized core) bound to 15 nucleotide RNA product. The structure exhibits the shape of letter λ which comprises two RNA/DNA helical duplexes forming an angle of 120° to each other and both lying above and tightly attached to the catalytic core. The catalytic domain consists of two stems (formed by two and four base pairs respectively) and two thymidine residues (T29 and T30) that interacts directly with A-1 and G1 nucleotides in the ligation junction (Figure 15a,c). The crystal structure contains three Mg^{2+} cations of which one is near the ligation junction. However, the structure could not explain the role of magnesium cations as cofactors for accelerating RNA ligation. The non-

bridging *pro*-Sp phosphate oxygen of dA13 is crucial in activating the 3'-hydroxy group to promote RNA ligation (Ponce-Salvatierra, Wawrzyniak-Turek et al. 2016).

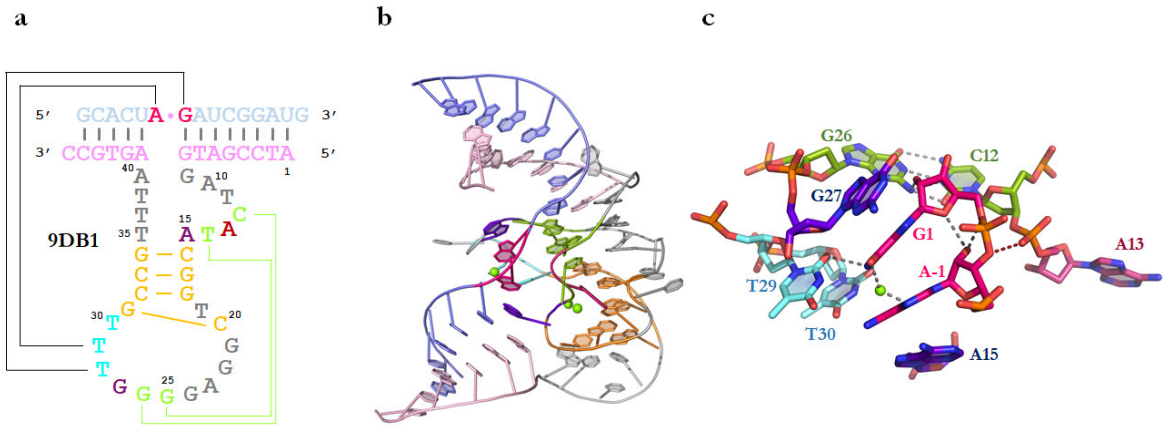


Figure 15. 9DB1 deoxyribozyme.

(a) Tertiary fold diagram of the 9DB1 deoxyribozyme (b) Overall crystal structure of the 9DB1. (c) The active site of the 9DB1. The G1 and A-1 (pink) are sandwiched between purine bases of G27 and A15. The catalytic dA13 (violet) phosphate, oriented by the long-range base pair G26:C12 (green) brings the *pro*-Sp non-bridging oxygen in proximity to the 3'-oxygen of A-1 (red dashed line). The Mg^{2+} cations are shown as green spheres.

In 2019, Aranda et al. described a possible ligation mechanism and suggested how the 9DB1 recruits magnesium cations to support the RNA ligation reaction (Aranda, Terrazas et al. 2019). They used molecular dynamics simulations and molecular interaction potential calculations, combined with high-level hybrid quantum mechanics/molecular mechanics (QM/MM) free energy calculations to identify magnesium binding sites on the pre-catalytic structure and describe its role in positioning reactive groups and stabilizing transition state. They proposed a mechanism where the O3' atom from A-1 attacks the α-phosphate group of the G1, and the leftover proton from the O3' atom is then transferred to the DNA (catalytic dA13 *pro*-Sp phosphate oxygen). The mechanism involves the active participation of two Mg^{2+} ions, not observed in the original crystal structure (Aranda, Terrazas et al. 2019, Höbartner 2019).

The goal of the study is to elucidate the crystal structure of the RNA-ligating deoxyribozyme in the pre-catalytic state of reaction to properly describe the ligation mechanism and explain the role of magnesium cations as cofactors for accelerating RNA ligation.

1.5.2 Fluorogenic RNA aptamers

The use of fluorescence proteins (FPs) allowed tracking tagged proteins *in vivo* in real-time and revolutionize cell biology (Day 2014). The GFP from *Aequorea victoria* (Shimomura, Johnson et al. 1962) was the first discovered fluorescent protein and became a starting point for a diverse fluorescent protein toolbox (Rodriguez, Campbell et al. 2017). In contrast to the wide range of tools for proteome studies, there are few fluorescent tags useful for visualizing the transcriptome in real-time. As there are no known naturally fluorescent RNAs, the invention of in vitro selection (SELEX) allowed to discover non-biological fluorescent RNA-dye complexes. The RNAs discovered by SELEX bind to exogenous, small-molecule fluorophores, and enhance their fluorescence (Trachman and Ferre-D'Amare 2019). They are termed fluorescent light-up probes, or fluorogen-activating aptamers (FLAPs). The leading examples of fluorogenic aptamers include the malachite green aptamer (Babendure, Adams et al. 2003), Spinach (Paige, Wu et al. 2011), Broccoli (Filonov, Moon et al. 2014), Corn (Warner, Sjekloca et al. 2017), Chili (Steinmetzger, Palanisamy et al. 2019), Mango (Dolgosheina, Jeng et al. 2014), SiRA (Wirth, Gao et al. 2019), Coral (Bouhedda, Fam et al. 2020) and Pepper (Chen, Zhang et al. 2019) aptamers. In the present study, I will focus on Chili, an RNA aptamer from the same family as Spinach and Corn aptamers. All these aptamers mimic green fluorescent proteins (GFP) that bind and activate derivatives of HBI (4-hydroxybenzylidene imidazolone), a small-molecule analog of the tripeptide chromophore scaffold present in GFP.

1.5.2.1 Spinach RNA aptamer

Spinach was the first artificial aptamer that mimics the green fluorescent protein (GFP) (Paige, Wu et al. 2011). Spinach binding to the DFHBI results in green fluorescence (Paige, Wu et al. 2011) while binding to DFHBI 1T and DFHBI 2T, inducing cyan and yellow fluorescence respectively (Song, Strack et al. 2014). DFHBI is nonfluorescent in its free, unbound form, and after its binding to the Spinach results in fluorescence. The Spinach-DFHBI complex is non-toxic when illuminated and therefore can be used as a genetically encoded tag to visualize RNAs in bacteria and mammalian cells (Paige, Wu et al. 2011).

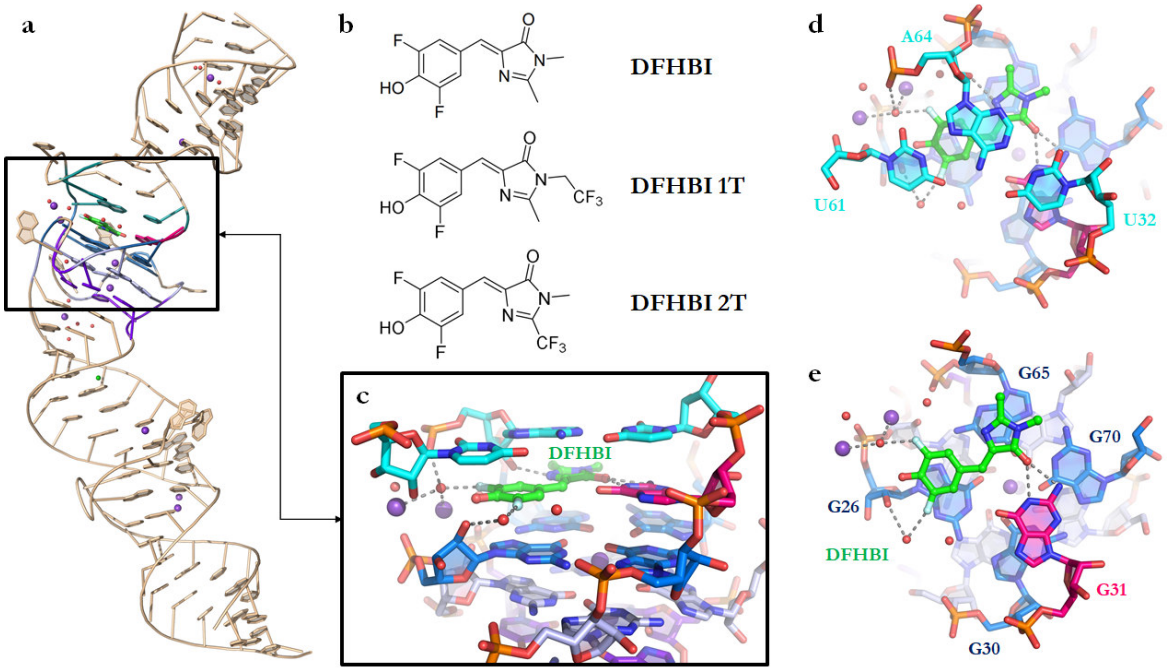


Figure 16. Spinach RNA fluorogenic-aptamer.

(a) Overall crystal structure of the Spinach (PDB: 4TS0). (b) Fluorophores of the Spinach aptamer. Side (c) and top (d,e) views of the DFHBI binding pocket. K^+ and water molecules are represented as purple and red spheres, respectively. Dashed, grey lines represent hydrogen-bonding of the DFHBI fluorophore to the Spinach RNA (also mediated by water and K^+ molecules).

The crystal structure of the Spinach aptamer in complex with DFHBI revealed that Spinach folds into a coaxial helical stack with a length of around 100 Å that is composed of a G-quadruplex core flanked by the two A-form duplexes (37,38). The G-quadruplex is formed by two G-quartets, which are glanced by a mixed based quadruple and fluorophore. DFHBI is sandwiched between base triplet (U32:A64*U61), G-quartet (G26/G30/G65/G70) (Figure 16d,e), and it is anchored to the unpaired guanine residue (G31). Mutational redesign of the Spinach core has confirmed that G-quadruplex, unpaired G31, and base triplet are important in the fluorescence activation (Warner, Chen et al. 2014). The Spinach G-quadruplex containing two G-quartets shows complex, mixed parallel, and antiparallel connectivity. Guanosines forming Spinach G-quadruplex are in the *syn* and *anti* conformations of the N-glycosidic bond angles and possess both, 2'-*endo* and 3'-*endo* sugar puckers. (Huang, Suslov et al. 2014, Warner, Chen et al. 2014).

The DFHBI aromatic rings adopt co-planar conformation when bound to the Spinach. Fluorophores containing two aromatic rings (like DFHBI) show improved fluorescence when they adopt such a conformation (Carlsson, Larsson et al. 1994). DFHBI forms multiple hydrogen-bonding interactions with the Spinach, with some of them being water and K^+ mediated (Figure 16c,d,e). This prevents DFHBI from dissociation from the complex and promotes the phenolate

state of the fluorophore when bound to Spinach. The crystal structure of the Spinach RNA in the absence of DFHBI have revealed that the DFHBI binding site collapses in the absence of fluorophore, due to the reorganization of base triplet, while the overall global architecture remains unaffected (Huang, Suslov et al. 2014). Crystal structures also allowed to miniaturize parent Spinach to 45 nucleotide long Baby-Spinach (Warner, Chen et al. 2014). Baby spinach has a similar fluorescence as parent Spinach, with improved folding properties (Okuda, Fourmy et al. 2017).

Spinach has several disadvantages that limit its *in vivo* applications. As a result of its relatively large size (98 nucleotides) and complex topology, the Spinach tends to misfold (Strack, Disney et al. 2013, Okuda, Fourmy et al. 2017). After photoexcitation, Spinach losses fluorescence very quickly, due to the isomerization of the fluorophore. If there is an excess of DFHBI in the solution, Spinach binds to non-photodamaged cis form or change its conformation to accommodate photoisomerized chromophore to fully recover its fluorescence (Han, Leslie et al. 2013). RNA Spinach was also found to bind to other sets of fluorophores, often with higher affinity than to the DFHBI (Jeng, Chan et al. 2016), which may limit the use of Spinach as a FRET donor/acceptor with other fluorogenic aptamers.

1.5.2.2 Corn RNA aptamer

Another fluorogenic aptamer, obtained in an in vitro selection experiment which binds the DFHO was named Corn (Figure 17a) (Song, Filonov et al. 2017). Corn is a light-up aptamer composed of 36 nucleotides. In contrast to the Spinach/DFHBI complex, the Corn-DFHO complex mimics the red fluorescent proteins by inducing red-shifted fluorescence. The DFHO fluorophore (3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime) is an extended derivative of DFHBI that carries an oxime sidechain (Figure 17b) and similarly to DFHBI, DFHO exhibits negligible fluorescence when it is free in solution. Corn is another example of the RNA aptamer containing G-quadruplex (Warner, Sjekloća et al. 2017). In contrast to Spinach, Corn is forming a homodimer that binds DFHO fluorophore at its interprotomer interface (Figure 17a,c).

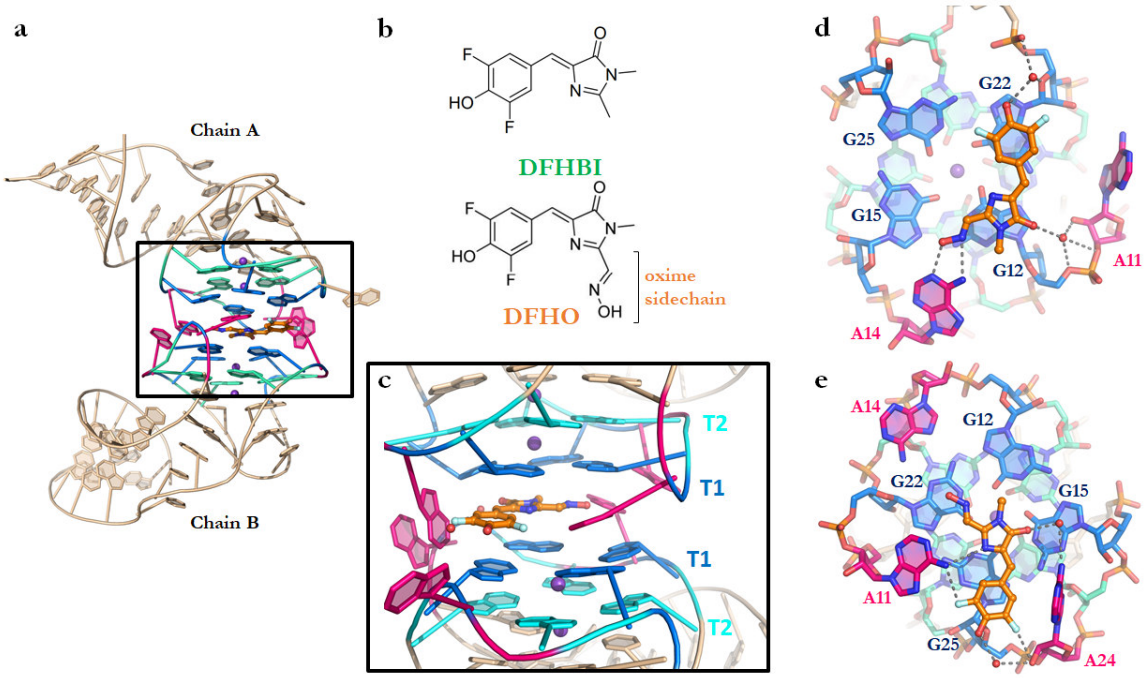


Figure 17. Corn RNA fluorogenic aptamer.

(a) Overall crystal structure of the Corn (PDB: 5BJO). The G-quartets are colored in blue and cyan. Each chain of Corn aptamer consists of two tetrads, forming a four-tiered G-quadruplex in a functional Corn homodimer. (b) Comparison of the DFHBI and DFHO chromophores. Side view of the DFHO binding core. Binding of the DFHO ligand to the chain A (d) and chain B (e) of the Corn aptamer. K^+ and water molecules are represented as purple and red spheres, respectively. Interfacial adenosines are colored in pink. Dashed, grey lines represent hydrogen-bonding of the DFHO fluorophore to the Corn RNA (also mediated by the water molecules).

The Corn homodimer is forming a four-tiered G-quadruplex flanked by the A-form helices. The G-quadruplex has antiparallel connectivity. The Corn G-quadruplex comprises two G-quartets stabilized by potassium cations. In contrast to Spinach, loops connecting G-quartets are small (1 nucleotide) with the longest composed of 5 nucleotides (Warner, Sjekloca et al. 2017). The DFHO chromophore is sandwiched between two G-quartets from each protomer (T1 - G12/G15/G22/G25) in a planar conformation (Figure 17c). Interfacial adenosines 11, 24, and 24 flanking the fluorophore binding site and are stabilizing the fluorophore in the binding site by forming hydrogen-bonding interactions (also water-mediated) with the ligand (Figure 17d,e).

DFHO binds between G-quartets in the plane of mirror symmetry of two protomers. Since Corn is chiral, it cannot match the symmetry of the bound ligand, therefore it breaks symmetry locally at the binding site. Except for the adenosines near the fluorophore binding site, the Corn aptamer is almost identical. Corn is unique among dimeric RNAs for two reasons – it is quasisymmetric and additionally lacks base-pairing interactions between two protomers (Jones and Ferre-D'Amare

2015). Moreover, the structure of the Corn aptamer was solved in the apo-state (ligand-free) and together with non-specific but G-quadruplex preferent chromophores - thioflavin-T (ThT) and thiazole orange (TO) (Sjekloća and Ferré-D'Amaré 2019).

The crystal structure of the Corn-TO complex revealed a single molecule of Thiazole Orange (TO) bound at the dimer interface (Figure 18a,d). Because the bound TO, similarly to the Corn-TO co-crystal structure, is traversed by a crystallographic 2-fold axis, the ligand exists in two distinct conformations in the binding site. Each asymmetric unit contains a single copy of TO with the crystallographic occupancy of 0.5 (adding to a full molecule of TO in the binding site of the crystallographic dimer) (Sjekloća and Ferré-D'Amaré 2019).

The crystal structure of the Corn-ThT complex revealed that the two ThT molecules occupy the binding site in an antiparallel, "head-to-tail" orientation in the same plane (Figure 18b,e). In the co-crystal structure, each ThT ligand has an occupancy of 0.5, therefore, a single protomer binds one molecule of the fluorophore. Moreover, because the crystallographic 2-fold axis does not coincide with the non-crystallographic 2-fold rotation relating the two bound ThT molecules, this results in two distinct alternate poses of the bound fluorophore molecules (Sjekloća and Ferré-D'Amaré 2019).

Interestingly, in the apo-form of the Corn aptamer (ligand-free), the fluorophore binding site has collapsed, so the G-quartets from the two protomers stacks directly on each other (Figure 18c,f). In this structure, the interfacial adenosines adopt conformations similar to those in the ThT complex structure, but different from those in the DFHO complex (Sjekloća and Ferré-D'Amaré 2019).

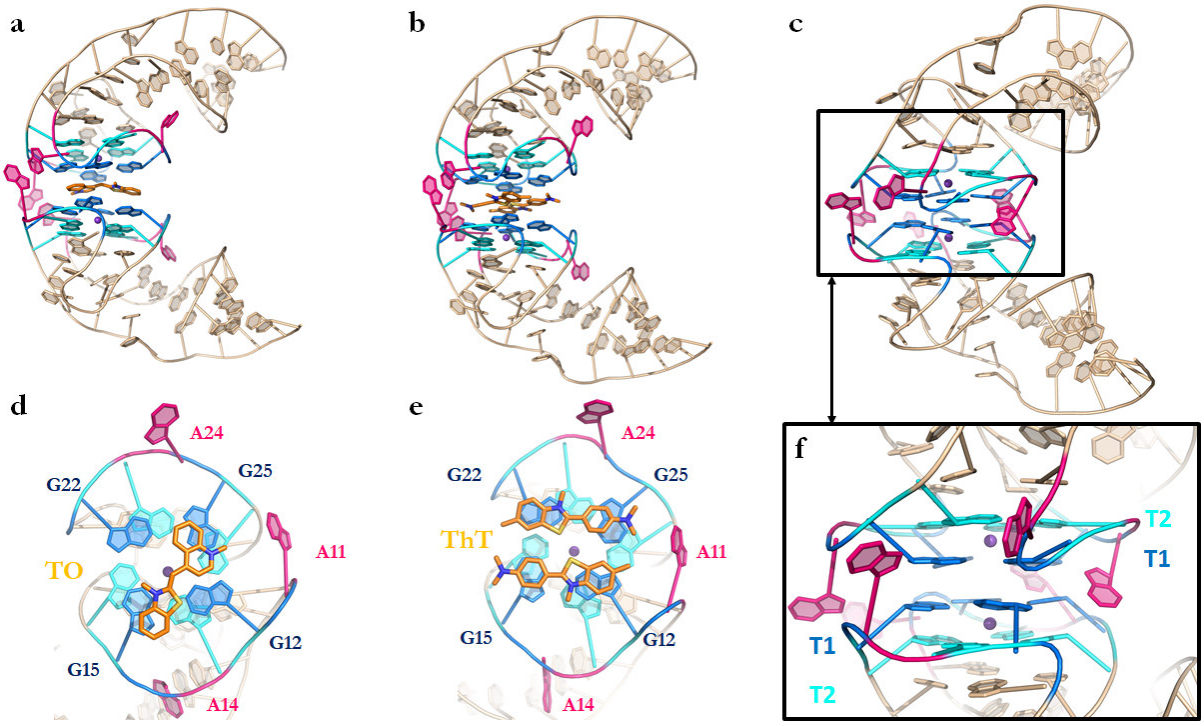


Figure 18. The Corn RNA aptamer.

Crystal structure of the Corn RNA aptamer in a complex with (a) TO, (b) ThT, and its apo-form. The ligands are sandwiched between T1 quartets of Corn aptamers. (d,e) Top view on the TO and ThT binding site. (f) Close-up view of the G-quadruplex of the apo-form. T1 and T2 quartets are colored in blue and cyan respectively. Interfacial adenosines are colored in pink. K^+ cations are shown as purple spheres.

1.5.2.3 Chili RNA aptamer

DFHBI and DFHO, the Spinach and Corn fluorophores, due to their low pK_a values, are easily deprotonated and exist predominantly as phenolates under physiological conditions (Paige, Wu et al. 2011, Song, Filonov et al. 2017). They are also known to function by the canonical twisted intramolecular charge transfer (TICT). The fluorogenic aptamers stabilize a planar conformation of bound ligands to maximize its fluorescence. The fluorogenic aptamer isolated from the Spinach selection, named 13-2 which was originally reported to activate fluorescence of DMHBI, exhibits a large Stoke shift of 137 nm (Paige, Wu et al. 2011). This may suggest incorporating excited-state proton transfer (ESPT) in the activation mechanism. Large Stoke shift fluorescent proteins (termed LSS FP's), such as LSSmOrange (Shcherbakova, Hink et al. 2012) and LSSmKate (Piatkevich, Hulit et al. 2010, Piatkevich, Malashkevich et al. 2010), have been engineered to exploit ESPT in the fluorescence turn-on mechanism. The proton is extracted from the chromophore hydroxy-group to the neighboring amino-acid (Piatkevich, Malashkevich et al. 2010).

Miniaturization of the 13-2 aptamer (around 60 nucleotides) resulted in the Chili aptamer (52 nucleotides) (Steinmetzger, Palanisamy et al. 2019). The Chili aptamer binds and activates dimethoxy-substituted analogs of DFHBI as their phenol form, termed DMHBI. Moreover, the addition of the cationic N-trimethylaniline moiety to the DMHBI resulted in the $DMHBI^+$ derivative. To mimic red fluorescent proteins, such as DsRed and its relatives, the oxime moiety was introduced as a stable analog of acylimine moiety integrated into the peptide backbone between the amino acid residue preceding the chromophore and the first residue of the chromophore (Day and Davidson 2009), resulting in the $DMHBO^+$ derivative (Figure 19). $DMHBO^+$ and $DMHBI^+$ resulted in the large stoke shifts - 136 and 129 nm respectively. $DMHBO^+$ is also a better binder (K_D=12 nM) than $DMHBI^+$ (Steinmetzger, Palanisamy et al. 2019). The Chili can also bind fluorophores with different substitution patterns, which results in green, yellow, and red fluorescence emission (Steinmetzger, Bessi et al. 2019).

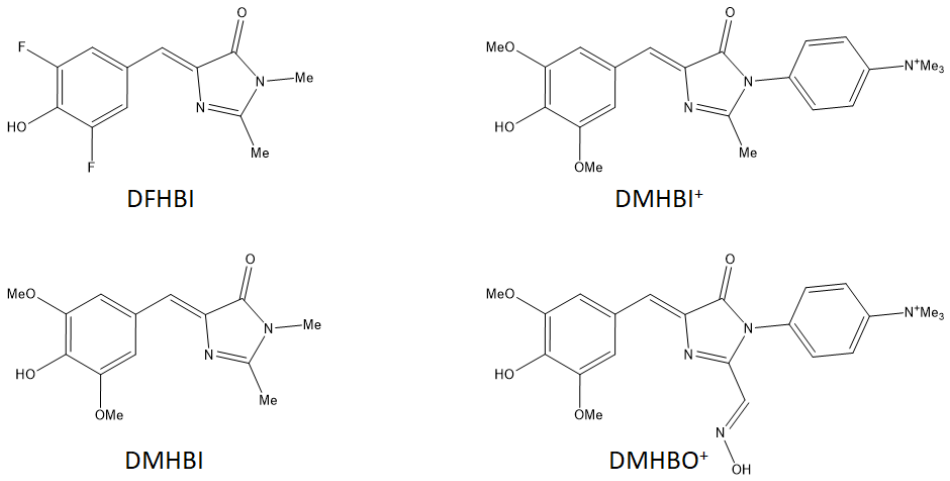


Figure 19. Comparison of the fluorophores.

The structure of the Chili aptamer in complex with its cognate ligand(s) remains unknown. The high guanine content of the Chili RNA sequence (25 guanosines out of 52 nucleotides) and NMR data suggest the presence of two-tiered, G-quadruplex (Warner, Chen et al. 2014, Steinmetzger, Bessi et al. 2019). There is an open question on how Chili mimics LSS FPs. The structural studies will help to reveal the identity and function of the proton acceptor and describe the difference between the binding affinity of $DMHBO^+$ and $DMHBI^+$.

2 MATERIALS AND METHODS

2.1 BIOCHEMISTRY OF NUCLEIC ACIDS

2.1.1 RNA and DNA preparation

Unmodified DNA oligonucleotides for crystallization experiments were purchased from Sigma Aldrich. Modified RNA and RNA with 5'-triphosphate were prepared using solid-phase synthesis and T7 RNA polymerase in vitro transcription respectively. If necessary, oligonucleotides were purified using denaturing polyacrylamide gel electrophoresis. The quality of purified oligonucleotides was assessed by mass spectrometry and HPLC analysis. The list of DNA and RNA oligonucleotides in this thesis can be found in section 2.1.5.4.

2.1.2 Solid-phase synthesis and deprotection of oligonucleotides.

Solid supports

TOM rA (n-acetyl) CPG 1000Å support	ChemeGenes
TOM rC (n-acetyl) CPG 1000Å support	ChemeGenes
TOM rG (n-acetyl) CPG 1000Å support	ChemeGenes
TOM rU CPG 1000Å support	ChemeGenes
deoxy Adenosine (n-bz) CPG 500Å support	ChemeGenes
deoxy Cytidine (n-acetyl) CPG 500Å support	ChemeGenes
deoxy Guanosine (n-ibu) CPG 500Å support	ChemeGenes
Thymidine CPG 500Å support	ChemeGenes

Phosphoramidites

5'-DMT-2'-TOM-ribo Adenosine (n-acetyl) OP	ChemeGenes
5'-DMT-2'-TOM-ribo Cytidine (n-acetyl) OP	ChemeGenes
5'-DMT-2'-TOM-ribo Guanosine (n-acetyl) OP	ChemeGenes
5'-DMT-2'-TOM-ribo Uridine OP	ChemeGenes
5'-DMT-2'-TBDMS-N6-Methyl-ribo Adenosine-CE Phosphoramidite	Homemade
5'-DMT-deoxy Adenosine (n-bz) CED phosphoramidite	ChemeGenes

5'-DMT-deoxy Cytidine (n-acetyl) CED phosphoramidite	ChemeGenes
5'-DMT-deoxy Guanosine (n-ibu) CED phosphoramidite	ChemeGenes
5'-DMT-Thymidine CED phosphoramidite	ChemeGenes
Solvents and reagents	
Acetonitrile for DNA synthesis (≤10 ppm H2O)	J.T. Baker/Roth
Acetonitrile (HPLC quality)	Merck
Dichloroacetic acid, 99%	Acros Organics
1,2-Dichloroethane	Carl Roth
Acetic Anhydride 99%+	Acros Organics
2,4,6-collidine	Acros Organics
Iodine	Acros Organics
5-(ethylthio)-1H-tetrazole (ETT)	Carbosynth
Deblock-TCA (ABI)	J.T. Baker
Capping A (ABI) [Acetic anhydride]	J.T. Baker
Capping B (ABI) [1-Methylimidazole]	J.T. Baker
Oxidizing (ABI) [Pyridine]	J.T. Baker
Molecular sieves (4 Å)	Sigma Aldrich
Deprotection solutions and reagents	
Ammonia solution 25% in H2O	Sigma Aldrich
Methylamine solution 40% in H2O	Sigma Aldrich
Triethylamine trihydrofluoride (TEA 3HF)	Sigma Aldrich
DMSO anhydrous ≥99.9%	Sigma Aldrich
In vitro transcription reagents and chemicals	
Magnesium chloride hexahydrate	Merck
Tris-HCl Roth	Roth
Dithiothreitol (DTT) Roth	Roth

Spermidine trichloride Sigma Aldrich	Sigma Aldrich
Nucleotide triphosphates: ATP, CTP, GTP, UTP	Jena Bioscience
Ethylenediaminetetraacetic acid (EDTA) Sigma Aldrich	Sigma Aldrich
T7 polymerase (homemade)	Homemade
Gel electrophoresis and EtOH precipitation	
Urea	Merck
Tris-HCl	Roth
Boric acid	Roth
Rotiophorese gel 40 (acrylamide/bisacrylamide 19:1, w/v)	Roth
EDTA	Roth
NaCl	Merck
Xylencyanole	Roth
Bromophenol blue	Roth
Formamide	Merck
EtOH absolute	Merck
N,N,N‘,N‘-tetramethyl ethylene diamine (TEMED)	Roth
Ammonium peroxodisulfate (APS)	Roth
Anion-exchange HPLC analysis	
Tris-HCl	Roth
NaClO4·H2O	Sigma Aldrich
Urea	Merck
DNAPack PA200 2x250 mm column	Dionex

Modified RNA oligonucleotides were synthesized using controlled pore glass (CPG) supports from ChemGenes on ABI 392 or Pharmacia Gene Assembler Plus synthesizers. The synthesis scale was set to 1 µmol with 5-(ethylthio)-1H-tetrazole (ETT) as the activator. Phosphoramidites were employed as 100 mM solutions in anhydrous acetonitrile and dried over activated molecular sieves (3Å).

Before synthesis, 0.8 - 1.0 µmol of solid support was loaded into a synthesis column. The synthesis was started by the initial detetrytilation of the solid support for 1 minute. The coupling time for TOM-protected RNA phosphoramidites was 4 minutes and 12 minutes for any modified phosphoramidite. Capping solutions A and B mixed in 1:1 ratio were used for capping of unreacted groups on the column. The oxidation step was achieved using iodine solution for 1 minute. HPLC quality acetonitrile and/or dichloroethane was used in between each synthesis step. Coupling efficiency was typically between 97 to 98%. After synthesis, columns with synthesized oligonucleotides were dried under vacuum for 15 minutes and then exposed to deprotection.

2.1.3 Deprotection of the oligonucleotides

RNA oligonucleotides were cleaved and deprotected using 1ml of AMA, which is a 1:1 mixture (v/v) of aqueous ammonium hydroxide and aqueous methylamine at 65°C for 1 hour. Afterward, CPG solid support was removed and the solvent was evaporated in speed-vac. The remaining RNA was re-dissolved in 200 µl anhydrous DMSO. If necessary, oligonucleotides were heated at 65°C for about 5 minutes to get into the solution. Next, 250 µl of triethylamine trihydrofluoride was incorporated, mixed, and heated for 2.5 hours at 65°C for complete removal of the 2' protecting group. After deprotection, the mixture was cooled down in the freezer briefly.

Desalting of the mixture was done using the butanol precipitation method. The RNA in DMSO/THF (triethylamine trihydrofluoride) solution was split into two tubes. 25 µl of 3 M Sodium Acetate in filtered MiliQ water was added to each RNA solution and mixed by vertexing. To precipitate RNA from the solution 1ml of butanol was subjected to each tube, mixed, and cooled at -70°C for 30 minutes, followed by centrifugation for 10 minutes at 13.000 RPM. Butanol was decanted, the RNA pellet was washed twice using ice-cold 70% ethanol and dried in speed-vac. The product was re-dissolved in 500 µl of MiliQ water. A portion of crude RNA product was analyzed using a Dionex PA-200 HPLC column with a sodium perchlorate gradient at 60°C. After determining crude quality, RNA was subjected to PAGE purification.

2.1.4 In vitro transcription of RNA

DNA templates for in vitro transcription were purchased from Sigma Aldrich. In vitro transcription reactions were performed with T7 RNA polymerase using the corresponding DNA template and T7 promoter strand (1 µM each) in an aqueous solution containing 40 mM Tris-HCl, pH 8.0, 30 mM MgCl2, 10 mM DTT, 4 mM of each NTP, and 2 mM spermidine at 37 °C for 6 hours. The transcription reaction was quenched using 50 mM ethylenediaminetetraacetic acid (EDTA) and purified by PAGE.

2.1.5 Oligonucleotides purification and quality validation

2.1.5.1 Polyacrylamide gel electrophoresis (PAGE)

Denaturing polyacrylamide gels (0.7×200×300 mm) were prepared from 10 - 20% acrylamide/bis-acrylamide 19:1 and 7 M urea and used with running buffer 1x TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3). The polyacrylamide gel percentage depended on the length of nucleic acid. N,N,N‘,N‘-tetramethylethylene diamine (TEMED), and ammonium peroxydisulfate (APS) were used to initiate gel polymerization. The gels were running at 35 W constant power in a 1X TBE buffer. The oligonucleotide sample was mixed with loading buffer (80% formamide, 1xTBE (89 mM each Tris and boric acid, pH 8.3), 50 mM EDTA, and 0.025% bromophenol blue and xylene cyanol) in 1:1 ratio and heat-denatured at 95°C for 3 minutes before loading on the gel. After the gel electrophoresis, products were visualized by UV shadowing on thin-layer chromatography (TLC) plate and extracted by „crush & soak” method. The gel pieces with the desired product were incubated in TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA pH 8.0, and 300 mM NaCl) in a 1:3 ratio and extracted for 2 hours at 37°C with light shaking. The second extraction was performed as the first one and combined supernatants were ethanol precipitated. Three volumes of ice-cold ethanol were combined with one volume of the sample, followed by mixing and freezing in liquid nitrogen for 3 minutes. Next, a sample was centrifuged at 13.000 RPM for 30 minutes at 4°C. After ethanol and TEN buffer mixture were decanted, oligonucleotides pellets were washed with ice-cold 70% ethanol, dried under vacuum, and dissolved in MiliQ water. Oligonucleotides concentration was measured using NanoDrop Microvolume Spectrophotometer with a dedicated application for single-stranded oligonucleotides.

2.1.5.2 HPLC analysis

Anion-exchange HPLC analyses were performed using an ÄKTA Purifier from GE Healthcare and Shimadzu HPLC systems with a DNAPac PA200 column (2x250 mm) from Thermo Scientific. 100 - 200 pmol of the sample was injected into the column equilibrated in Buffer A (25 mM Tris-HCl pH 8.0, 6M Urea). The run was performed at 60°C and a sample was eluted by an increasing gradient of Buffer B (25 mM Tris-HCl pH 8.0, 6 M Urea, 0.5 M $NaClO_4$) in Buffer A with the steepness of 4% for column value (CV). Oligonucleotides were detected by measuring absorbance at 260 nm.

2.1.5.3 Mass Spectrometry

Purified oligonucleotides were analyzed by electrospray ionization (ESI) mass spectrometry using Q-ToF Ultima from Waters by Uwe Pleßmann and Monika Raabe from the Bioanalytical Mass Spectrometry Group located in the Max Planck Institute for Biophysical Chemistry led by Prof. Henning Urlaub.

2.1.5.4 List of oligonucleotides

Table 1. List of the DNA oligonucleotides used in the study.

D008	CTGTAATACGACTCACTATA
D1086	GGCTAGCCGACCGCACCCAACCACCAGCGAACTGGCGCCCCTCCAGCTAGCCT ATAGTGAGTCGTATTACAG
D1342	ATCCGAC GGATCATACGGTCGGAGGGGTCTGCCGTTTA AGTGCC*
D1739	CGGAAG GGGTCTCCAGCTGGACGTTA TTCATT
D1740	TCGGAAG GGGTCTCCAGCTGGACGTTA TTCAT
D1745	CGGAAG GGGTCTCCAGCTGGACGTTC TTCATT
D1776	GGAAG GGGTCTCCAGCTGGACGTT CTTCAT
D1777	GTCGAA GGGGTCTCCAGCTGGACGTT CTTCA
D1778	GCCTGGC GGGTCTCCAGCTGGACGTT CGTGCCT
D1841	TAGTTACGGAAG AAAGGGCGGGCAAACTCTG CTTCATTCAGT
D1842	TAGTTACGGAAG AAAGGGCGGGCAAACTCTGA TTCATTCAGT
D1960	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGCAA
D1961	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGCA
D1962	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGC
D1963	ATCCGA GGGTCTCCAGCTGGACGTTA TAGTGCC
D1964	ACTCAAAG GGGTCTCCAGCTGGACGTTA CAATGCA
D1965	ACTCAAAG GGGTCTCCAGCTGGACGTTA CAATGC
D1966	CCTCAAAG GGGTCTCCAGCTGGACGTTA CAATG
D1967	CATCCGA GGGTCTCCAGCTGGACGTTA TAGTGC
D2016	TCTCAAAG GGGTCTCCAGCTGGACGTTA CAATG
D2017	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGG
D2018	CCTCAAAG GGGTCTCCAGCTGGACGTTA CAATGC
D2019	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGCG
D2020	CTCAAAG GGGTCTCCAGCTGGACGTTA CAATGC
D2021	CTCTCAAAG GGGTCTCCAGCTGGACGTTA CAAT
D2128	TTCTCAAAG GGGTCTCCAGCTGGACGTTA CAA
D2129	CCTCTCAAAG GGGTCTCCAGCTGGACGTTA CAAT

*** Red color represents catalytic core sequence of the deoxyribozyme**

Table 2. List of RNA oligonucleotides used in the study.

Tr108	GGCUAGCUGGAGGGGCGCCAGUUCGCUGGUGGUUGGGUGCGGUCGGCUAGCC
Tr185	pppGGUCGGAUG
R423	GCACUA
R556	pGGUCGGAUG
R593	GCACU(dA)
R597	AAUGAAG(dG)ACUUCCG
R616	AUGAAG(dG)ACUUCC
R617	UGAAG(dG)ACUUCGAC
R618	AGGGACG(dG)AGCCAGGC
R633	AUGAAG(dG)(m6A)CUCC
R634	AUGAAGG(m6A)CUUCC
R663	UGCAUUG(dG)ACUUUGAGU
R664	UGCAUUG(dG)ACUUUGAG
R665	GCAUUG(dG)ACUUUGAG
R666	GCACUA(dG)AUCGGAUG
R694	ACAUUG(dG)ACUUUGAG
R695	CAUUG(dG)ACUUUGAGC
R696	GGCAUUG(dG)ACUUUGAG
R697	GCAUUG(dG)ACUUUGAGC
R698	GCCAUUG(dG)ACUUUGA
R699	GAUUG(dG)ACUUUGAGA
R727	AUUG(dG)ACUUUGAGA
R728	GAUUG(dG)ACUUUGAGAG
R744	UGCAUUG(dG)(m6A)CUUUGAG

2.2 X-RAY CRYSTALLOGRAPHY METHODS

Structure determination of macromolecules using X-ray crystallography requires different consecutive steps that are shown in Figure 20. The flow diagram provides a general overview of the steps in the structure determination process. All experimental laboratory work is expressed as a blue shaded box, while all steps after data collection are conducted *in silico*.

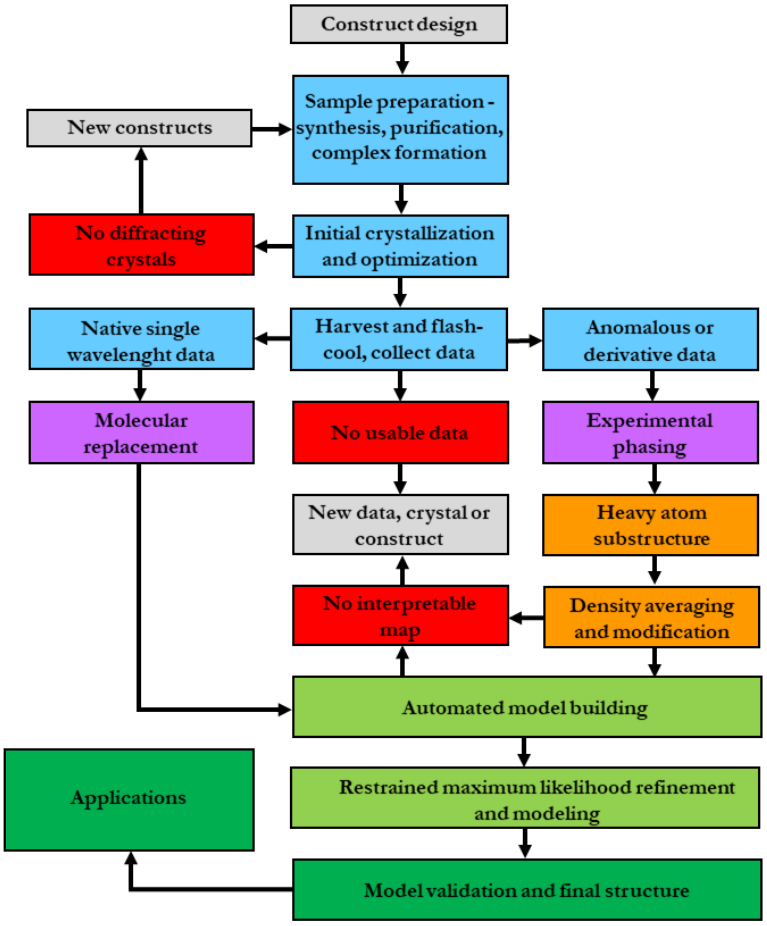


Figure 20. Key stages in the X-ray structure determination process.

2.2.1 Sample preparation

2.2.1.1 Deoxyribozyme complex preparation

Crystallization samples were formed by mixing the DNAzyme (in ddH_2O), RNA substrate (in ddH_2O) in a 1:1 molar ratio. The single deoxynucleotide residue mutation was introduced in the cleavage site to prevent RNA cleavage during crystallization, thus mimic the pre-catalytic state of reaction. Complexes were analyzed by HPLC and based on the molar extinction coefficient (ε), the ratio between the peaks was calculated and the ratio DNA/RNA was adjusted necessary. Complexes were concentrated using Vivaspin 3000 MWCO ultrafiltration spin columns at 12.500

RPM to ~1 mM. Next, HEPES pH 8.0 was added to obtain a final concentration of 10 mM. The sample was denatured at 95°C for 3 minutes and cooled down for 15 - 30 minutes. Following, $MgCl_2$ was added to get a final concentration of 20 mM. Complexes were stored at -20°C before the crystallization.

2.2.1.2 Chili RNA aptamer complex preparation

Crystallization complexes were formed by mixing RNA (in ddH_2O) and ligand (in DMSO) at a 1:1.2 molar ratio in 10 mM HEPES pH 8.0, 50 mM KCl, and 1.5% DMSO buffer. The complex was concentrated using Vivaspin 3000 MWCO ultrafiltration spin columns at 12.500 RPM to ~ 0.5 mM. The sample was denatured at 95°C for 3 minutes and cooled down for 15 - 30 minutes. Following, $MgCl_2$ was added to get a final concentration of 5 mM. Complexes were stored for 16 hours at 4°C before the crystallization.

2.2.2 Crystallization of the macromolecules

X-ray crystallography has evolved into a very powerful tool to determine the three-dimensional structure of macromolecules, however, the first prerequisite for structure determination is the ability of macromolecule to form well ordered, single crystal. Crystallization of macromolecules is a complex task and may be influenced by many variables and factors, which are presented in Table 3. Therefore, usually is necessary to rationally design crystallization constructs and search for optimal crystallization conditions to obtain well-ordered, nicely diffracting crystals.

Table 3. Variables and factors influencing the crystallization experiment.

Biochemical and chemical variables that could or do affect crystallization	Physical variables that could or do affect crystallization
▪ Purity of the sample ▪ Concentration of the sample ▪ pH and buffer ▪ Metals ▪ Ionic strength ▪ Conformational flexibility of the sample	▪ Vibration and sound ▪ Temperature ▪ Rate of equilibration ▪ Time ▪ Method of crystallization ▪ Volume of the sample and reagent

The macromolecule crystallization is in general controlled precipitation process. It involves a transition from the liquid state into the solid, crystalline state. A crystallization experiment begins with the sample in a stabilizing solution of water and other additives such as buffer and salt. Before

mixing with the crystallization solution, the macromolecule sample is undersaturated, therefore no crystal can be formed. The spontaneous homogeneous nucleation takes place in the Labile Zone of supersaturation and the crystal further grows in the Metastable Zone of supersaturation. The phase diagram, showing transitions that take place in the process of crystallization is shown in Figure 21.

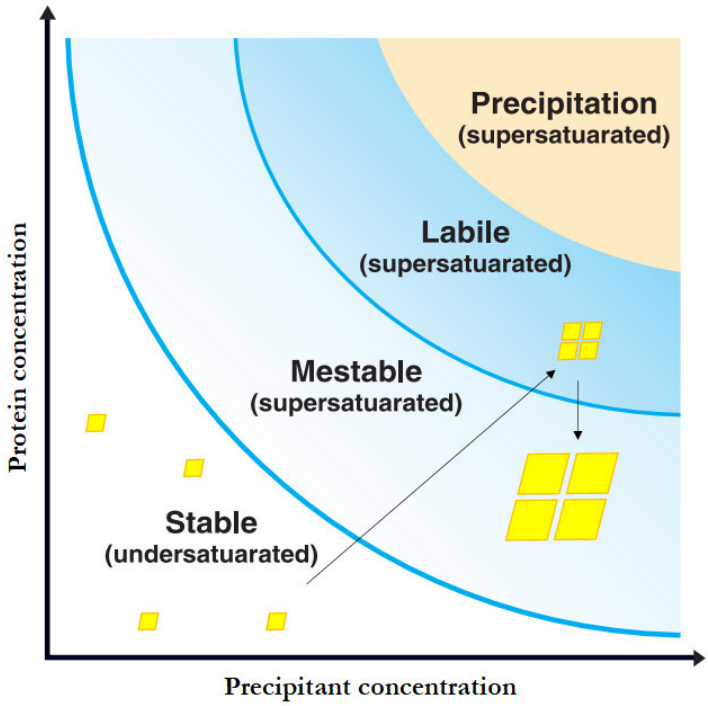


Figure 21. The phase diagram.

2.2.2.1 Setting up initial trials: High-Throughput crystallization screening

Finding crystallization conditions for a new macromolecule requires a lot of trials. The first step in high-throughput crystallization screening is to set up as many conditions as possible. For this purpose, commercially available screens are used. This leads to determine „hits" that points to conditions that can be further optimized. The most common way to optimize the „hit" is to fine-tune the crystallization condition by varying concentrations of precipitant, pH, or temperature. The most common technique at this stage is vapor diffusion. It can be done in two ways - using a sitting or hanging drop. The drop containing macromolecule sample mixed with crystallization cocktail is exposed to a reservoir with a large volume of crystallization cocktail. The drop is losing water in the diffusion process until the concentration of all reagents in it matches that in the reservoir.

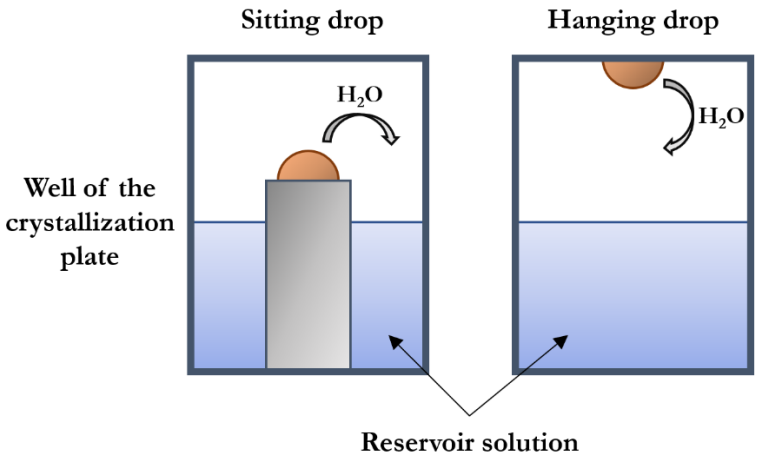


Figure 22. Overview of vapor diffusion technique.

Initial crystallization screen was performed using the sitting drop vapor-diffusion technique on 96-well MRC plates (Molecular Dimensions). Drops consisting of 60 nl or 120 nl of the sample were mixed with 60 or 120 nl of the reservoir were dispensed using 8-Channel Cartesian Nanodispenser Robot (Zinser Analytic). To prevent evaporation during setting crystallization plates, the humidity was controlled above 90% in the robot chamber. MRC plates were sealed using Crystal Clear Sealing Tape (Hampton Research) and stored in the RockImager (Formulatrix) with the temperature set to 20°C. The plates were imaged at days: 0, 1, 3, 5, 10, 20, and 30, and images were visualized using the RockMaker software (Formulatrix).

Commercially available crystallization screens used for initial screening

Nucleix	Qiagen
Natrix	Hampton Research
Helix	Molecular Dimensions
MIDAS	Molecular Dimensions
Wizard Classic 1+2	Jena Bioscience
Wizard Classic 3+4	Jena Bioscience

2.2.2.2 Optimization of initial crystallization condition

Initial crystals from crystallization screening are often small and because of that, not suitable for diffraction experiments. Therefore, initial crystals have to be reproduced and optimized. Sometimes, it may be difficult to reproduce initial crystals, so it is important to optimize as many variants of the original condition as possible. As mentioned before, optimization of crystallization may be done condition by varying concentration of precipitant, pH, drop size, macromolecule

sample concentration, or crystallization temperature. Additionally, optimization could be performed through the addition of additives or detergent to the original crystallization cocktail and by seeding or other techniques.

Initial hits from crystallization screening were upscaled and set manually on 24-well VDX plates (Hampton Research) using the hanging drop vapor diffusion technique. Crystallization drops were pipetted on a circular cover slide and consist of 1 µl of macromolecule sample mixed with 1 µl of crystallization cocktail. The volume of the reservoir in the crystallization well was 450-500 µl. Individual crystallization well was sealed with circular cover slide with crystallization drop, stored at 20°C, and inspected manually using Leica 125M stereomicroscope.

Chemical and reagents

PEG 2000 monomethyl ether (MME)	Hampton Research
PEG 1000	Sigma Aldrich
PEG 400	Sigma Aldrich
2-Methyl-2,4-pentanediol (MPD)	Sigma Aldrich
Calcium acetate	Qiagen
Lithium sulfate	Qiagen
Magnesium sulfate	Qiagen
Magnesium chloride hexahydrate	Qiagen
Potassium chloride	Qiagen
Lithium chloride	Qiagen
Sodium chloride	Qiagen
Hexamine Iridium (III) chloride	Homemade
Hexamine Cobalt (III) chloride	Hampton Research
DMSO	Sigma Aldrich
HEPES	Qiagen
Imidazole	Sigma Aldrich
Bis-Tris	Hampton Research
2-(N-morpholino)-ethanesulfonic acid (MES)	Hampton Research

Spermine tetrahydrochloride	Sigma Aldrich
Glycerol	Merck
Perfluoropolyether Cryo Oil	Hampton Research

2.2.2.3 Cryoprotection of the macromolecule crystal

Macromolecule crystals contain on average 50% of solvent. Because of that, they are often very fragile and they have to be protected before exposing them to the X-rays. Low-temperature X-ray diffraction methods, also called cryocrystallography, can minimize the rate of radiation damage to a crystal (Garman 2010) and increase its lifetime during data collection. During cryocrystallography, the macromolecule crystal is cooled to around 100 Kelvin. To minimize crystal damage during the cooling process, cryoprotectants are added to the crystallization solution to prevent crystalline ice formation (Rubinson, Ladner et al. 2000). Choosing a good cryoprotectant is one of the crucial steps to get a nice diffraction pattern and it takes many trials to find the perfect cryoprotectant that fits the sample.

The crystals were harvested from the crystallization drops using CryoLoops (Hampton Research) or LithoLoops (Molecular Dimensions) of the appropriate size, transferred to the cryoprotectant solution, and flash-frozen in liquid nitrogen. The loops were placed in previously cooled down UniPucks (Molecular Dimension) and transferred to a puck holder inside the transportation dry shipping dewar.

2.2.2.4 Heavy metal atom derivatization

In the perfect case scenario, in the structure determination process, we only need to collect a single dataset from the native crystal. Then, the model coordinates and electron density maps may be determined using the molecular replacement method. Unfortunately, in many cases, a molecular replacement cannot be used for initial phase determination. During the diffraction experiment, we record the intensities of reflections but the phase information gets lost. This is called a crystallographic phase problem and it is one of the bottlenecks in the structure determination procedure.

In the absence of a homology model, heavy-atom derivatization is one of the most popular methods for determining substructure location and obtaining phase information for macromolecule crystals. Typically, heavy metal salt or heavy metal cluster (eg. Tantalum Bromide Cluster) is derivatized via soaking or co-crystallization. During soaking, heavy metal is introduced

to the crystallization drop containing crystals. During co-crystallization, heavy metal is introduced to crystallization drop before crystallization of the macromolecule.

The heavy metal derivatization was transferring macromolecule crystal to the drop containing hexamine iridium (III) chloride or hexamine cobalt (III) chloride and cryoprotecting agent to stabilize the crystal during the soaking procedure. Soaking time varies from few minutes to few hours. After soaking, crystals were transferred to cryoprotection solution to remove excess heavy atoms. This procedure is called a back-soak and allows the removal of heavy atoms that do not contribute to the anomalous signal but increase the absorption coefficient and therefore the X-ray dose absorbed by the crystal.

2.2.2.5 Improving crystals quality by dehydration and ion exchange

A lot of RNA crystals diffract X-rays to a low-resolution range, which is insufficient to provide biochemical insights. To rescue low-quality crystals, many different post-crystallization treatments like annealing, dehydration, supplementing ligands have been developed to improve diffraction properties. During the dehydration process, we try to overcome the problems of loose packing of molecules and large solvent content which lead to low-resolution diffraction.

Metal cations are required for the proper folding of the DNA and RNA. Each nucleotide carries a negative charge on the phosphate group. As a result, folding of the nucleic acid results in a huge build-up of negative charges. To overcome this, many different mono- and divalent counterions are introduced within crystallization screening to compensate for the negative charge of the phosphate backbone and to promote proper folding and stabilize a folded nucleic acid structure.

Dehydration and ion exchange for VMC10 deoxyribozyme in complex with RNA substrate was performed using the already described protocol with slight modifications (Zhang and Ferre-D'Amare 2015). D1966/R665 crystals grown in original condition 0.2 M Lithium chloride, 0.05 M Bis-Tris pH 7.0, and 15-20% PEG 2000 MME were harvested and transferred to Crystal Treatment Solutions. Different concentrations of the primary precipitant – 20-40% PEG 2000 MME were tested to achieve a range of final solvent contents. D1965/R664 crystals grown in original condition 0.01 M magnesium chloride, 50 mM HEPES pH 7.0, and 4 M lithium chloride were harvested and transferred to Crystal Treatment Solutions. Different concentrations of the primary precipitant – 4-6 M of lithium chloride were tested to achieve a range of final solvent contents. D2020/R699 crystals grown in original condition, 0.05 M magnesium sulfate, and 50 mM HEPES pH 7.0 and 1.6 M lithium sulfate were harvested and transferred to Crystal Treatment Solutions. Different concentrations of the primary precipitant – 1.6-2.5 M of lithium sulfate tested to achieve a range

of final solvent contents. Additionally, different divalent cations in particular the alkaline earth metals - Mg^{2+} and Sr^{2+} - were used for post-crystallization treatment.

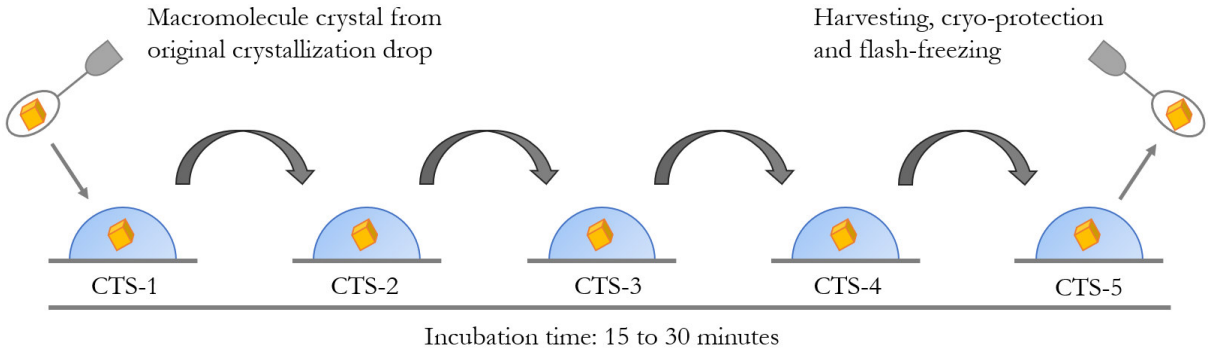


Figure 23. Dehydration and ion exchange procedure.

2.2.3 X-ray data collection and structure determination

A crystal structure is described as an ordered arrangement of atoms, ions, or molecules in the crystalline material. In the crystal, molecules are organized in the lattice, composed by the unit cell, which is described by its axes, or edges (a, b, c) and angles (α, β, γ). The asymmetric unit is the smallest part of a crystal structure to which symmetry operations can be applied to generate the complete unit cell (Figure 24). The most common symmetry operators to crystals of biological molecules are translations, rotation, and screw axes. Application of crystallographic symmetry operations to an asymmetric unit results in a complete unit cell that is further translated into three dimensions that build up the entire crystal. Combinations of rotation and translation are known as space groups. In macromolecular crystallography, there are 65 possible space groups.

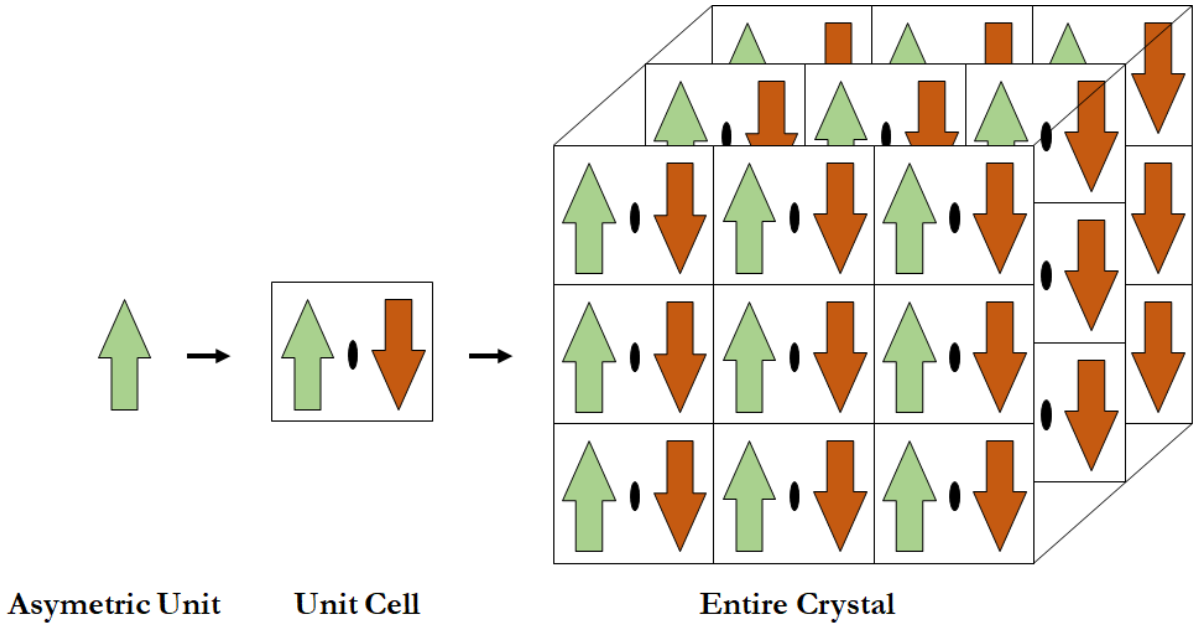


Figure 24. The basic concept of crystal symmetry.

The asymmetric unit molecule (green arrow) is rotated 180 degrees about a two-fold crystallographic symmetry axis to produce a symmetry-related molecule (red arrow). Together the two molecules comprise the unit cell. The unit cell is further translationally repeated in three directions to make an entire crystal.

2.2.3.1 Data collection

When the crystal is exposed to the X-ray beams with a fixed wavelength and at certain incident angles, and the scattered X-rays wavelengths interfere constructively, the intense reflected X-rays are produced. This physical phenomenon is called X-ray diffraction. For the waves to interfere constructively, the differences in the travel path must be equal to integer multiples of the wavelength. When this constructive interference occurs, a diffracted beam of X-rays will leave the crystal at an angle equal to that of the incident beam. The relationship between the wavelength of the incident X-ray beams, the spacing between the crystal lattice planes, and the incidence angle is known as Bragg's Law, and it is shown in Figure 25.

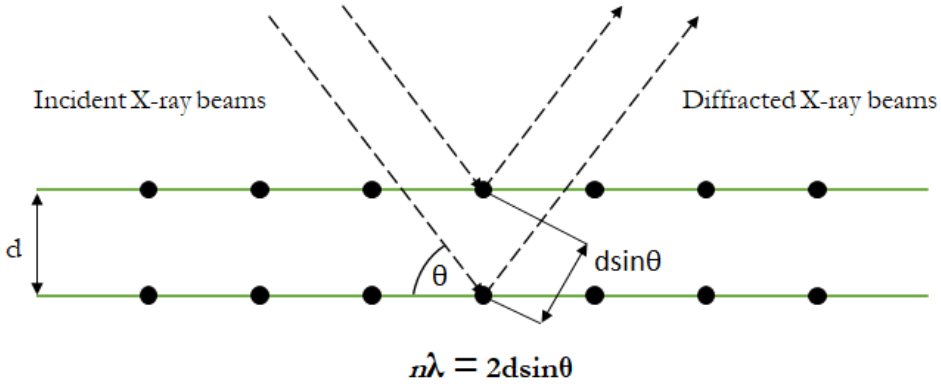


Figure 25. Bragg's Law.

When the two incident X-ray beams with identical wavelength and phase hit the crystal, the diffraction of that radiation occurs on each of its atoms. The Bragg condition assumes a deflection on the planes on which the crystal atoms are arranged. With the known interplanar distance and wavelength, Bragg's law determines the angle at which a wave must strike the crystal to interference constructively. In the final form of this law, n is a positive integer, λ is the wavelength of the incident wave, d is an interplanar distance and θ is the glancing angle.

X-ray sources and diffraction equipment

The reliable, stable beam of X-ray is a prerequisite for any X-ray diffraction experiments. Two general types of devices are used for the generation of X-rays: laboratory X-ray sources and synchrotron X-sources. Laboratory X-ray sources rely on the emission of characteristic radiation

from target materials after electron bombardment. In the synchrotrons, electrons are produced in the injector and are pre-accelerated in the booster ring and injected into the storage ring. The polychromatic X-rays are monochromated to a narrow band in the beamline optics and reach the experimental hutch with diffraction equipment. Additionally, the high brilliance of the X-ray beam allows the data collection from the very small crystals, which generally don't show diffraction in reasonable time on laboratory X-ray sources.

Diffraction data were collected at 100 K on EIGER16M and PILATUS 6M detectors at the X10SA (SLS – Swiss Light Source) or P11 (DESY - Deutsches Elektronen-Synchrotron) beamlines. Crystals were automatically mounted on the goniometer with an automatic sample changer. Diffraction images were visualized with ALBULA software. The data collection wavelength for native datasets was set to 1Å. The anomalous data collection was performed at Iridium L-III edge and Cobalt and Zinc K-edge.

2.2.3.2 Data processing and structure determination.

As mentioned before, crystals can be described by repeating units in the three-dimensional space. This space is known as real space. The real space can be described by the electron density - $\varrho(xyz)$, a function defined by each point of the unit cell of coordinates (xyz). When the X-ray interacts with the macromolecular crystal, and more precisely with the atoms forming a crystal, results in a diffraction pattern, also known as reciprocal space, with the properties of a lattice (reciprocal lattice) owning certain symmetry. Each point of this reciprocal lattice contains the information of the diffraction intensity.

Before the determination of the initial phases, diffraction images have to be reduced. Reflections from each diffraction image have to be processed into a single file. Crystallographic data processing is divided into several general steps:

a) Indexing and integration – During the indexing, it is possible to define the unit cell parameters and to determine the space group from the relative position of the different diffractions spots. During integration, the intensity of each reflection is determined.
b) Scaling and merging – During this scaling process, multiple observations of reflections are scaled together (measured more than once). Merging means combining multiple observations into an average intensity.

There are several criteria to estimate data quality:

a) R-merge – a measure of agreement between multiple measurements of a given reflection, used to assess data quality. An overall R-merge value should not exceed 10% for a good data set.
b) I/σI – refers to the signal to noise ratio. It indicates the strength of the signal. It results from dividing the averaged measured intensity of the reflections by the background noise.
c) Redundancy/Multiplicity – refers to multiple measurements of the same or symmetry equivalent, reflection. High redundancy allows lowering the measurement error of each reflection. A redundancy above 3.0 is acceptable.
d) Completeness – represents data collected compared to what is the unique data for the given crystal symmetry. The completeness should be at least over 90% (ideally 100%) for an overall dataset, and above 75 - 80% for the high-resolution shell.
e) CC ½ – Correlation Coefficient ½ - a percentage of correlation between intensities from a random half-datasets. CC ½ is the primary indicator for selecting a high-resolution cut-off for data processing. The rule of thumb is that data are usually insignificant if CC ½ below 15% or lower.

Crystallographic phase problem

Between the real and reciprocal there is a holistic relationship, where every detail in the direct space depends on the total information in the reciprocal space and vice versa. This relationship is known as the electron density function (Figure 26).

$$\rho\,(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} \underbrace{|F_{hkl}|}_{\text{Amplitudes}} e^{-2\pi i\,(hx+ky+lz-\underbrace{\phi\{hkl\}}_{\text{Phases?}})}$$

Real space — Reciprocal space

Figure 26. The function defining electron density within the unit cell is given by the coordinates (x,y,z).

In the formula above, F(hkl) represents diffracted beams of all atoms within the unit cell in a given direction. These magnitudes (amplitudes, actually waves), one for each diffracted beam are known as structure factors. The amplitude can be easily calculated by taking the square root of the intensity, but the phase is lost during the data collection. Because of that, it is not possible to solve the

equation directly, therefore the structure determination process is always associated with the "phase problem".

There are different methods to find initial phases in the macromolecular crystallography :

- Molecular replacement (MR) – uses the homologous structural model of sufficient identity, which structure has been solved.

-Single/Multiple Isomorphous Replacement (SIR/MIR) – uses heavy atom derivative and compares the diffraction pattern of the crystal to the diffraction pattern of the native crystal (without heavy atom).

- Single/Multiple Anomalous Dispersion (SAD/MAD) – uses the introduced heavy atom that behaves as anomalous scatters for the incident wavelength.

-SIRAS/MIRAS with Anomalous Dispersion – uses the combination of single/multiple isomorphous replacement and anomalous dispersion.

Molecular replacement

When a homologous model with sufficient conservation is available, this can be used to approximate initial phases. Experimental intensities are used for the Patterson map calculation and then compared to the theoretical Patterson map of the homologous model structure coordinates in different orientations. Then, a correctly oriented known model can be properly positioned by translating it to the correct coordinates within the asymmetric unit (Figure 27). When the model is placed, the electron density map is calculated using the experimental amplitudes of our molecule of interest, and the phases are obtained from the homology model molecule. The success in phase determination using the molecular replacement method is determined by the Translation Function Z-score (TFZ). For the TFZ above 8, we have a definite solution, between 7-8 probable solution and between 6-7 possible solution. The final quality of the electron map is verified by the experimenter.

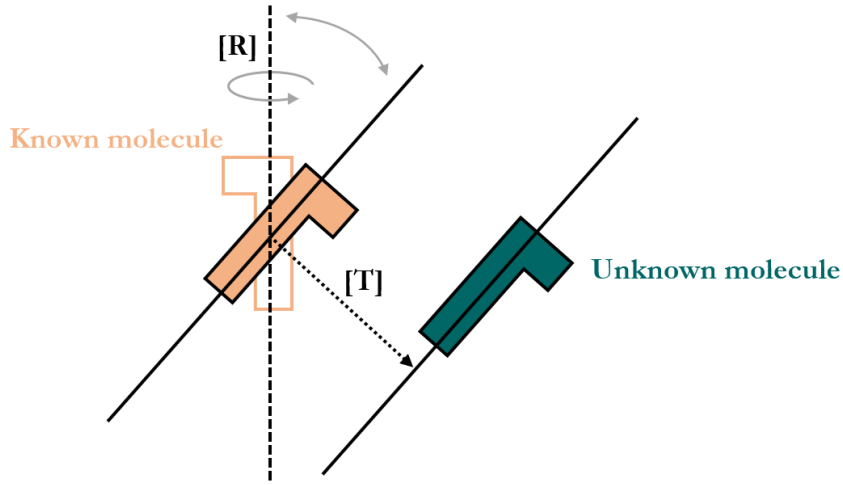


Figure 27. Molecular replacement (MR) method.

Single anomalous dispersion

The introduction of the heavy metal into the macromolecule crystal changes the intensity of the diffraction data. The atom is called an "anomalous scatterer" when the incident X-ray radiation has a frequency close to the natural vibration frequency of the electrons in a given atom and it is producing an energy transition, which is responsible for the anomalous dispersion effect. This fact is producing some changes in the atomic scattering factor, which is modified by the two terms – f' and f'' which corresponds to its real and imaginary components, respectively. The normal scattering factor f_0 is independent of the wavelength and is what contributes to the normal diffraction. The anomalous scattering factors f' and f'' changes with the wavelength. Atoms that contribute to the anomalous signal, its scattering factor is described in Figure 28.

$$f = f_0 + f' + if''$$

Figure 28. Total scattering factor equation.

When the anomalous scattering is present, f_0 has to be modified by the addition of f' and f''.

The dispersive term f' modifies the normal scattering factor, whereas the absorption term f'' is 90° advanced in phase. In the presence of anomalous scatter Friedel's Law breaks down ($|FPH(+)| \neq |FPH(-)|$), bringing up anomalous differences that can be used to locate anomalous scatters (Taylor 2010) (Figure 29a).

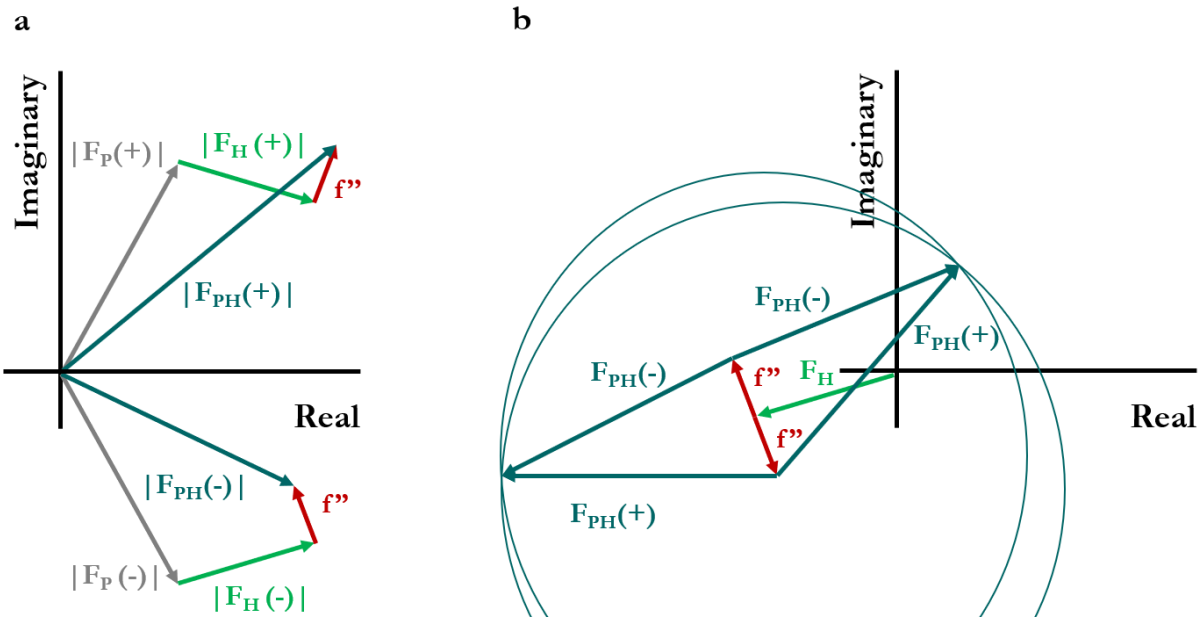


Figure 29. Overview of the single anomalous dispersion (SAD) experiment.
Adopted with modification from (Taylor 2010).

The single anomalous dispersion experiment only provides measurements of the anomalous or Bijvoet differences $\Delta F\pm=|F_{PH}(+)|-|F_{PH}(-)|$. The differences allow estimating the heavy-atom contribution to the scattering. Then, heavy-atom substructure can be derived using direct or Patterson methods. The SAD experiment can be illustrated using the Argand diagram to draw Harker construction (Figure 29b). A single reflection from a hypothetical SAD experiment shows that once the heavy-atom substructure is known, the calculated amplitude and phase of this contribution can be drawn (FH). Afterward, the phase ambiguity is broken through density-modification procedures (Taylor 2010).

All data were indexed, integrated, and scaled with XDS (Kabsch 2010) and reduced with Pointless, Aimless, and Ctruncate within the CCP4 package (Evans 2011).

2.2.3.2.1 Crystal structure determination of the Chili RNA aptamer

The structure of Chili RNA Aptamer was solved using SAD data collected at the Iridium L-III peak wavelength. Crystals containing heavy metal were prepared as described in section 2.2.2.4. Initial phases were determined by single anomalous dispersion (SAD) using AutoSol Wizard in Phenix (Adams, Afonine et al. 2010). Electron density map from AutoSol enabled the automated building of an initial model with AutoBuild module in Phenix (Terwilliger, Grosse-Kunstleve et al. 2008). This initial model was modified by iterative rebuilding in COOT (Emsley and Cowtan 2004) and refinement with Phenix.refine (Afonine, Grosse-Kunstleve et al. 2012). The partially built model from the Ir co-crystallized crystal was used as a molecular replacement search model

in Phaser against a high-resolution native dataset (Chili-DMHBO+), previously corrected for anisotropy using the STARANISO server (Tickle 2018). The model was completed by manual rebuilding in COOT. The location of the $DMHBO^+$ ligand was confirmed using both unbiased mFo-DFc and polder omit maps (Liebschner, Afonine et al. 2017). Ligand restraints were prepared using eLBOW in Phenix. The final structures of Chili-DMHBI+ and Chili DMHBO+ (Ir soak) were solved by molecular replacement using a single 51 nucleotides long RNA model built with native Chili-DMHBO+ data.

2.2.3.2.2 Crystal structure determination of VMC10 non-cleavable RNA substrates – R616, R633 and R698

Structures of VMC10 RNA substrates (R616 and R633) were solved using SAD data collected at the Cobalt and Zinc K-edge peak wavelength. In the case of the R616 structure, derivatization of heavy metal was performed using a soaking procedure, described in section 2.2.2.4. In the case of the R633 structure, Zinc ion was present in the original crystallization condition. Initial phases were determined by single anomalous dispersion (SAD) using AutoSol Wizard in Phenix. Electron density map from AutoSol enabled manual model building in COOT. Models were modified by iterative rebuilding in COOT and refinement with Phenix.refine. The model built from the Co derivatized crystal was used as a molecular replacement search model in Phaser against a native dataset.

The structure of the third VMC10 RNA substrate R698 was solved using molecular replacement using the solution structure of the RNA hairpin loop (PDB code: 1HS8) as a search model in Phaser. Phaser found 4 MR solutions with TFZ=7.1 and LLG=96.15. Electron density map from AutoSol enabled the automated building of an initial model with AutoBuild module in Phenix. This initial model was completed by iterative rebuilding in COOT and refinement with Phenix.refine.

2.2.3.2.3 Crystal structure determination of 9DB1 mutant

The structure of 9DB1 mutant in complex with ligated RNA product was solved using molecular replacement using a crystal structure of 9DB1 deoxyribozyme (PDB code: 5CKK) as a search model in Phaser. Phaser found 1 MR solution with TFZ=16.9 and LLG=243.79. This initial model was modified by iterative rebuilding in COOT and refinement with Phenix.refine.

2.2.3.3 Structure refinement

The initial model and electron density map obtained from the experimental or molecular replacement phases are not usually accurate. This is due to the fact, that the atomic positions from

the initial solution are not the direct result of the diffraction experiment, but the interpretation of the electron density function calculated from measured intensities and pre-determined phase angles. Better phases can be calculated from the atomic positions, which allows to re-determine electron density function with higher precision. Refinement is the iterative process, wherefrom the new electron density map, more precise atomic positions can be derived, which results in better phase angles and so forth. In every cycle of the refinement, we introduce changes to the atomic model: changing atom types, introducing missing atoms, and removing subsidiary atoms. The model parameters that are adjusted by the refinement software includes, for each atom, its [x,y,z] coordinates, B-factors (ADPs – atomic displacement parameters or temperature factor) which reflects atom mobility or smearing in place, and occupancies which represents the fraction of the particular atom in space. The simulated annealing approach is also very common, where the whole model is "heated" and then "cooled down" to get a less distorted model.

One of the most important criteria which reflects model quality is the crystallographic R-factor (R_{work}) – a measure of agreement between the model and the original diffraction data. The crystallographic R-factor is calculated using the following formula:

$$R = \frac{\Sigma||Fobs|-|Fcalc||}{\Sigma|Fobs|}$$

Equation. 1.

Here, each $|F_{obs}|$ is derived from a measured intensity of a reflection in the diffraction pattern, and $|F_{calc}|$ is the intensity of the same reflection calculated from the model. The random set of atoms will give the R-factor value of about 0.6, where a perfect fit would have a value of 0. In a real case, the perfect fit is not possible and typical values are around 0.2.

To avoid possible over-modeling of the data, there was a necessity to introduce a cross-validation method, based on the calculation of the second agreement factor, known as Free R-factor (R_{free}). The free R factor is calculated in the same manner as R-factor, but only a small fraction of randomly chosen reflections (also known as a test set), which are set aside during refinement. During refinement, R_{free} measures how well the current atomic model predicts a subset of the measured reflection intensities that were not included in the refinement, whereas R measures how well the current model predicts the entire data set that produced the model. For the perfect model which is not over-interpreting the data, the Free R-factor will be similar to the R-factor. Typically, it is a little higher, with a value of about 0.26.

Structure refinement was done using phenix.refine software. The R_{work} and R_{free} factors were monitored during the refinement to determine the proper refinement strategy. The refinement procedure includes refinement of [x,y,z] coordinates, B-factors, TLS parameters (refines anisotropic displacements for large groups of atoms, typically individual domains or chains), and occupancies. The targets of refinements included optimization of geometry and ADP weights, NCS-restraints, and reference model restraints if necessary.

The structural model is considered as a final, when it matches its electron density map and when "good" R_{work} and R_{free} values are achieved. As the rule of thumb, the good R_{free} for models with a resolution of 2.0 Å or better, R_{free} should not exceed (resolution/10) by more than 0.05. In practice, if the resolution is 2.0 Å, R_{free} should not exceed 0.25. For resolutions around 3.0 Å, the R_{free} should not exceed the resolution of the data divided by 10. The final model should also obey minimal requirements for correct geometry – stereochemistry, clashes between atoms, torsion angles, and so on. Additionally, a structural model is validated by MolProbity (Chen, Arendall et al. 2010) software implemented in the refinement program.

3 RESULTS

3.1 STRUCTURAL INVESTIGATION OF THE RNA-CLEAVING VMC10 DEOXYRIBOZYME

Nucleic acid molecules can be crystallized using combinations of methods originally developed for protein crystallography. Additionally, The DNA and RNA can be easily synthesized and purified in quantities and with the quality required for macromolecular crystallography. One of the bottlenecks in nucleic acid crystallography is to obtain high-quality crystals with good diffraction properties. On the one hand, the formation of the high-quality crystal may be difficult, because nucleic acids have few sites for making intermolecular contacts. Thus, nitrogenous bases are mostly forming base pairs and exceptionally form intermolecular contacts. On the other hand, the negatively charged backbone may lead to unspecific binding sites and packing, which may introduce defects in the crystal lattice (Mooers 2009). To overcome those limitations, the rational design of the crystallization construct is necessary. For instance, double-stranded oligonucleotides can be designed with blunt or sticky ends to promote the formation of a semi-continuous helix in the crystal lattice (Shah and Brunger 1999, Wedekind and McKay 1999). Additionally, the length and the sequence are varied to identify constructs that may form crystals. The various constructs are tested for their biochemical activity and subjected to crystallization screening. The construct design strategy is presented in Figure 30. Thus, to identify DNA-RNA complexes that form well-diffracting crystals, I designed variations of constructs with blunt or sticky ends (a), modifications of the sticky end (b), different ratios of the binding arms (c), and different binding arms lengths (d).

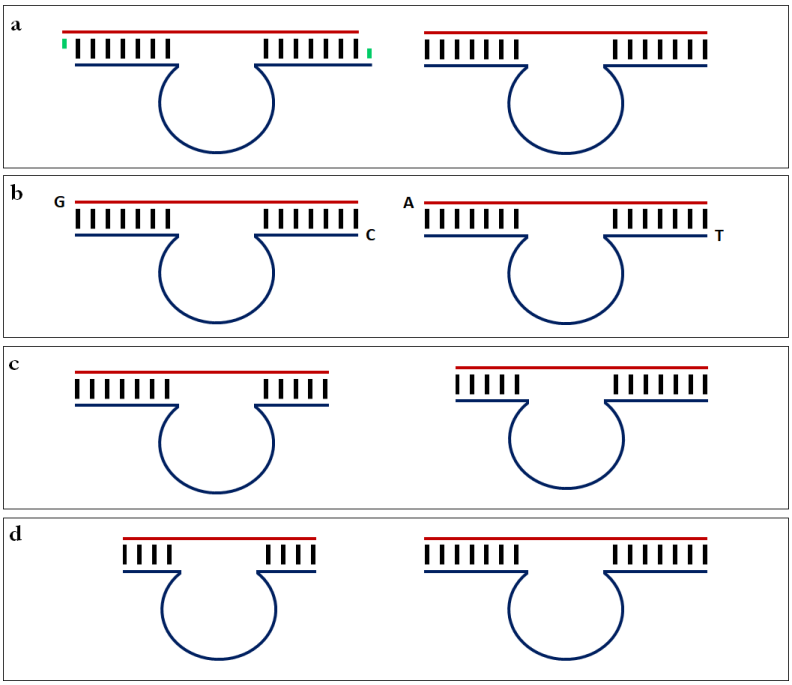


Figure 30. Rational construct design strategy.

To investigate the structure of the N6-methyl sensitive RNA-cleaving VMC10 deoxyribozyme, I crystallized the DNAzyme with its RNA substrate, thus mimicking the pre-catalytic state of reaction. The RNA strand contains the cleavage site, which is positioned between guanosine and adenosine (Figure 31). To trap the VMC10 deoxyribozyme in its pre-catalytic state for crystallization, a single nucleotide mutation in the cleavage site was introduced to inhibit the cleavage of the RNA substrate. With the loss of its 2'-hydroxyl, guanosine in the cleavage site can no longer serve as a nucleophile for VMC10 to carry out phosphoester transfer reaction.

3.1.1 Design and crystallization of VMC10 variants

To investigate the crystal structure of the VMC10 deoxyribozyme I have designed five generations of the DNAzyme in complex with its substrate. Constructs from generations 1-3 yielded in the structure determination of the RNA substrates. To overcome the dissociation of RNA from the deoxyribozyme, I have designed two new variants of the VMC10. Constructs from generations 4 and 5 produced crystals of VMC10 in complex with RNA substrate diffracting to 3.85 Å. Additionally, the RNA substrate from a single construct dissociated from the DNAzyme during the crystallization and formed the RNA crystals.

3.1.1.1 First generation of the VMC10 DNAzyme constructs

For the initial crystallization screening, I prepared two variants of constructs. The 15 nucleotides long RNA substrate with corresponding DNAzyme binding arms was used to generate complex with "blunt" and "sticky" ends (Figure 31).

Figure 31. The first generation of the VMC10 DNAzyme constructs.

(a) D1739/R597 (b) D1740/R597 complexes used for initial crystallization screening

Crystallization screening

Complexes from the first-generation of VMC10 constructs, prepared as described in section 2.2.1.1, were subjected to crystallization screening at 20 °C. Four commercially available crystallization screens – Nucleix, Natrix 1+2, and MPD suite were used for initial screening. The crystallization drops consist of 60 nl of a 2 mM complex solution and 60 nl of crystallization buffer. D1739/R597 produced initial crystals in condition A6 of the MPD Suite crystallization screen after one week.

Optimization of the crystallization conditions of the first-generation of VMC10 constructs

D1739/R597 complex at 2 mM concentration was mixed with solutions containing a gradient of MPD concentration (30-45%) and 0.2 M sodium chloride. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 1 µL of DNAzyme complex and 1 µL of crystallization buffer. Crystallization was performed using-hanging drop vapor diffusion method at 20°C.

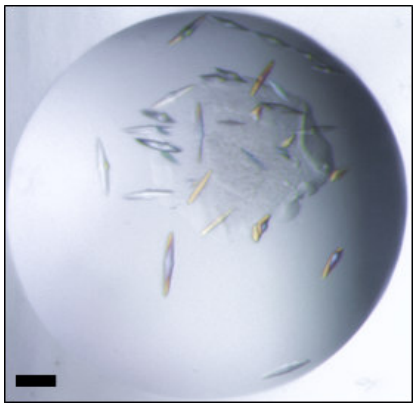


Figure 32. Crystals produced in the A6 MIDAS condition.

D1739/R597 complex, crystallization conditions: 35% MPD, 0.2 M sodium chloride. The black bar represents the length of around 100 μm.

Diffraction experiments

The crystals of the VMC10-substrate complex (D1739/R597) were harvested from the drop, cryoprotected with crystallization solutions with 40% MPD, and flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K on PILATUS 6M detectors at the X10SA (SLS). The crystals of the D1739/R597 do not diffract. An example of the diffraction image is shown in Figure 33.

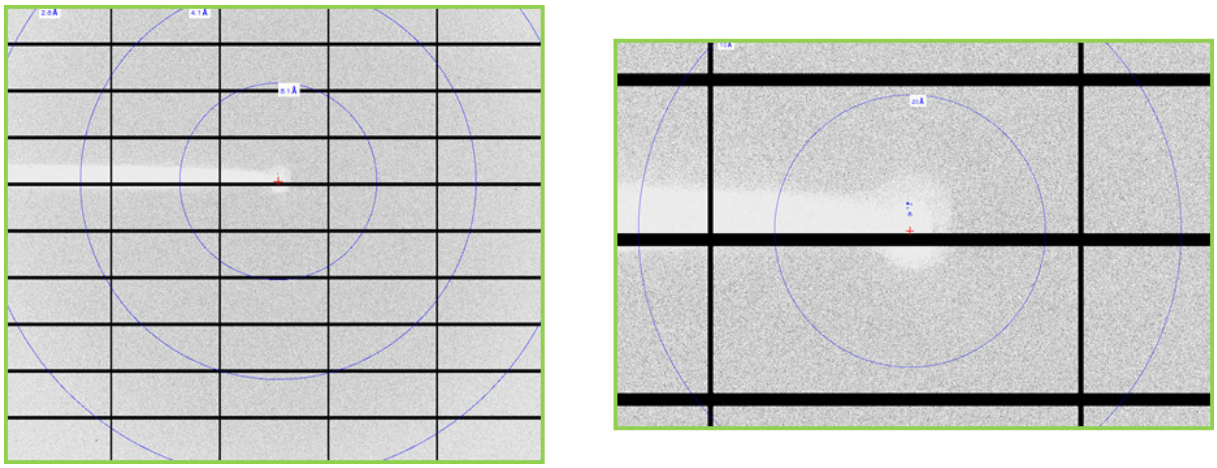


Figure 33. Diffraction image of the D1739/R597 crystals (left) and its close-up view (right).

3.1.1.2 Second generation of the VMC10 DNAzyme constructs

The first generation VMC10-substrate constructs did not form diffracting crystals, therefore the second generation of VMC10 constructs was generated. I designed four new RNA substrates with different lengths, varying from 13 to 16 nucleotides. Additionally, the adenosine at position 20 in the catalytic core was mutated to guanosine, to form an additional base pair next to the cleavage site (Figure 34).

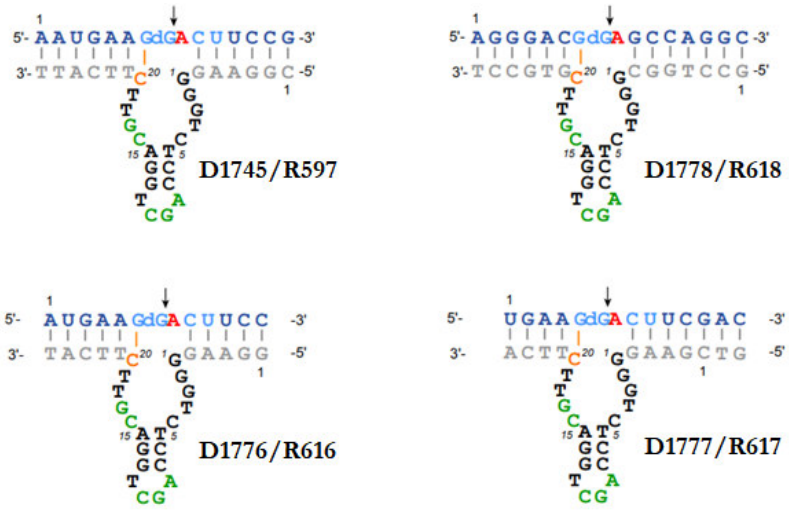


Figure 34. The second generation of the VMC10 DNAzyme constructs.

Crystallization screening

Crystallization screening of the third generation of VMC10 and VMA deoxyribozymes complexes was performed in the same way as for the second generation of DNAzyme constructs. Four commercially available crystallization screens – Nucleix, Natrix 1+2 and Wizard Classic 1+2, and Wizard Classic 3+4 were used for initial screening. The crystallization drops consist of 120 nl of a 1 mM complex solution and 120 nl of crystallization buffer. Construct D1776/R616 produced initial crystals in condition A12 of the Wizard Classic 1+2 crystallization screen after one-two days.

Optimization of crystallization conditions of the second generation of VMC10 constructs

D1776/R616 complex at 1 mM concentration was mixed with solutions containing gradient of PEG 1000 concentration (10-30%), 0.2 M calcium acetate and 0.1 M imidazole pH 8.0. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 1 μL of DNAzyme complex and 1 μL of crystallization buffer. Crystallization was performed using the hanging drop vapor diffusion method at 20ºC.

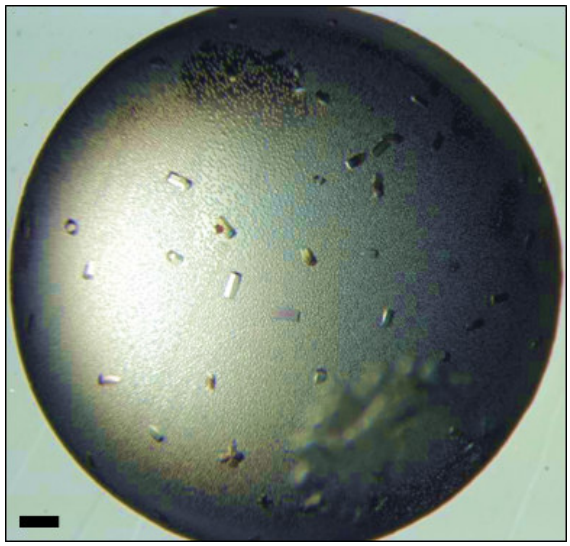


Figure 35. Crystals produced in the A12 condition of Wizard Classic 1+2.

D1776/R616 complex, crystallization conditions: 20% PEG 1000, 0.2 M calcium acetate and 0.1 M imidazole pH 8.0. The black bar represents the length of around 150 μm.

3.1.1.3 Third generation of the VMC10 DNAzyme constructs

Given the high potential of the D1776/R616 construct from the previous generation, I decided to generate new constructs, with the m6A mutation in the cleavage site and by introducing new DNAzymes – VMA8 and VMA15 (Figure 36).

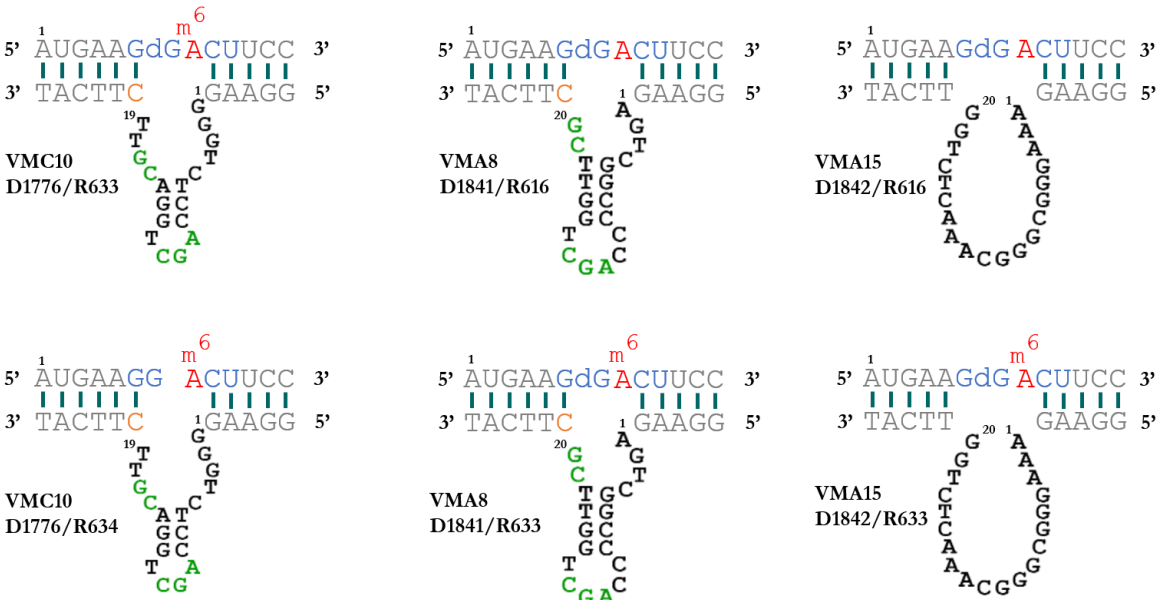


Figure 36. The third generation of the VMC10 DNAzyme constructs.

Crystallization screening

I performed crystallization screening of the third generation of VMC10 and VMA deoxyribozymes complexes in the same way as for the second generation of DNAzyme constructs. The complexes D1841/R616 and D1842/R616 formed the crystal in the same conditions as D1776/R616 – A12 Wizard Classic 1+2 (20% PEG 1000, 0.2 M calcium acetate, and 0.1 M imidazole pH 8.0), with very similar crystal morphology (hexagonal-shaped rods). The complexes D1776/R633 and D1842/R633 crystallized in condition D12 (20% PEG 1000 0.2 M zinc acetate 0.1 M sodium acetate pH 4.5) and H4 (20% PEG 3000 0.2 M zinc acetate and 0.1 M sodium acetate pH 4.5) of Wizard Classic 1+2 crystal screen respectively.

Optimization of crystallization condition of the third generation VMC10 constructs

D1841/R616 and D1842/R616 complexes at 1 mM concentration were mixed with solutions containing a gradient of PEG 1000 concentration (10-30%), 0.2 M calcium acetate, and 0.1 M imidazole pH 8.0. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 1 µL of DNAzyme complex and 1 µL of crystallization buffer. Crystallization was performed using the hanging drop vapor diffusion method at 20ºC.

D1842/R633 complex at 1 mM concentration was mixed with solutions containing gradient of PEG 3000 concentration (10-30%), 0.2 M zinc acetate and 0.1 M sodium acetate pH 4.5. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 1 µL of DNAzyme complex and 1 µL of crystallization buffer. Crystallization was performed using the hanging drop vapor diffusion method at 20ºC. However, I could not reproduce crystals during the optimization process.

3.1.1.4 Heavy metal derivatization

Heavy metal derivatization was performed as described in section 2.2.2.4. Before the anomalous data collection, D1776/R616 crystals were soaked with cobalt (III) hexamine (1-10 mM) for a few minutes, back-soaked, and frozen in liquid nitrogen. The zinc ion present in the crystallization conditions of D1842/R633 serves as an anomalous scatter, therefore derivatization with other heavy metals was not necessary.

3.1.2 Data collection and structure determination of the second and third generation of the VMC10 DNAzyme constructs

Diffraction data of the second and third generation of VMC10 DNAzymes constructs were collected at 100 K on PILATUS 6 M detector at the X10SA (SLS). The anomalous data of the D1776/R616 crystals soaked with cobalt (III) hexamine was collected at the cobalt K-edge peak wavelength (Table 4). The anomalous data of the D1842/R633 was collected at zinc K-edge peak wavelength (Table 5). Data were processed as described in section 2.2.3.2.

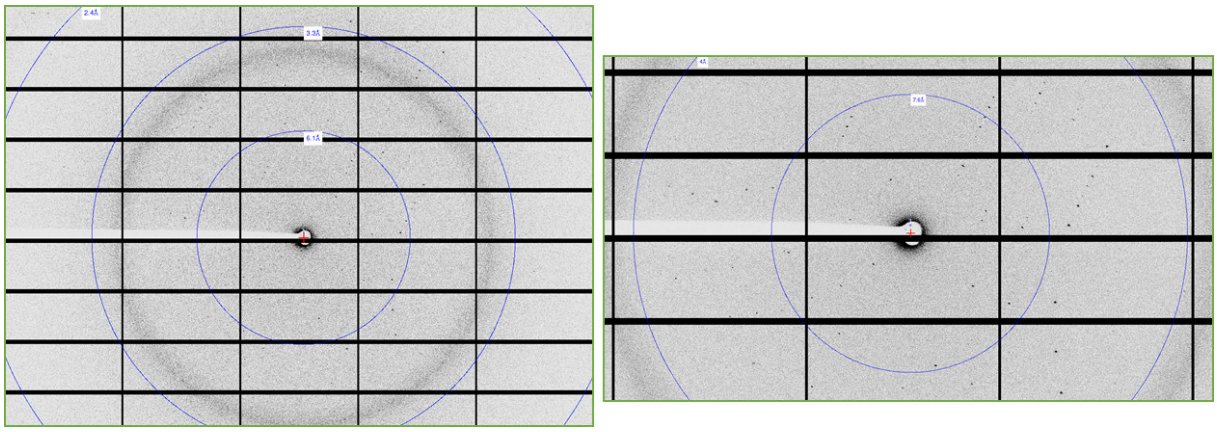


Figure 37. Diffraction image of the D17776/R616 crystals (left) and its close-up view (right).

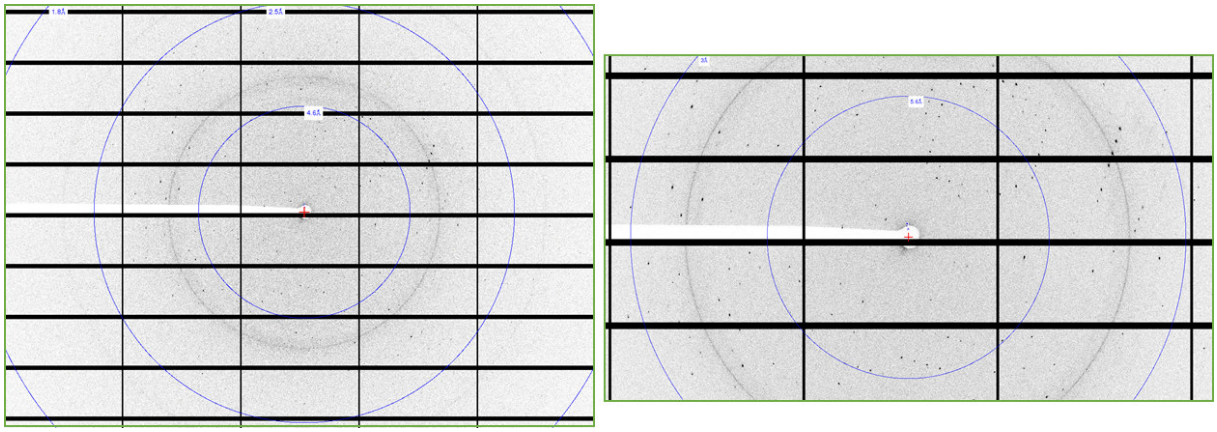


Figure 38. Diffraction image of the D1842/R633 crystals (left) and its close-up view (right).

Initial phases were determined by single anomalous dispersion (SAD) using AutoSol Wizard in Phenix using anomalous data collected at cobalt and zinc wavelength. Electron density map from AutoSol enabled the automated building of an initial model with AutoBuild module in Phenix (Terwilliger, Grosse-Kunstleve et al. 2008, Terwilliger, Adams et al. 2009). The initial models were modified by iterative rebuilding in COOT (Emsley and Cowtan 2004)and refinement with Phenix.refine(Afonine, Grosse-Kunstleve et al. 2012). Data collection and refinement statistics are summarized in Tables 4 and 5.

3.1.3 VMC10 RNA substrate structure analysis

During the model building and refinement, it turned out that the RNA substrate dissociates from the DNAzyme during the crystallization and formed RNA-only containing crystals. however, I decided to refine structures, as this can provide interesting insights about the folding of nucleic acids and/or bringing new insight into how nucleic acids packs in the crystal lattice.

3.1.3.1 Structure analysis of the VMC10 RNA substrate (R616)

Overall crystal structure

The crystallographic asymmetric unit (ASU) contains 2 copies of 13-nucleotide long VMC10 RNA substrate (Figure 39a). In the unit cell, individual copies of the asymmetric unit molecules are forming two, A-form double helices (RNA duplex) with their symmetry mates (Figure 39b). The structures of individual RNA duplexes in the unit cell are very similar, as they can be superimposed with a root-mean-square deviation of 0.57 Å (Figure 40).

The VMC10 RNA substrate folds into a single A-form helix with a length of approximately 50 Å. The RNA duplex contains two, consecutive non-canonical dG7:A8 base pairs, in which the sugar edge of dG7 contacts the major groove of the A8 (Figure 41). The basepairs dG7:A8 are flanked by the stems, containing four canonicals base pairs (nucleotides G6:C9, A5:U10, 4A:11U, and G3:C12), overall forming ten (cWW/tHS) hydrogen bonds. There are three unpaired, additional nucleotides, two at the 5' site – A1 and U2, one on the 3' site – C13, which are not forming any hydrogen within the single RNA duplex.

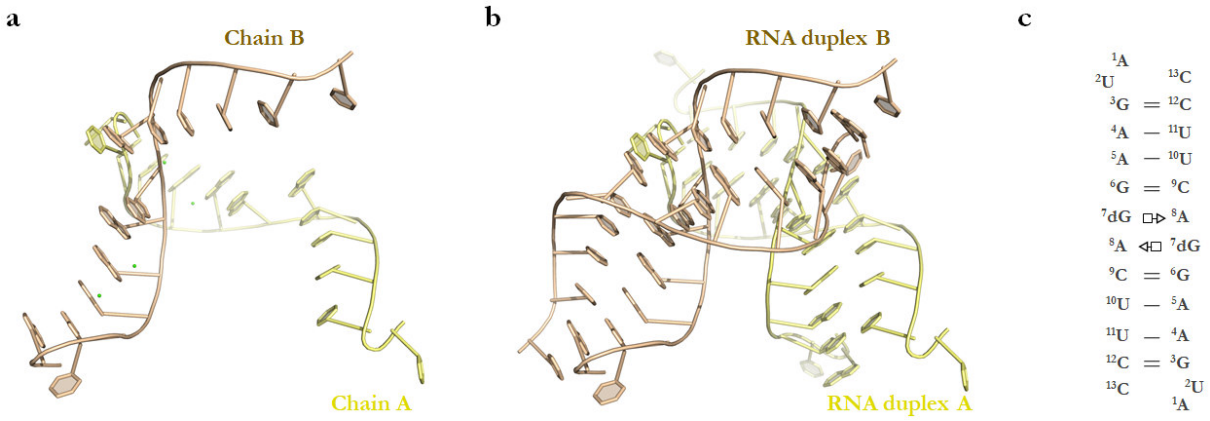


Figure 39. Overall crystal structure of the R616 VMC10 RNA substrate.

(a) Two 13-nucleotide long RNA molecules forming the asymmetric unit. (b) Two double-stranded RNA A-form helices present in the unit cell. The RNA duplexes A and B are colored in yellow and beige respectively. (c) Organization of the nucleotides in the single RNA duplex. The single and double bars

represents WC base pairs. The symbol between 7dG and 8A represents the trans Hoogsteen/Sugar edge base pair (tHS).

Table 4. Data collection and refinement statistics of the R616.

	Non-cleavable VMC10 RNA substrate (R616)	**Non-cleavable VMC10 RNA substrate Cobalt derivative (R616)**
Data collection		
Wavelength (Å)	1.000	1.604
Resolution range	45.95 - 2.80 (2.95 - 2.80)	46.10 - 2.40 (2.49 - 2.40)
Space group	$P6_2$ 2 2	$P6_2$ 2 2
Unit Cell		
a, b, c (Å)	53.06 53.06 100.91	53.24 53.24 100.61
α, β, γ (°)	90 90 120	90 90 120
R-merge	0.093 (>1.0)	0.114 (>1.0)
No. reflections	76051 (8100)	122133 (8367)
No. unique	2371 (331)	3669 (358)
I/σI	8.2 (0.4)	19.3 (1.9)
Multiplicity	26.1 (1.4)	22.1 (0.4)
Completeness (%)	99.8 (99.7)	99.8 (99.5)
CC1/2	1.000 (0.879)	1.00 (0.561)
Refinement		
Resolution range	45.950 - 2.851 (2.952 - 2.851)	46.104 - 2.406 (2.492 - 2.406)
R-work/R-free (%)	21.38/25.30	21.49/23.62
Reflections used in refinement	2126 (216)	3477 (262)
No of atoms	550	600
RNA	546	546
Ligands/Ions	4	43
Water	-	11
r.m.s deviations		
Bond lengths	0.002	0.002
Bond angles	0.41	0.55
Average B-factor	109.09	105.53
RNA	108.82	100.49
Ligands/Ions	147.00	172.38
Water	-	94.19

*** Statistics for the highest-resolution shell are shown in parentheses**

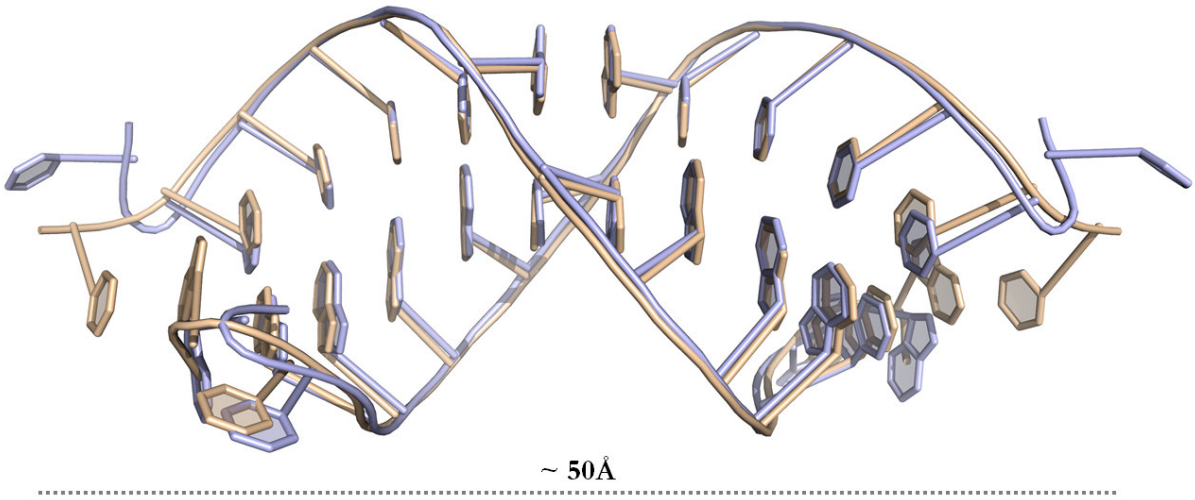


Figure 40. Superimposed models of the RNA duplexes.

(a) Superimposed single copies of RNA duplexes. Individual copies of RNA duplexes A and B are colored in blue and beige, respectively. The grey dashed line indicates the approximate length of the RNA duplex.

Characterization of the sheared dG:A base pair

The two, consecutive dG7:A8 base pairs are formed by trans Hoogsteen/Sugar edge hydrogen-bonding pattern. The Hoogsteen edge of A8 is bonding with the minor groove, sugar edge of dG7. In the RNA duplex A (Figure 41a), a single hydrogen bond is formed between N2 of dG7 and N7 of A8 with the N-N distance of 3.2 Å. In the RNA duplex B (Figure 41b), the distance between N2 of dG7 and N7 of A8 is a little bit shorter, with a distance of 3.0 Å. Additionally, a new hydrogen bonding interaction is formed between N3 of dG7 and N6 of A8, with the N-N distance of 3.6 Å. Both dG7 and A8 nucleotides forming tHS base pair possess C3'-*endo* sugar puckers conformation.

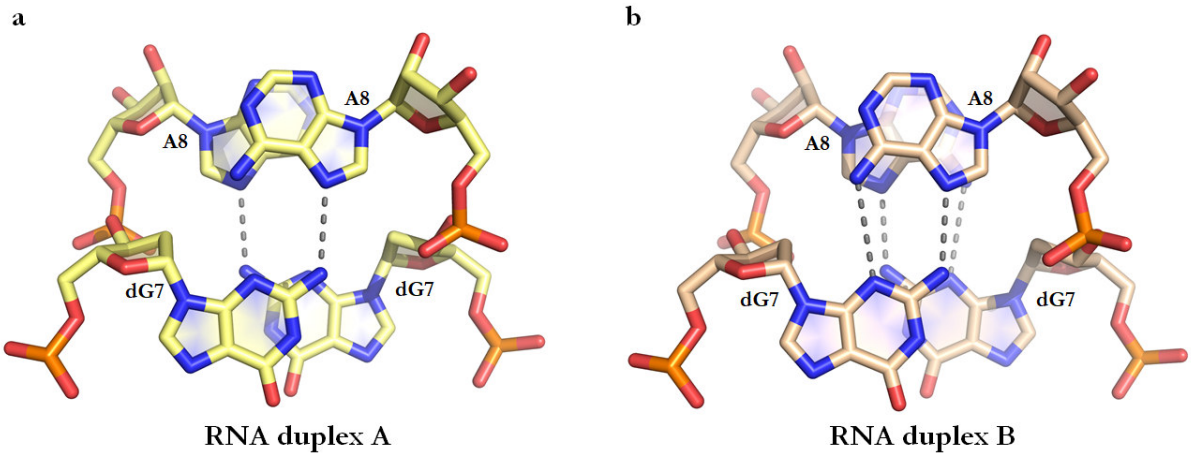


Figure 41. The non-WC trans Hoogsteen/Sugar edge base-pairing of the dG7:A8.

The tHS dG7:A8 base pairs present in the RNA duplex A (a) and RNA duplex B (b).

Crystal packing of the R616 VMC10 RNA substrate

The previously characterized unpaired nucleotides – A1, U2, and C13 turned out to be responsible for the crystal packing interactions. Inside the unit cell, the two RNA duplexes (Figure 42) are oriented to each other by the 90-degree rotation and held in place by the two DNA/RNA triplets and two hydrogen bonds between 2'-OH and phosphate groups. An additional interaction between the A1 and the C13 allows the forming of a semi-continuous helix in the crystal (Figure 44a,b). The DNA/RNA triplets are formed by the dG7:A8 base pairs with U2 from the neighboring molecule forming a canonical, trans Watson-Crick base pair, with the distances between N1/ N6 of the A8 and N3/O2 of the U2 with an average distance of 2.9 Å (Figure 44a,b). All nucleotides forming the DNA/RNA triplet are in the *anti* – conformation. Two intermolecular hydrogen bonds between 2'-OH of G6 and OP1 of G3 are in the average distance of 2.35 Å (Figure 44a,b).

An intermolecular interaction, which allows forming a semi-continuous structure is formed between N6 of A1 and OP1 of C13 from the neighboring molecule, with a distance of 3.0 Å (Figure 44c). The A1 adopts a *syn* – conformation with the torsion angle χ of 95°.

Figure 42. Crystal packing of the R616 VMC10 RNA substrate.

The RNA duplex A (teal) is forming a semi-continuous helix in the crystal lattice. Intermolecular contacts formed by the RNA duplex B (red) are connecting semi-continuous helix formed by the RNA duplex A.

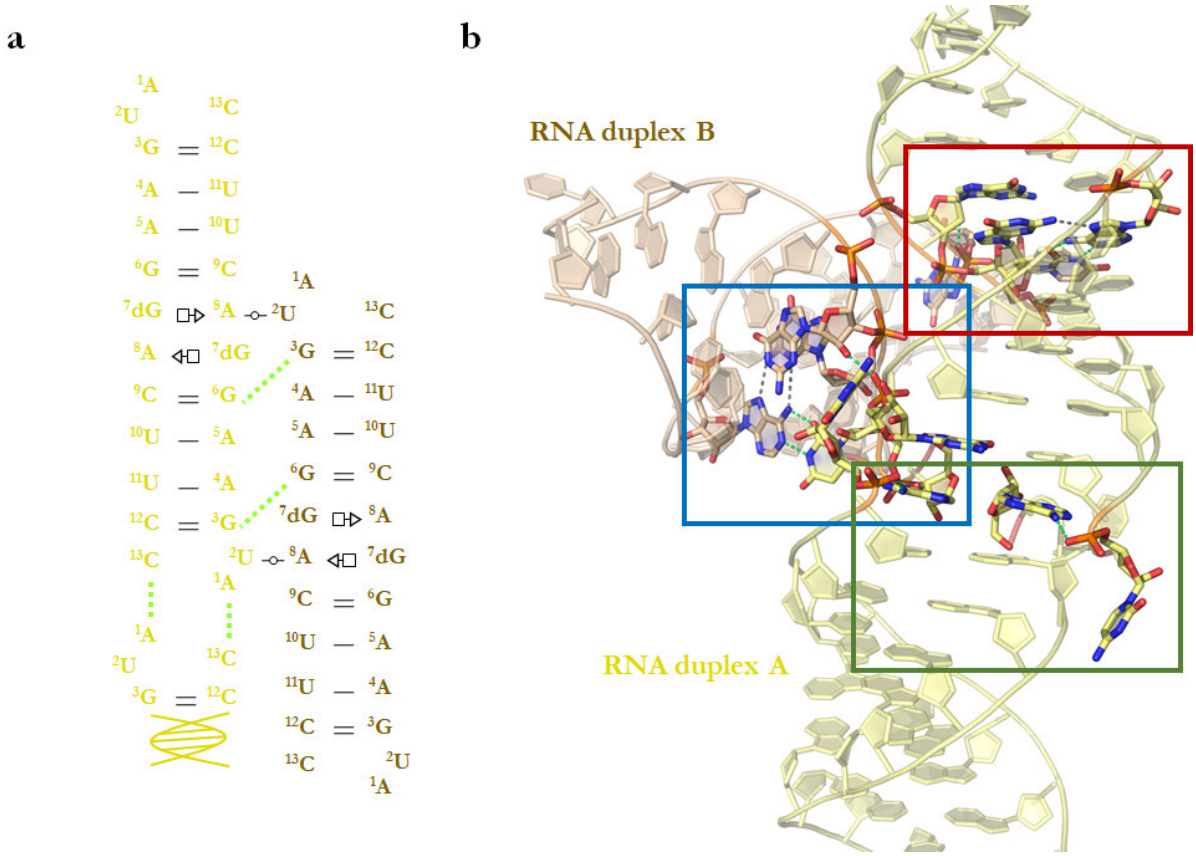


Figure 43. Intermolecular contacts formed by the R616 VMC10 RNA substrate.

(a) Secondary structure model of the double-stranded RNA with intermolecular contacts between individual duplexes. The RNA duplexes A and B are colored in yellow and beige, respectively. Two DNA/RNA triplets and two hydrogen bonds between 2-OH groups and phosphate oxygens are responsible for 90° rotation orientation. The interaction between A1 and C13 allows for the formation of a semi-continuous helix structure in the crystal (b) Intermolecular contacts between the RNA duplexes. Dashed green lines represent intermolecular hydrogen bonding interactions between the RNA duplexes.

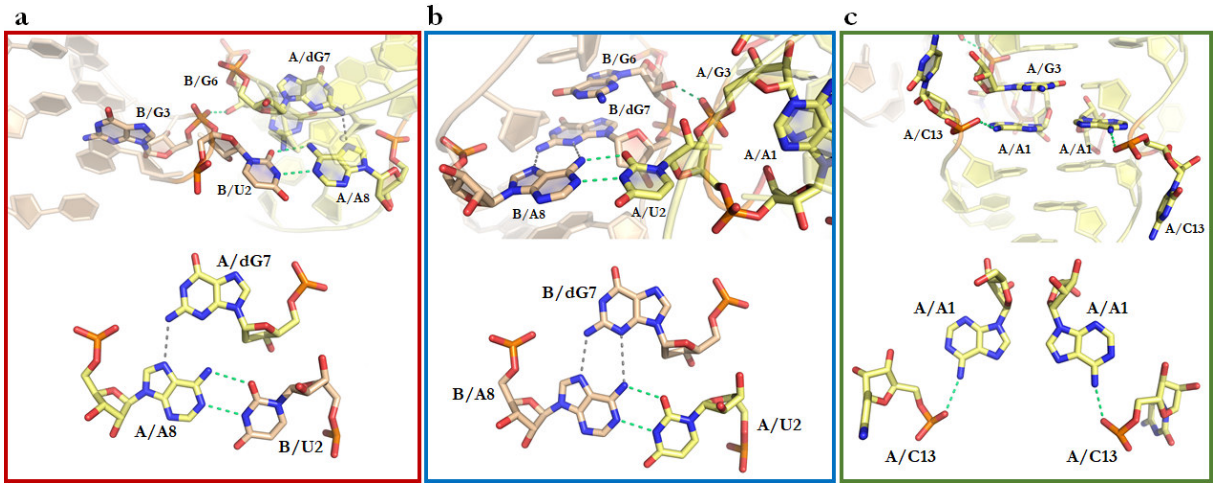


Figure 44. The close-up view for the intermolecular contacts between neighboring RNA duplexes.

(a,b) Close up view for the intermolecular contacts between dG7:A8 base pairs and U2 forming DNA/RNA triplets. Two intermolecular hydrogen bonding interactions between 2'OH of G6 and oxygen-phosphate (OP1) of G3 helps to stabilize two RNA duplexes (c) Close up view for the intermolecular contacts between A1 and C13. This interaction allows forming a semi-continuous helix structure in the unit cell. Dashed green lines represent intermolecular hydrogen bonding interactions between the RNA duplexes.

3.1.3.2 Structure analysis of the methylated VMC10 RNA substrate (R633)

Overall crystal structure

The crystallographic asymmetric unit (ASU) contains 6 copies of 13-nucleotide long methylated, VMC10 RNA substrate, containing N6-methyladenosine modification RNA substrate (Figure 45). In the unit cell, four copies of the asymmetric unit molecules (Chains – A, B, C, and D) are forming tetrameric-RNA complex structure. The second tetrameric-RNA complex is formed by the two asymmetric unit molecules (Chains – E and F) with its symmetry mates. In the tetrameric-RNA, we can distinguish two, triple-RNA helixes formed by chains A, B, C and A, B, D (Figure 46b). The structures of individual tetrameric-RNA complexes in the unit cell are very similar, as they can be superimposed with a root-mean-square deviation of 0.36 Å (Figure 47). Because of that, tetrameric-RNA structure built from chains – A, B, C, and D were used for structure visualization and measurements. The methylated VMC10 RNA substrate is forming tetrameric-complex with the length and width of approximately 70 Å and 45 Å respectively. The core of the tetrameric RNA, sandwiched by the base and apical stems and is formed by the six RNA triplets – four C:G:C^+, two U:A:U triplets, and by two non-canonical dG7:m6A8 base pairs, in which the sugar edge of dG7 contacts the major groove of the A8 (Figure 48).

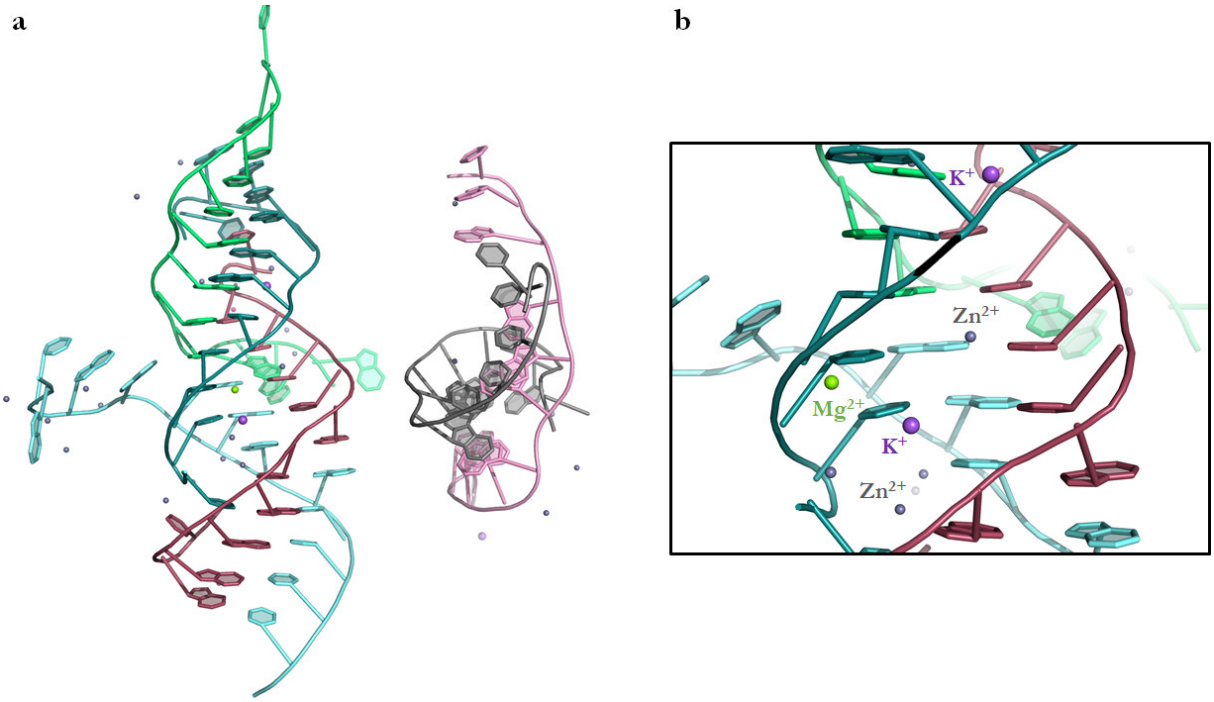


Figure 45. Overall crystal structure of the methylated R633 VMC10 RNA substrate.

(a) Overall crystal structure of the R633 VMC10 RNA substrate and its (b) close up view. Individual chains on the methylated VMC10 RNA substrate are colored separately. Chains A, B, C, and D are colored in red, cyan, teal, and green, respectively. Chains E and F are colored in grey and pink, respectively. Grey spheres indicate zinc ions. Purple and green spheres represent sodium and magnesium ions, respectively. For better visualization, water molecules were removed from the model.

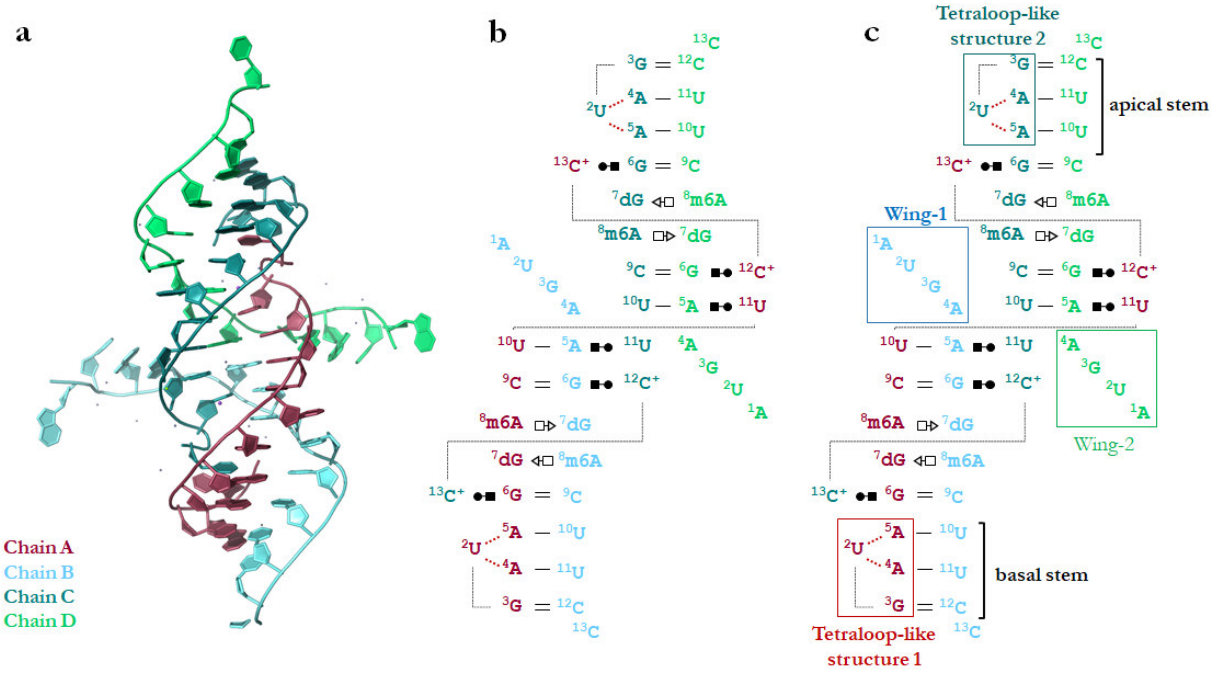


Figure 46. Crystal structure of the methylated VMC10 RNA substrate.

(a) Structure of the single copy of the tetrameric-RNA (Chains – A, B, C, D). (b) Organization of the nucleotides in the tetrameric RNA. The complex comprises two triple RNA helixes – formed between

chains A, B, C and A, C, D. Chains B, and D are not forming any hydrogen bonding interactions between each other. (c) Structural motifs observed in the tetrameric-RNA. The basal and apical stems contain three WC base pairs and the tetraloop-like structure. The core of the complex is formed by six RNA triplets (four C:G:C$^+$ and two U:A:U) and four deoxy-guanosine and N6-methyladenosine basepairs. Four unpaired nucleotides at the 5' site of the chains B and D are forming two wings. The single and double bars represents WC base pairs. The symbol between C$^+$ and G represents the cis Watson-Crick/Hoogsteen (cWH) base pair. The symbol between 7dG and 8A represents the trans Hoogsteen/Sugar edge base pair (tHS) base pair.

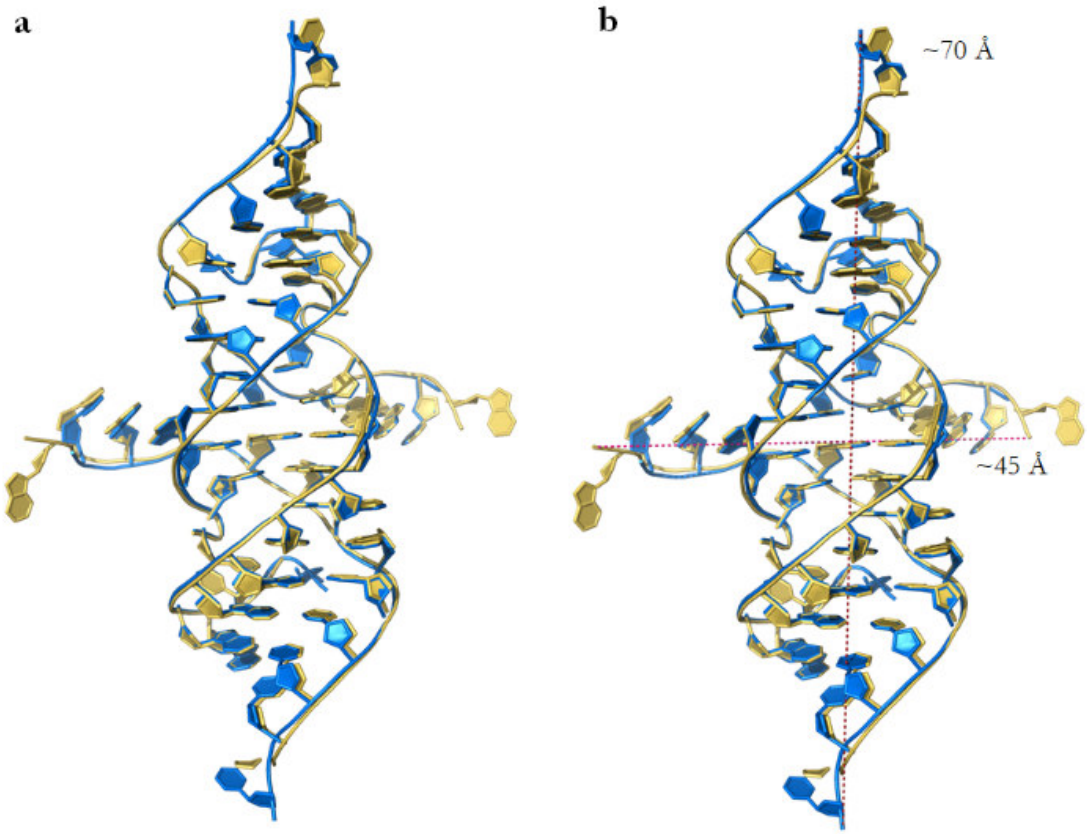


Figure 47. Superimposed models of the tetrameric-RNA complexes.

(a) Superimposed tetrameric-RNAs. Individual molecules of tetrameric-RNA are colored in blue and yellow. (b) Superimposed tetrameric-RNA complexes. Red and pink dashed lines indicate the approximate length and width of the RNA tetramer, respectively.

Table 5. Data collection and refinement statistics of the R616.

	Non-cleavable VMC10 m6A RNA substrate (R633)
Data collection	
Wavelength (Å)	1.28097
Resolution range	44.54 - 1.68 (1.71 - 1.68)*
Space group	P 3_2 2 1
Unit Cell	
a, b, c (Å)	76.50 76.50 60.16
α, β, γ (°)	90 90 120
R-merge	0.093 (>1.0)
No. reflections	385235 (3450)
No unique	22034 (719)
I/σI	30.04 (1.3)
Multiplicity	17.5 (4.8)
Completeness (%)	93.9 (61.0)
CC1/2	1.00 (0.669)
Refinement	
Resolution range	44.54 - 1.68 (1.71 - 1.68)
R-work/R-free (%)	18.69/20.44
Reflections used in the refinement	22011 (1527)
No of atoms	1851
RNA	1572
Ligands/Ions	29
Water	250
r.m.s deviations	
Bond lengths	0.016
Bond angles	1.74
Average B-factor	33.81
RNA	32.53
Ligands/Ions	53.41
Water	39.61

***Statistics for the highest-resolution shell are shown in parentheses.**

The C:G:C^+ triplet is formed at low pH and dissociate at high pH. Under such conditions, protonation of the N3 of cytosine base occurs and stabilizes RNA triple helix. The protonated cytidine is forming the RNA triplet with the canonical G:C base pair via cis Watson-Crick/Hoogsteen (cWH) hydrogen-bonding pattern, where the protonated N3 of cytidine in forming a hydrogen bond with the N7 of guanosine, with the distance of ~2.8 Å. An additional hydrogen bond, which helps to stabilize the triple RNA helix is formed between 2'-OH of the protonated cytidine and oxygen-phosphate (OP2) of the guanosine, with the distance of ~2.6 Å (Figure 48a,d).

In the U:A:U triplet, the triple RNA helix is stabilized by the residual uridine hydrogen bonding with the A:U base pair via cis Watson-Crick/Hoogsteen (cWH) bonding pattern (Figure 48e,f).

The first hydrogen bond is formed between N3 of uridine and N7 of adenosine, N6 of adenosine, and the second bond is formed between O4 of uridine and N6 of adenosine. The cis Watson-Cricks/Hoogsteen edge hydrogen bonds have a distance of ~2.9 Å.

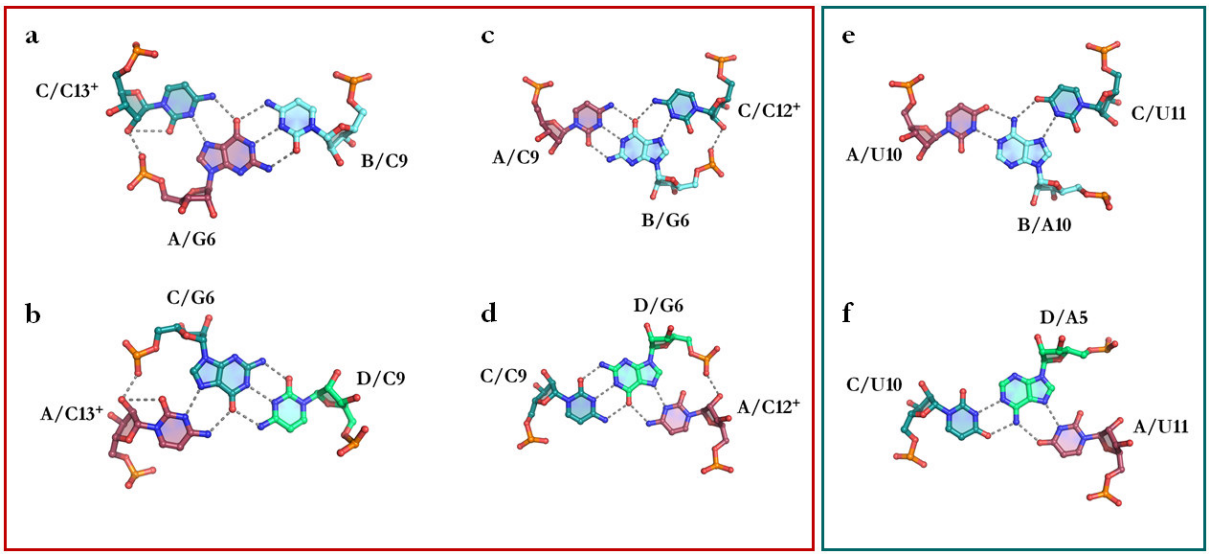


Figure 48. RNA triplets forming the tetrameric-RNA core.

Four C:G:C+ (a-d) and two U:A:U (e-f) RNA triplets are stabilizing the core structure of the tetrameric-RNA.

The complex comprises basal and apical stem containing three base pairs – G3:C12, A4:U11, and A5:U10. The four, unpaired nucleotides – A1, U2, G3, and A4 from chains B and D, are forming "wing" structures (Figure 49). The wings are stabilized by a few hydrogen bonding interactions. The first bond is formed between 2'-OH of G6 and 2'-OH of U11 with a distance of 2.6 Å. The second bond is formed between the N1 of the A4 and 2'-OH of U10 with a distance of 2.6 – 2.8 Å. An additional hydrogen bonding interaction stabilizing a triple RNA helix is formed between 2'OH of C12 and oxygen-phosphate (OP2) of G6, with a length of 2.6 – 2.7 Å. The nucleotides U2 and G3 are additionally stabilized by the two zinc atoms (Figure 49).

Additionally, nucleotides 2U, 3G, 4A, and 5A from chains A and C, are forming tetraloop-like structure, besides creating the Watson-Crick base pairs with U10, U11, and C12 from chains B and D (Figure 50).

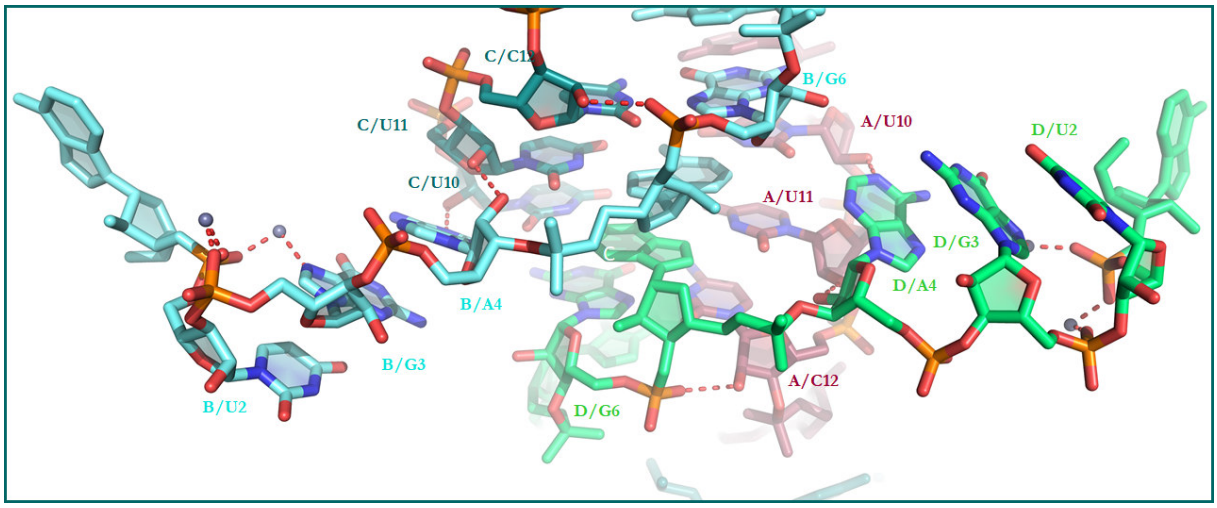


Figure 49. A close-up view of the wings.

Close up view of the wings and their hydrogen-bonding interactions. Two RNA triple-helixes are formed within tetrameric-RNA.

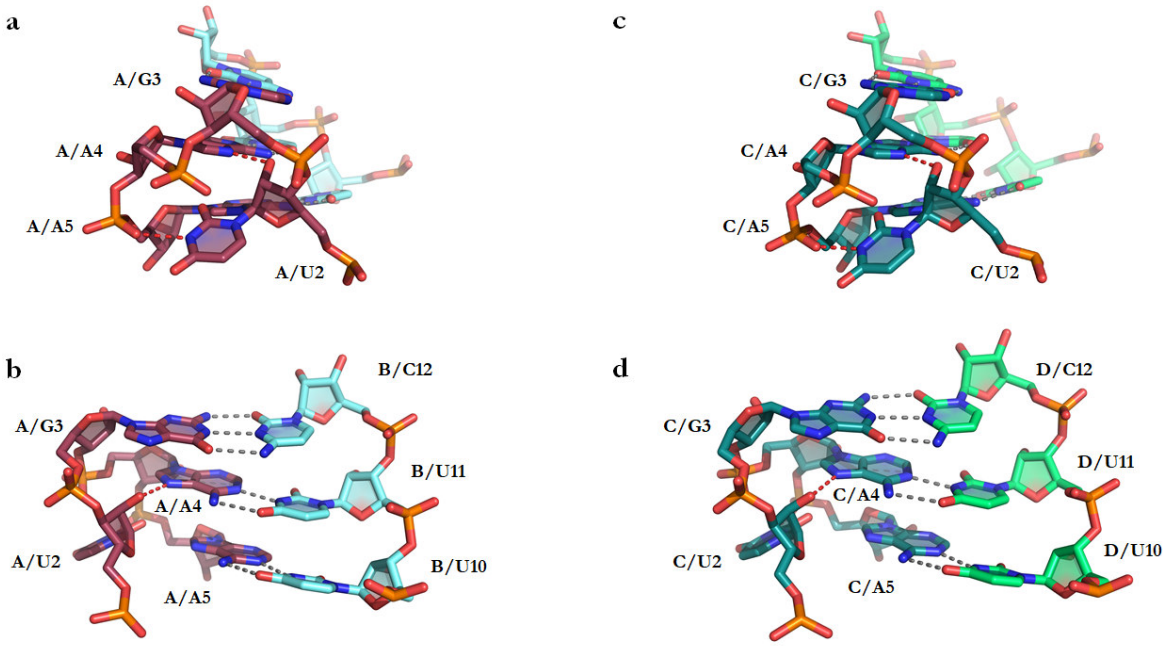


Figure 50. Tetraloop-like/Watson-Crick base pair interactions.

Tetraloop-like/WC base pair interaction in the basal (a,b) and apical (c,d) stems from the tetrameric-RNA. Dashed red lines indicate hydrogen-bonding interaction within the tetraloop-like structure. Dashed grey lines represent canonical Watson-Crick base-pairing.

Characterization of the sheared dG:m6A base pair

The two, consecutive dG7:A8 base pairs are formed by trans Hoogsteen/Sugar edge hydrogen bonding interaction (Figure 51). The hydrogen bond formed between N2 of dG7 and N7 of A8 has an approximate length of 2.9-3.0 Å. Additionally, a second hydrogen bond can be formed

between N2 of dG7 and oxygen-phosphate (OP1) of m6A8 with an average length of 3.2 – 3.5 Å, except the B/dG7:A/m6A8 base pair (Figure 51b), where the distance between the N2 and OP1 has a length of 3.8 Å and because of that, hydrogen bond cannot be formed. All nucleotides forming tHS dG:m6A base pairs possess C3'-*endo* sugar puckers conformation.

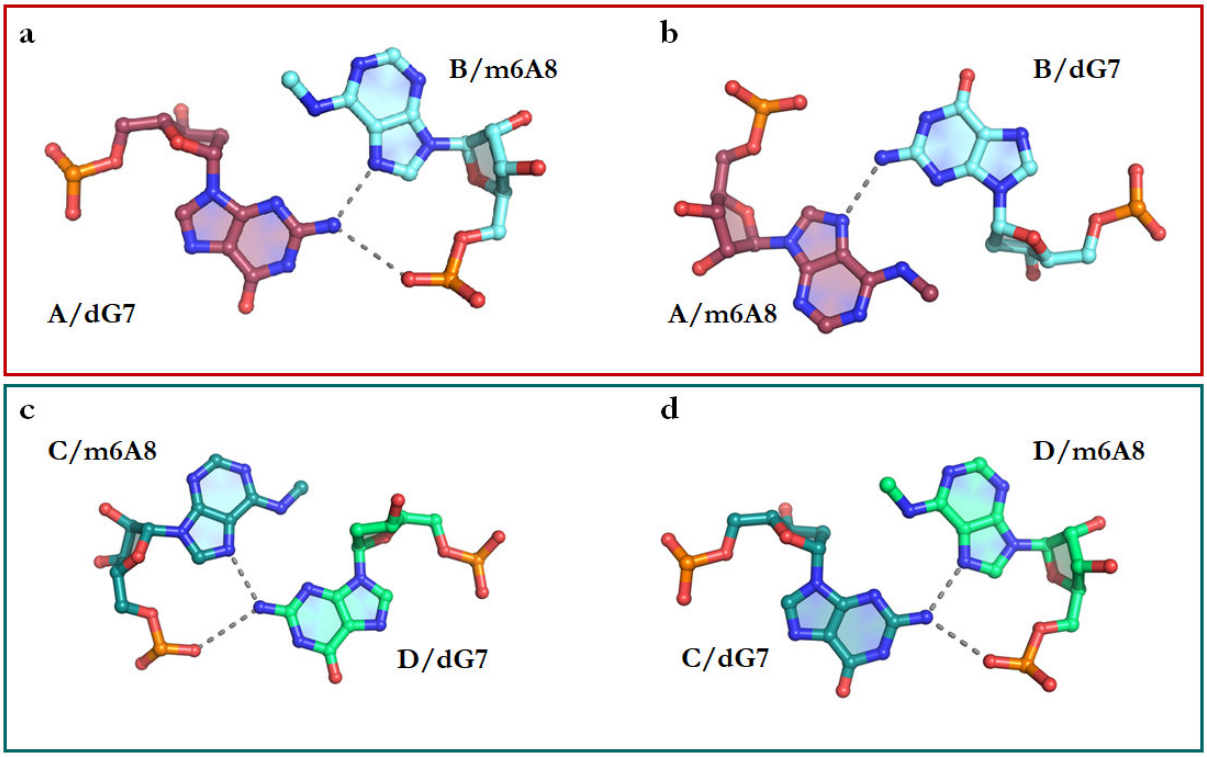


Figure 51. The trans Hoogsteen/Sugar base pairing of the dG7:m6A8.

The tHS base pairs of dG7:m6A8 formed next to the basal (a,b) and apical (c,d) stems from the tetrameric-RNA.

Crystal packing of the R633 methylated VMC10 RNA substrate

The R633, methylated VMC10 RNA substrate is forming a double, semi-continuous helix in the crystal (Figure 52). The individual molecules of the tetrameric-RNA are forming semi-continuous via interactions between basal and apical stems. The wings are forming intermolecular contacts between two separate semi-continuous helixes in the crystal (Figure 53).

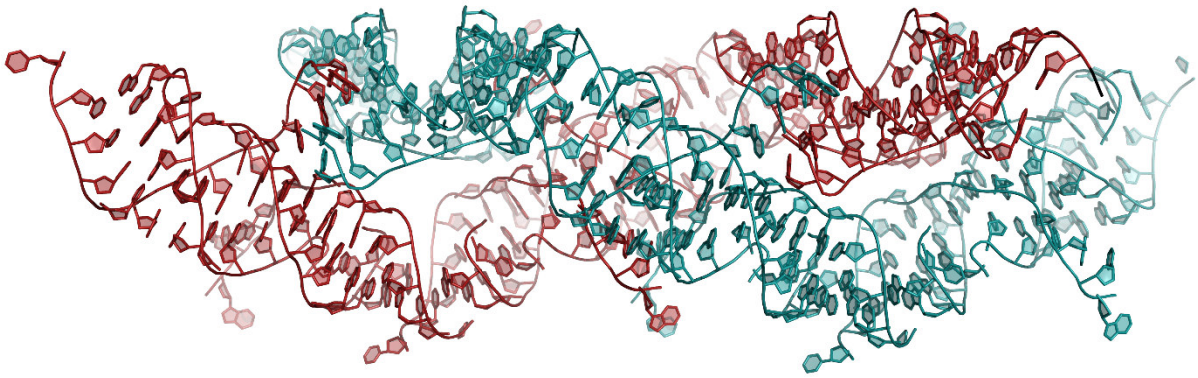


Figure 52. Crystal packing of the R633 methylated VMC10 RNA substrate.

The VMC10-RNA substrate is forming a double, semi-continuous helix in the crystal (red and teal). The individual molecules of the tetrameric-RNA are connected by the interactions between basal and apical stems.

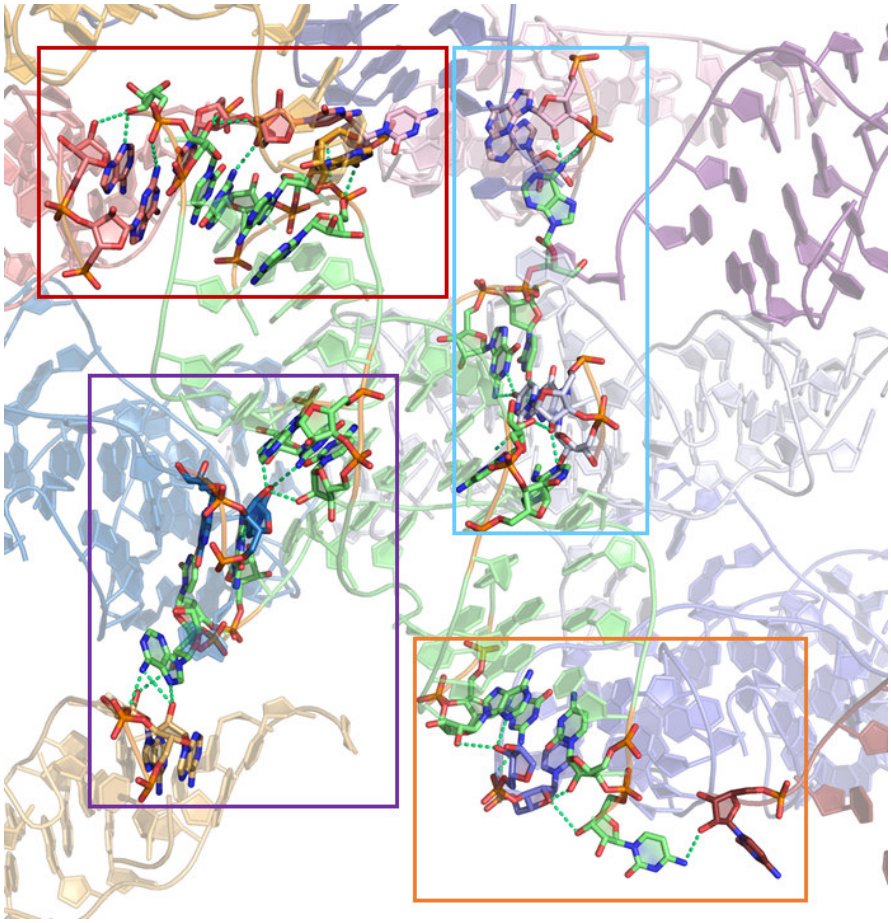


Figure 53. Intermolecular contacts of the R633 methylated VMC10 RNA substrate.

The green molecule represents RNA-tetramer forming the asymmetric unit. Red and the orange boxes represent intermolecular interactions formed by the basal and apical stems of the RNA-tetramer. Blue and purple boxes show intermolecular interactions formed by the wings.

Intermolecular contacts formed by the apical and basal stems

In the basal and apical stem, we can observe many different intermolecular contacts involving both nucleotide bases and 2'-hydroxy groups. In the apical and basal stem, we can observe intermolecular hydrogen bonds between 2'OH groups of cytidine 12 and 13 from the chains B and D with the symmetry-related cytidine 12 from chains D' and B', with the O-O a distance of 2.4 Å and 3.4 Å respectively. N2 of G3 (chain A) is forming a hydrogen bond with the oxygen-phosphate 2 of the symmetry-related C13 (chain D') with the N-O distance of 2.9 Å, while the N2 of G3 (chain C) is forming a hydrogen bond with the oxygen-phosphate 2 of the symmetry-related C12 (chain B') with the N-O distance of 3.4 Å. The N3 and 2'OH of A4 from the apical stem (chain C) is hydrogen bonding with symmetry-related C13 (chain B') with the N-O distance of 2.8 Å and O-O a distance of 3.4 Å, while the A4 from the basal stem (chain A) is making hydrogen bonding with N7 or N6 of symmetry-related A1 from the chain D' (N-O distance 2.9 Å and 3.4 Å respectively). An extra, intermolecular hydrogen bond in the basal stem is formed between 2'OH of symmetry-related C13 (chain F') and O5' of A5 (chain A, O-O distance 3.4 Å). C13 In the apical stem, we can observe an extra hydrogen bond formed between N4 of C13 (chain D) with the symmetry-related 2'OH group of C13 (chain F', N-O distance 3.3 Å) (Figure 54).

Intermolecular contacts formed by the wings

Several intermolecular hydrogen bonds involving nucleotides forming the core of the tetrameric-RNA helps to stabilize wings. The N1 of the m6A8 and 2'OH group of the C9 (chains – A and C) are hydrogen bonding with the 2'OH of the symmetry-related U2 (chains – B' and D') with the N-O and O-O a distance of around 2.8 Å and 2.7 Å, respectively. Additionally, the N2 of G6 (chains – B and D) is forming an intermolecular hydrogen bond with symmetry-related O2 of U2 (chains – D' and B') with the N-O distance of around 2.6 Å. The wings are stabilized via intermolecular contacts between U2 and G3 (chains – B and D) with the same, symmetry-related U2 and G3 (chains B' and D'). Wing-1 (chain B) is in contact with symmetry-related Wing-2 (chain D') and analogously Wing-2 (chain D) is in contact with symmetry-related Wing-1(chain B'). The O4 of U2 is forming hydrogen bonds with N1 and N2 of symmetry-related G3 (N-O distances 2.7 Å and 3.0 Å respectively.) The same bonding pattern is observed for the G3 and symmetry-related U2 ((N-O distance 2.7 Å and 3.3 Å respectively). The 5'-terminal A1 in the Wings adopts two different conformations. The A1 from the Wing-2 is in *syn*-conformation (χ = 52.5º) while the A1 from the Wing-1 is in adopts *anti*-conformation (χ = -141.3º). In the Wing-1, the N6 of A1 forms a hydrogen bond with 2'OH of the symmetry-related A4 (N-O distance 2.2 Å) and O5' of the symmetry-related A5 (N-O distance 3.1 Å) from the chain E'. In the Wing-2, The N7 of A1 is

hydrogen bonding with 2'OH of symmetry-related A4 (N-O distance 2.9 Å). The N6 of the same adenosine is forming intermolecular interactions with 2'OH of A4 and O5' of A5 (N-O distance 3.4 and 3.2 Å respectively). An additional hydrogen bond is formed between 2'OH groups of A1 and symmetry-related A6 (O-O distance 3.6 Å).

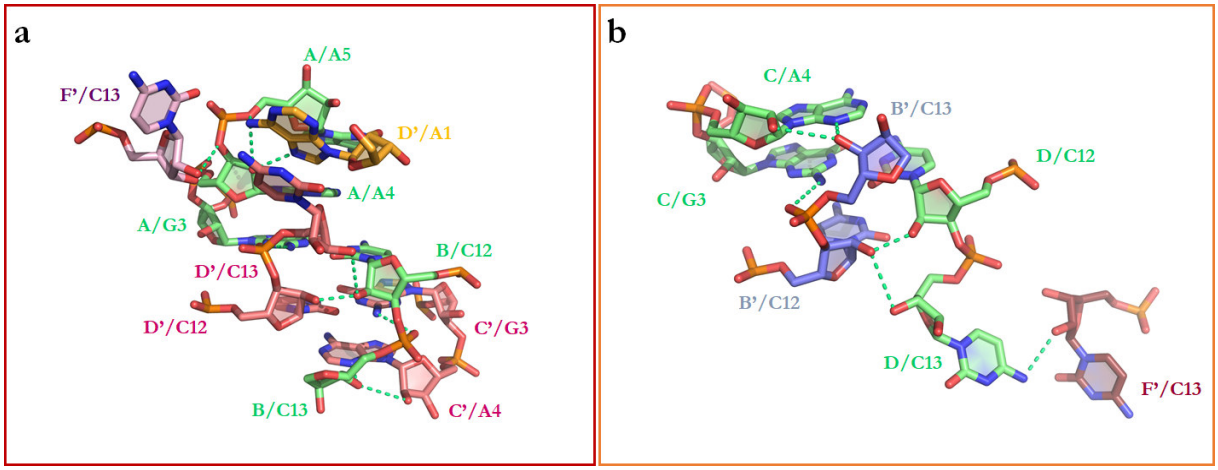


Figure 54. A close-up view for the intermolecular contacts between neighboring RNA-tetramers formed by the apical (a) and the basal (b) stems nucleotides.

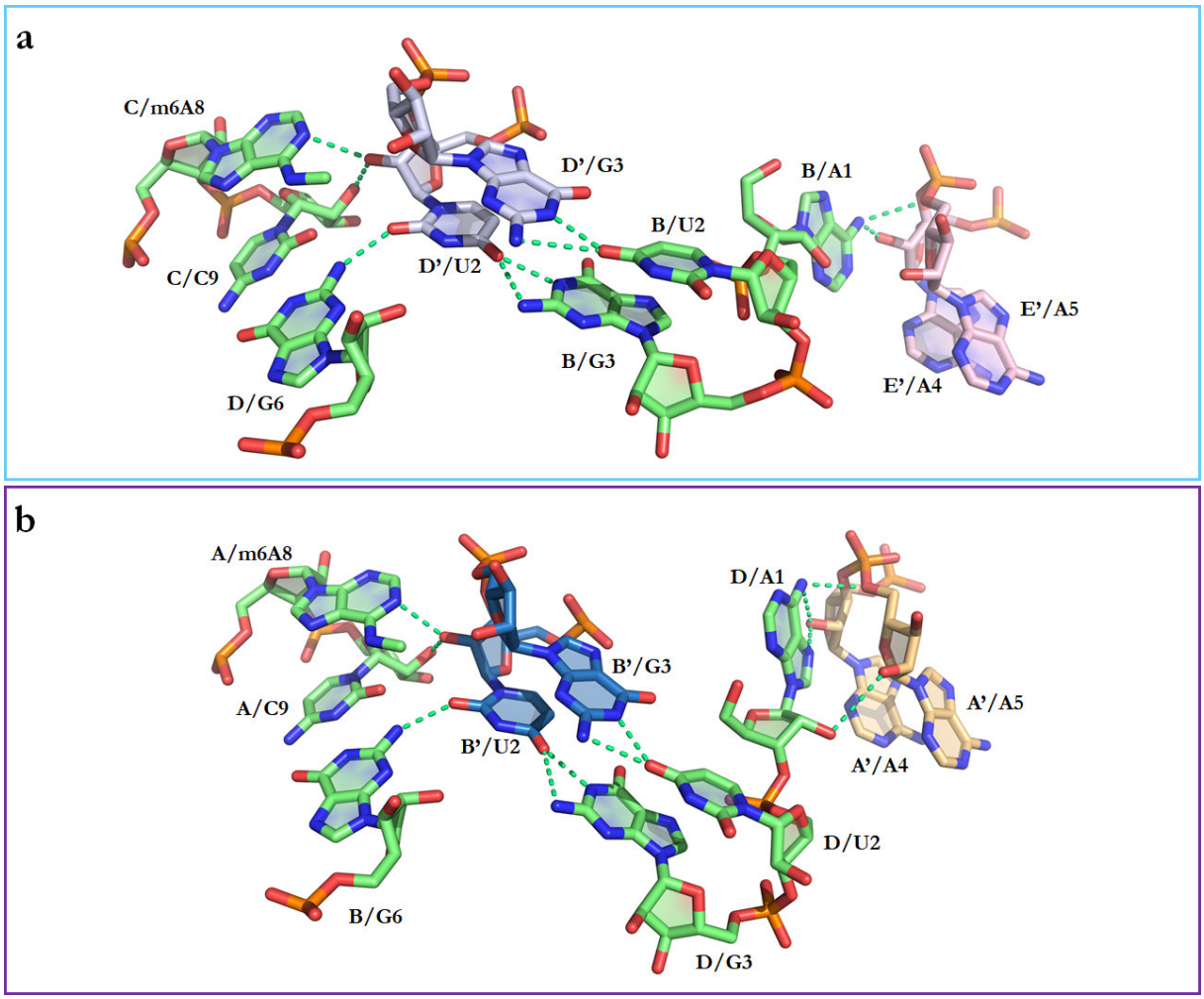


Figure 55. A close-up view for the intermolecular contacts between neighboring RNA-tetramers formed by the wing-1 (a) and the wing-2 (b).

3.1.4 Design and crystallization of new VMC10 variants

3.1.4.1 Fourth generation of the VMC10 DNAzyme constructs

Given that the RNA substrate dissociated from the DNAzyme, I re-designed the RNA substrate sequence to avoid dissociation from the DNAzyme during crystallization and prevent the formation of the RNA duplex and tetrameric RNA. The new generation of the VMC10 DNAzyme included different lengths of the RNA substrate (15-17 nucleotides) and complexes both with "sticky" and "blunt" ends (Figure 56).

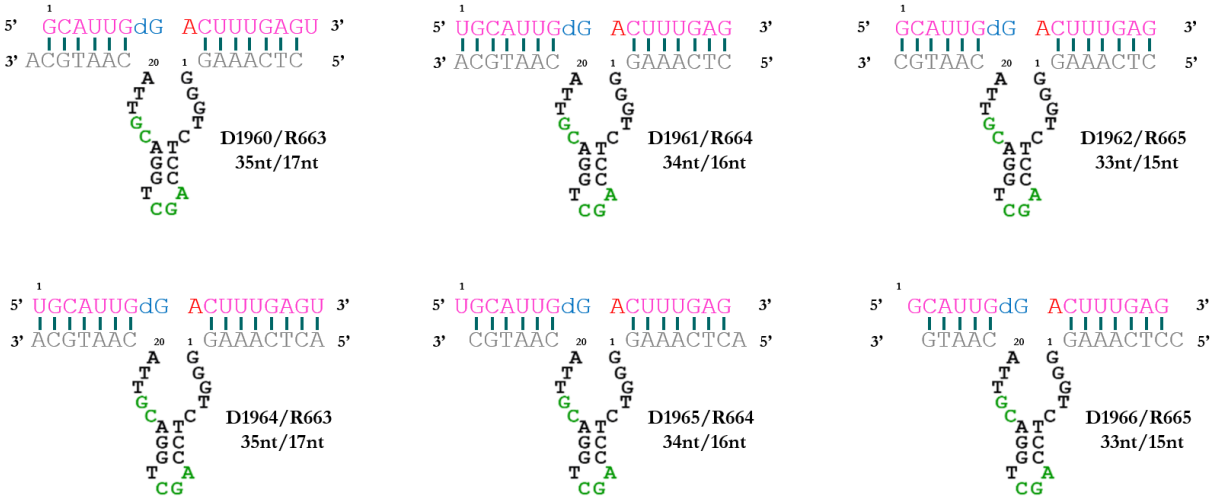


Figure 56. The fourth generation of the VMC10 DNAzyme constructs.

The re-designed sequences of the RNA substrate and corresponding substrate-binding arms are colored in pink and grey, respectively.

Crystallization screening

I performed the crystallization screening of the fourth generation of VMC10 deoxyribozyme complexes in the same way as for the previous generation of DNAzyme constructs. Two commercially available crystallization screens – Natrix 1+2 and Helix were used for initial screening. The crystallization drops consist of 120 nl of a 1 mM complex solution and 120 nl of crystallization buffer. Constructs D1965/R664 and D1966/R665 produced initial single crystals in the condition C5 of the Natrix 1+2 crystallization screen and D11 of the HELIX crystallization screen respectively. Crystals appeared after around one week.

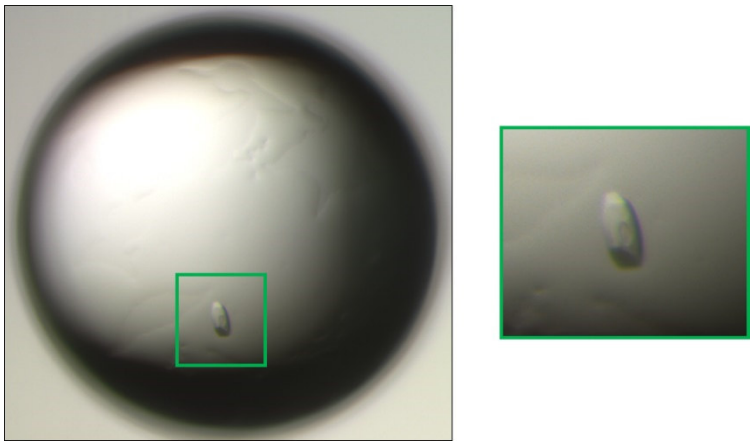


Figure 57. Representative crystals produced in the C4 Natrix 1+2 crystallization screen condition.

D1964/R664 complex, crystallization condition: 1.6 M lithium sulfate monohydrate, 0.01 M Magnesium chloride hexahydrate, 0.05 M HEPES pH 7.0.

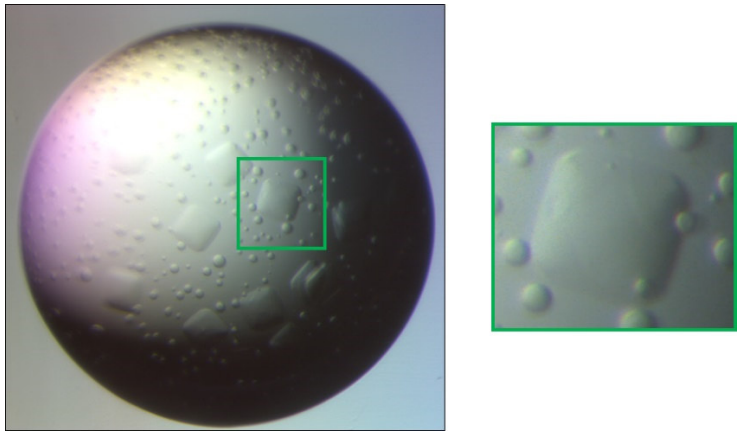


Figure 58. Representative crystals produced in the D11 crystallization screen condition.

D1965/R665 complex, crystallization condition: 0.2 M lithium chloride, 22% PEG 2000 MME, 0.05 M Bis-Tris pH 7.0.

Optimization of crystallization condition of the fourth-generation VMC10 constructs

D1965/R664 complex at 0.5 – 1 mM concentration was mixed with solutions containing a gradient of lithium chloride (3-5 M), 0.01 M magnesium chloride, and 0.05 M HEPES pH 7.0. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 1 µl of DNAzyme complex and 1 µl of crystallization buffer. Crystallization was performed using the hanging drop vapor diffusion method at 20ºC.

D1966/R665 complex at 0.5 – 1 mM concentration was mixed with solutions containing gradient of PEG 2000 MME (12-17 %), 0.2 M lithium chloride and 0.05 M Bis-Tris pH 7.0. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 120

nl of DNAzyme complex and 120 nl of crystallization buffer. Crystallization was performed using the sitting drop vapor diffusion method at 20ºC.

3.1.4.2 Fifth-generation of the VMC10 DNAzyme constructs

The VMC10-substrate complex - D1966/R665 from the 4th generation produced the biggest and well-formed, single three-dimensional crystals. Because of that, I generated additional constructs, very similar to the D1966/R665 complex, to increase the chance of getting high-quality crystals (Figure 59).

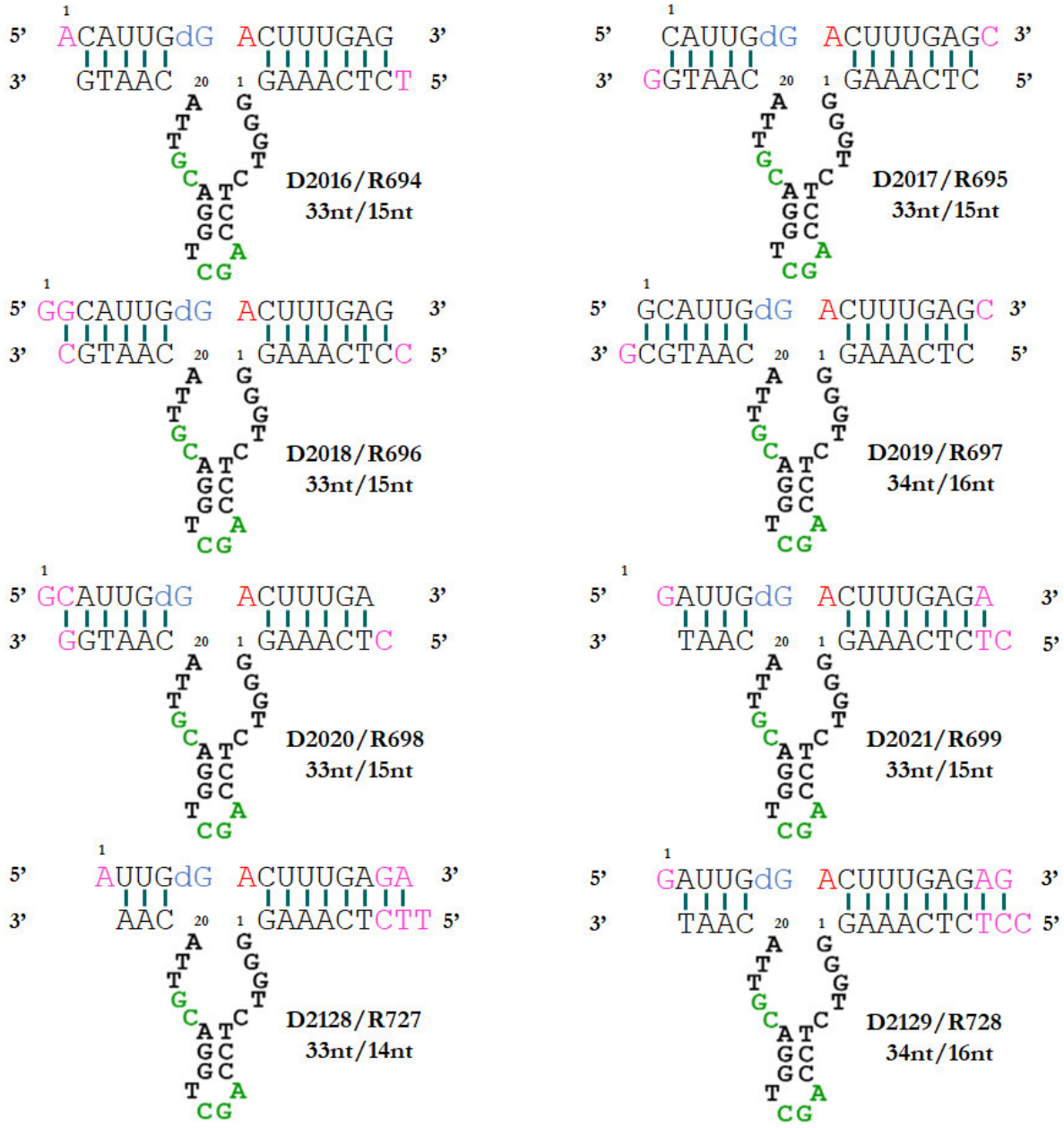


Figure 59. The fifth generation of the VMC10 crystallization constructs.

Crystallization screening

I performed the crystallization screening of the fourth-generation of VMC10 deoxyribozyme complexes in the same way as for the previous generation of DNAzyme constructs. Two commercially available crystallization screens – Natrix 1+2 and Helix were used for initial screening. The crystallization drops consist of 120 nl of a 1 mM complex solution and 120 nl of crystallization buffer.

Complex D2017/R695 produced initial single crystals in the condition H10 of the Helix. crystallization screen (0.05 M lithium sulfate, 0.05 M Bis-Tris pH 8.0, 0.03 M magnesium sulfate and 15% PEG 3350). Complex D2018/R696 produced initial single crystals in the condition A4 of the Helix crystallization screen (0.2 M potassium chloride, 0.005 M hexamine cobalt (III) chloride, 0.05 M MES pH 6.5, and 25% PEG 400). Complex D2019/R697 produced initial single crystals in the condition E10 of the Helix crystallization screen (0.05 M Bis-Tris pH 7.0 and 1.5 M lithium sulfate). Complex D2020/R698 produced initial single crystals in the condition F6 of the Helix crystallization screen (0.1 M Sodium chloride, 0.01 M lithium chloride 0.05 M HEPES pH 6.5, and 1.8 M ammonium sulfate). Complexes D2021/R699 and D2129/R728 produced initial single crystals in the condition C5 of the Natrix 1+2 crystallization screen (0.05 M magnesium sulfate, 50 mM HEPES pH 7.0, and 1.6 M lithium sulfate). All of the constructs (except D2020/R698) produced crystals after 1-2 weeks. D2020/R698 produced crystals after around one month. Complex D2016/R694 wasn't subjected to the crystallization screening.

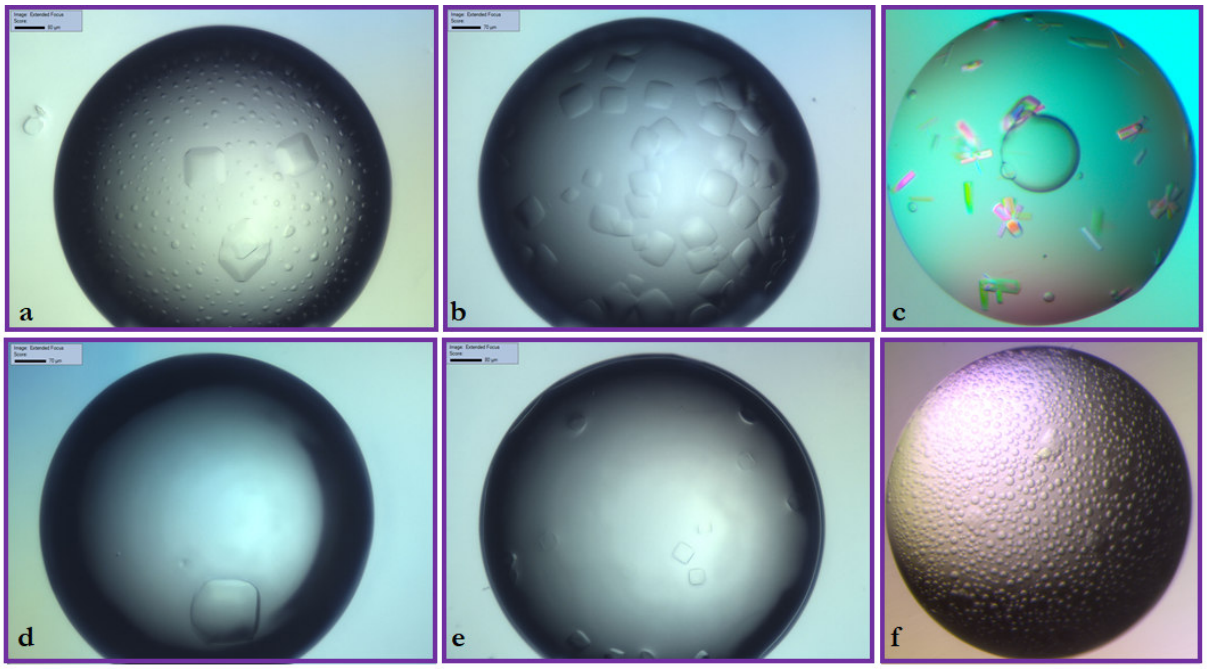


Figure 60. Representative crystals of the 5th generation of the VMC10 DNAzyme constructs.

Crystals of the (a) parent 4th generation D1966/R665 (b) D2017/R695 (c) D2020/R698 (d) D2019/R697 (e) D2020/R699 (f) D2018/R696 DNAzyme complexes.

Optimization of crystallization condition of the fifth-generation VMC10 constructs

The DNAzyme constructs from the 5th generation were subjected to the optimization procedure to produce bigger crystals for the diffraction analysis. The complexes at 0.5 – 1 mM concentration were mixed with solutions containing a gradient of the primary precipitant and pH. VMC10 complex and crystallization solutions were mixed in a 1:1 ratio. Crystallization drops consist of 120 nl of DNAzyme complex and 120 nl of crystallization buffer. Crystallization was performed using the sitting drop vapor diffusion method at 20ºC.

3.1.4.3 Diffraction experiments of the fourth and fifth generation of the VMC10 DNAzyme constructs

Diffraction data of the fourth and fifth generation of VMC10 DNAzymes constructs were collected at 100 K on PILATUS 6M or EIGER 16M detectors at the X10SA (SLS). Data were processed as described in section 2.2.3.2. The complexes from the 4th and 5th generation of VMC10 were diffracting with different power (Table 6).

Table 6. Summary of the diffraction properties of the best crystal forms.

DNAzyme complex	Resolution	Space group
D1965/R664	4.9-5 Å	$P4_1\ 2_1\ 2/P4_3\ 2_1\ 2$
D1966/R665	5.5-6 Å	$P4_1\ 2_1\ 2/P4_3\ 2_1\ 2$
D2017/R695	10-8 Å	N.D.
D2018/R696	10-8 Å	N.D.
D2019/R697	10-8 Å	N.D.
D2020/R698	2.7 Å	$P6_1\ 2\ 2/P6_5\ 2\ 2$
D2021/R699	4.76 Å	$P4_1\ 2_1\ 2/P4_3\ 2_1\ 2$
D2129/R728	8.5-9 Å	N.D

All of the complexes, except the D2129/R728 complex, possess similar morphology and crystallized in the primitive tetragonal space group. The complex D2020/R698 formed different crystals than the rest of the complexes and crystallized in a primitive hexagonal space group. The R616 VMC10 RNA substrate from the previous generation also formed primitive hexagonal crystals with similar unit cell parameters (Table 6). To check the presence of both DNA and RNA in the crystal I analyzed the crystal used for the data collection on the HPLC. The chromatogram confirmed that only the RNA molecule is present in the crystal (Figure 62). Surprisingly, this crystal form was formed only in solutions containing ammonium sulfate salt in high concentrations.

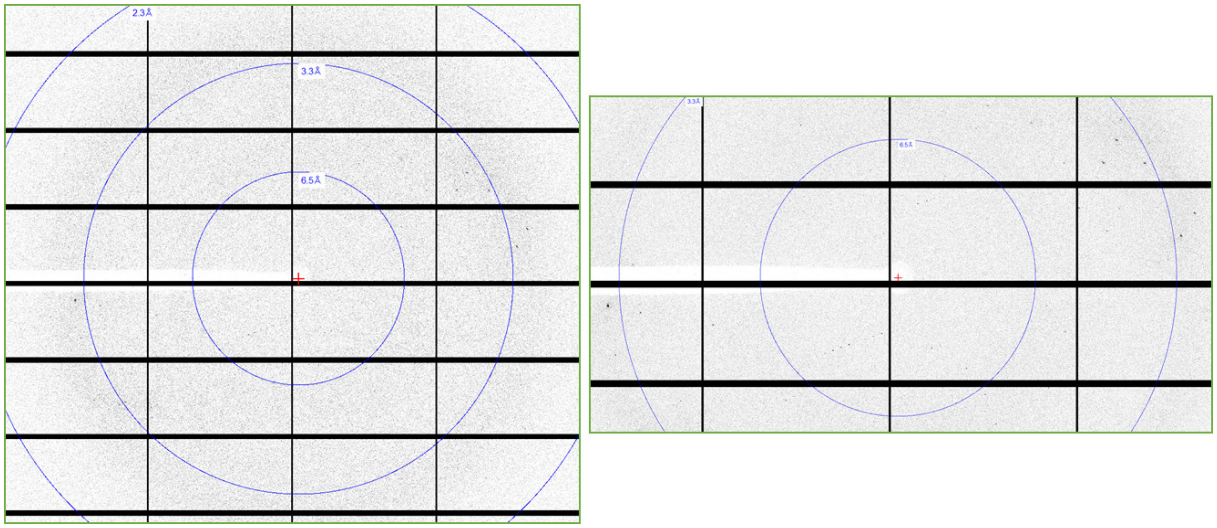


Figure 61. Diffraction image of the D2020/R698 crystal (left) and its close-up view (right).

Table 7. Comparison of the R616 and D2020/R698 crystal parameters.

	R616 structure	D2020/R698 complex
Space group	P 6_2 2 2	$P6_1\ 2\ 2/P6_5\ 2\ 2$
Unit cell dimensions		
a, b, c (Å)	53.06 53.06 100.92	58.39 58.39 104.60
α β γ (°)	90 90 120	90 90 120

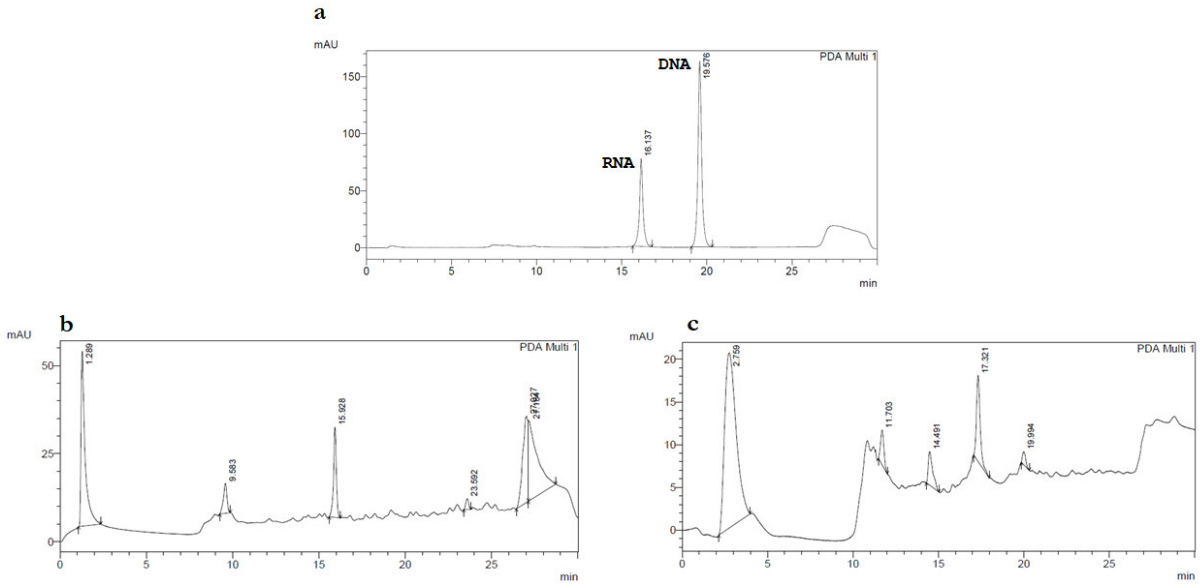


Figure 62. HPLC analysis of the D2020/R698 crystal.

(a) D2020/R698 sample used for crystallization. The RNA molecule is eluted around 16 minute and the DNA around 19 minute. (b) Hexagonal crystal formed at condition - Natrix 1+2 well F12 (0.02 M magnesium chloride, 0.05 M MOPS pH 7.0, 2.0 M ammonium sulfate, 0.0005 M Spermine) (c) Crystal used for the diffraction experiment from the condition Helix F6.

3.1.5 Post-crystallization improvement of the VMC10 crystals diffraction quality

To improve the diffraction properties of the VMC10 crystals, I have selected three of the most promising constructs (forming the biggest and reproducible crystals) for the post-crystallization treatment (Figure 63). The crystal dehydration combined with ion exchange was performed as described in section 2.2.2.5.

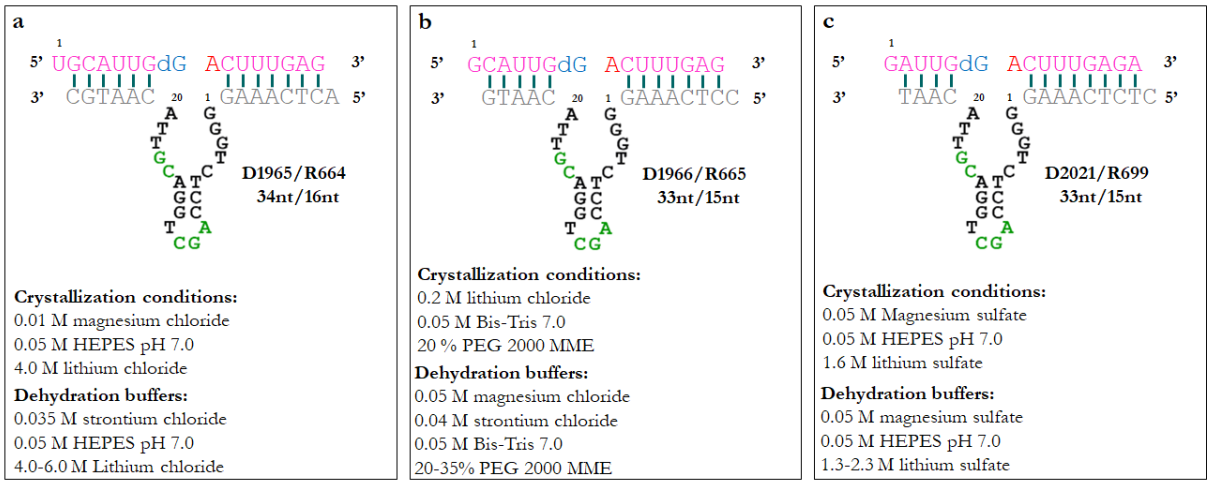


Figure 63. VMC10/RNA complexes used for the post-crystallization improvement of diffraction quality.

The crystal dehydration combined with the ion-exchange of the D1965/R664 and D2021/R699 complexes did not improve the diffraction properties of the complex crystals. The best effect was achieved with the D1966/R665 complex (Figure 64). The best diffraction properties were achieved by the crystal dehydration reaching the final PEG 2000 MME concertation of 35% and ion-exchange with 40 mM strontium chloride. Before the data collection, the D1966/R665 crystal was cryoprotected with the PFPE oil. Data collection statistics are presented in Table 8. To check the presence of both DNA and RNA in the crystal, D1966/R665 crystals formed in condition D11 were analyzed by the HPLC. The chromatogram confirms that both DNA and RNA are present in the crystal (Figure 65).

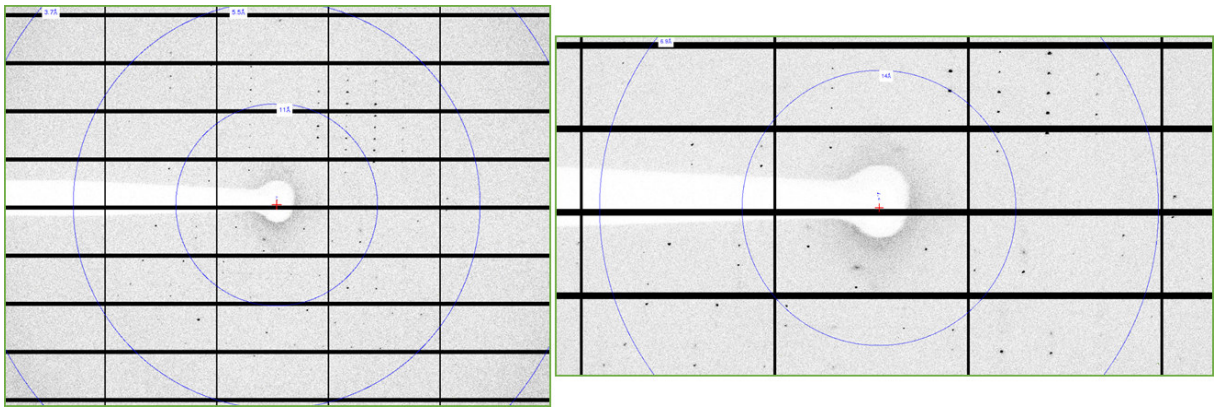


Figure 64. Diffraction image of the dehydrated D1966/R665 crystal (left) and its close-up view (right).

Table 8. Data collection statistics of the post-crystallization treated D1966/R665 crystal.

	D1966/R665 complex
Data collection	
Wavelength (Å)	1.000
Resolution range	48.40 - 3.83 (4.28-3.83)
Space group	P 4_1 2_1 2 / P 4_3 2_1 2
Unit Cell	
a, b, c (Å)	87.44 87.44 96.80
α, β, γ (°)	90.00 90.00 90.00
R-merge	0.05 (>1.0)
No. reflections	92868 (25129)
No. unique	3930 (1046)
I/σI	20.5 (1.3)
Multiplicity	23.6 (24.0)
Completeness (%)	99.4 (98.1)
CC1/2	1.00 (0.68)

***Statistics for the highest-resolution shell are shown in parentheses.**

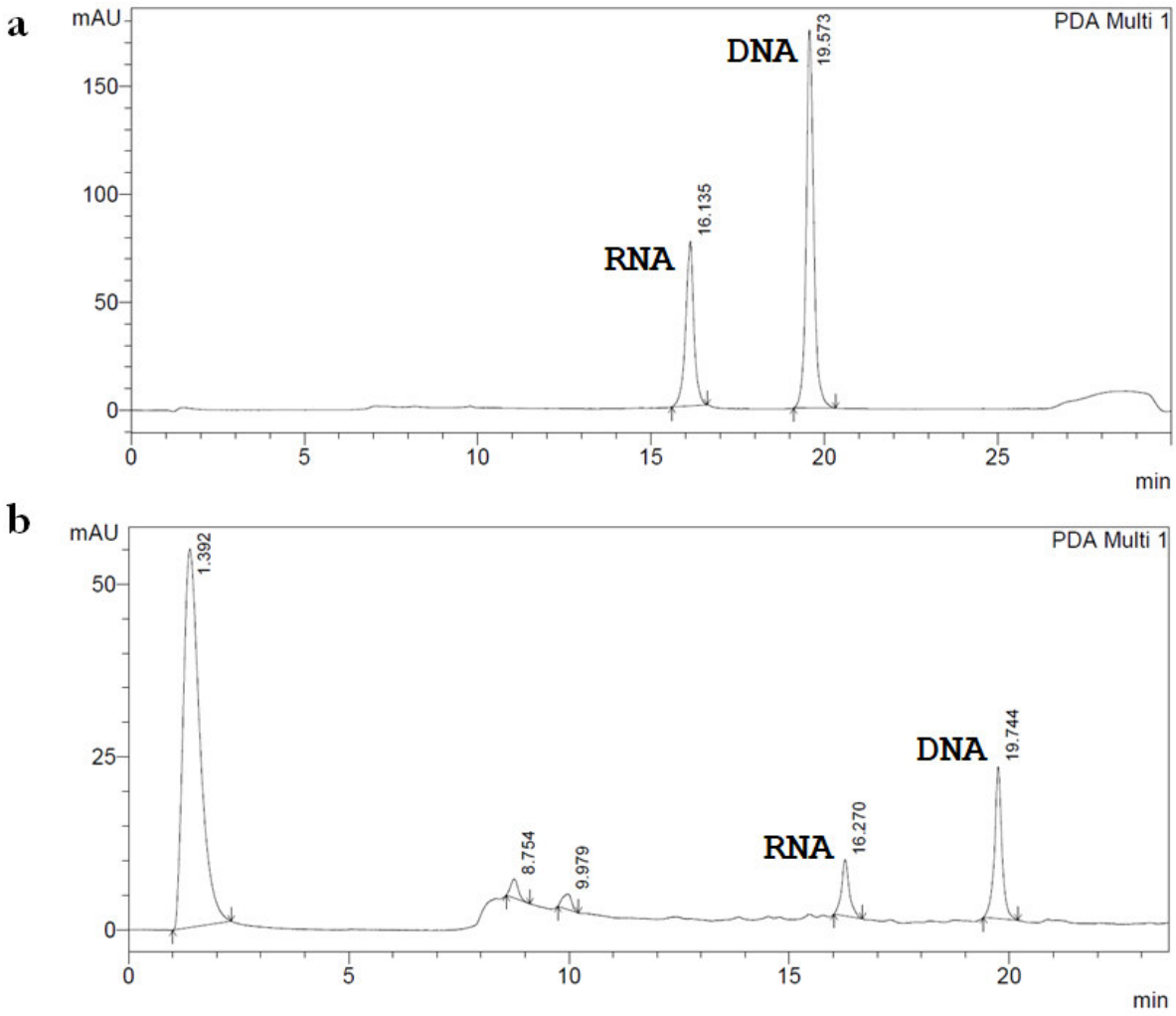


Figure 65. HPLC analysis of the D1966/R665 crystal.

(a) D1966/R665 sample used for crystallization. The RNA molecule is eluted around 16 minute and the DNA around 20 minute. (b) Crystals formed in the Helix D11 condition.

3.1.6 Structure analysis of the VMC10 RNA substrate (R698)

Overall crystal structure of the R698 VMC10 RNA substrate - new generation

The crystal structure of the R698 VMC10 RNA substrate was determined using the procedure presented in section 2.2.3.2.2. The crystallographic asymmetric unit (ASU) contains 1 copy of 15-nucleotide long non-cleavable VMC10 RNA substrate. In the unit cell, The RNA is forming a single A-form double helix flanked with two hairpin-like structures with a length of approximately 50 Å (Figure 66).

The RNA duplex contains 10 canonical Watson-Crick base pairs (two of each pair – G1:C10, C2: dG8, C3:G7, A4:U6, and U11:A15), two bulges (A9), U5-U5 mismatch, and two terminal hairpins.

Two terminal base pairs (A15:U11) are in the trans base pair configuration, while the rest base pairs are in the cis configuration. Additionally, A15 is in the *syn*-configuration with the $\chi = 70.6^{o}$.

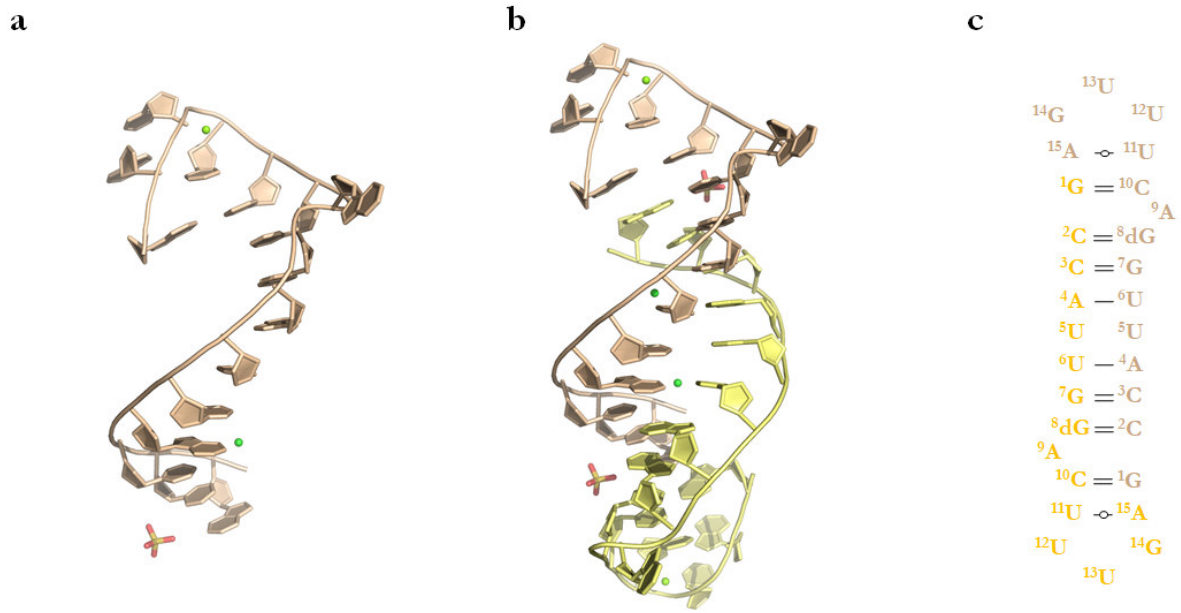


Figure 66. Overall crystal structure of the non-cleavable VMC10 RNA substrate.

(a) A single 15-nucleotide long RNA molecule forming an asymmetric unit (b) The double-stranded RNA A-form helix present in the unit cell. (beige – ASU molecule, yellow – symmetry-related molecule) (c) Organization of the nucleotides in the double helix. The single and double bars represents WC base pairs. The symbol between U11 and A15 represents the trans WC base pair.

Table 9. Data collection statistics of the R698 crystal.

	Non-cleavable VMC10 RNA substrate (R698)
Data collection	
Wavelength (Å)	1.000
Resolution range	45.53 - 2.70 (2.83 - 2.70)*
Space group	P 6_5 2 2
Unit Cell	
a, b, c (Å)	58.39 58.39 104.60
α, β, γ (°)	90 90 120
R-merge	0.072 (>1.00)
No. reflections	58605 (7700)
No. unique	3254 (403)
I/σI	17.2 (0.5)
Multiplicity	18.0 (19.1)
Completeness (%)	99.6 (99.2)
CC1/2	1.00 (0.66)
Refinement	
Resolution range	45.528 - 2.90 (3.00 - 2.90)

R-work/R-free (%)	24.24/25.97
Reflections used in the refinement	2498 (237)
No of atoms	321
RNA	314
Ligands/Ions	7
r.m.s deviations	
Bond lengths	0.002
Bond angles	0.54
Average B-factor	142.22
RNA	141.13
Ligands/Ions	191.18
Water	142.22

*** Statistics for the highest-resolution shell are shown in parentheses**

Organization of the nucleotides forming hairpin-like structure

The last five nucleotides at 3' (U11, U12, U13, G14, and A15) formed a hairpin-like motif with a three nucleotides loop. In a hairpin-like structure, the A15 at the 3' end is trans base-paired with U11 and two additional hydrogen bonds are formed. The first bond is formed between N3 of U12 and the oxygen-phosphate (OP2) of A15, and the second one between 2'OH of U12 and N7 of G14. The magnesium ion is stabilizing the RNA backbone. The A9 is flipped-out from the helix and it is forming a single nucleotide bulge (Figure 67).

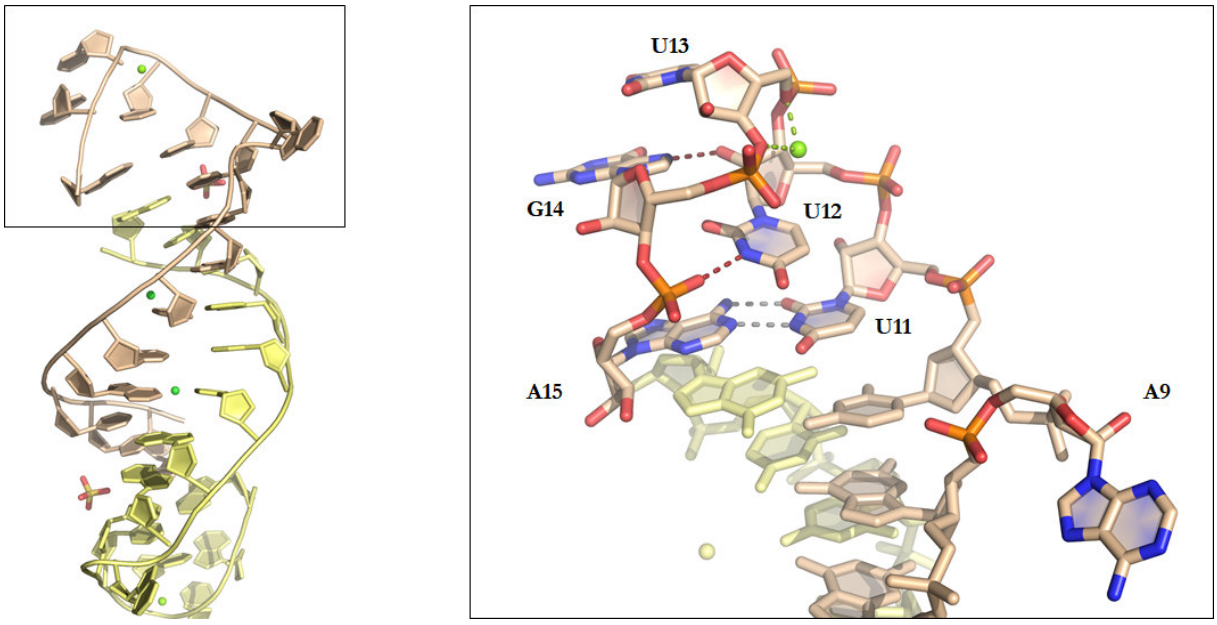


Figure 67. Hairpin-like motif observed in the R698 structure.

Five nucleotides from the 3' site are forming a hairpin-like motif. U11 base pairs with terminal A15. Two additional hydrogen bonds between U12 and G14 and oxygen-phosphate of A15, together with the magnesium ion (green sphere), are stabilizing the hairpin-like structure.

Crystal packing of the R698 VMC10 RNA substrate of the new generation

The new generation, non-cleavable VMC10 RNA substrate is forming a semi-continuous helix in the crystal (Figure 68). Formation of the semi-continuous helix is possible due to the kissing stem-loop interactions (Nowakowski and Tinoco 1997) between hairpins formed at the ends of the double helixes (Figures 69, 70). The bulged A9 forms an intermolecular contact between two separate semi-continuous helixes in the crystal (Figures 69,70).

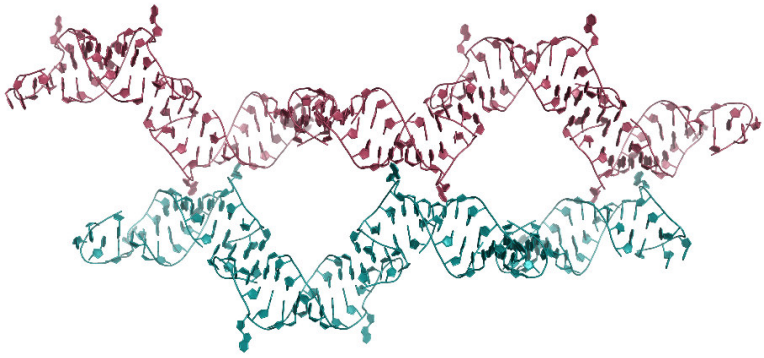


Figure 68. New-generation VMC10 RNA substrate forming semi-continuous helix in the crystal.

The formation of a semi-continuous helix in the crystal is possible due to the intermolecular interactions between hairpins from the two different ds-RNA molecules and they form a kissing stem-loop. In the new-generation non-cleavable VMC10 RNA substrate, the kissing loop is formed by two U:A:U triplets and two G-U pairs. The U13 residues form the U:A:U triplet by pairing with the Hoogsteen edge of A15, base-paired with U11, forming cis Watson-Crick/Hoogsteen (cWH) interaction pattern. The N3 and O4 of U13 make hydrogen bonds with the N7 and N6 of A15 (N-N distance 2.7 Å and 3.0 Å respectively). The G-U pairs are formed between two U12 and two G14 residues, where the O2 of U12 can form a hydrogen bond with N1 or N2 of G14 (both O-N distances 2.9 Å).

The bulged A9 is responsible for the contacts between two separate semi-continuous helixes. The N1 and N7 of the adenosine are forming hydrogen bonds with the 2'-OH groups with U6 and G7 of a neighboring molecule. The distance between N1 of A9 and 2'O of U6 is 2.8 Å and between N7 of A9 and 2'O of G7 is 3.2 Å. Additionally, the 2'OH group may form hydrogen bonds with N3 and a 2'-OH group of neighboring A4 (O-N distance 2.5 Å; O-O distance 2.8 Å).

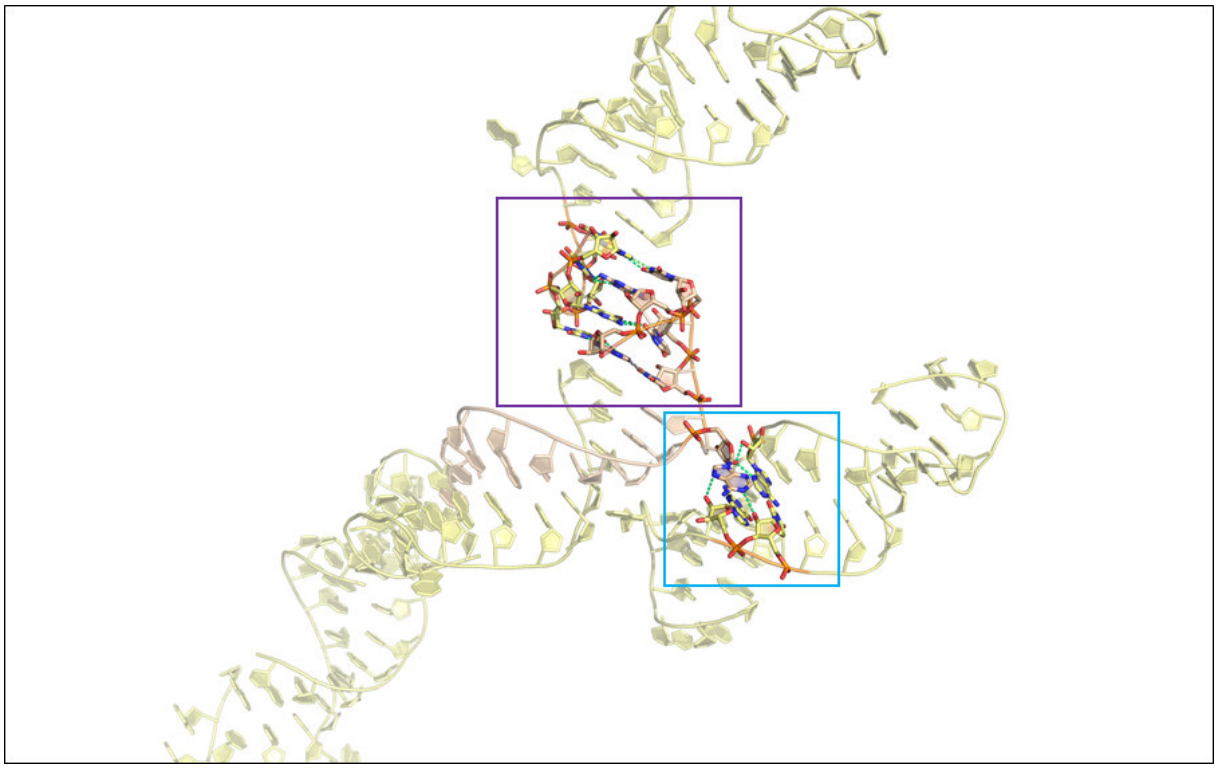


Figure 69. Intermolecular contacts of the new-generation VMC10 RNA substrate.

The kissing stem-loop intermolecular interactions formed at the ends of ds-RNA (purple box) allow for the formation of a semi-continuous helix in the crystal. The bulged A9 is responsible for the intermolecular contact between two separate semi-continuous helixes.

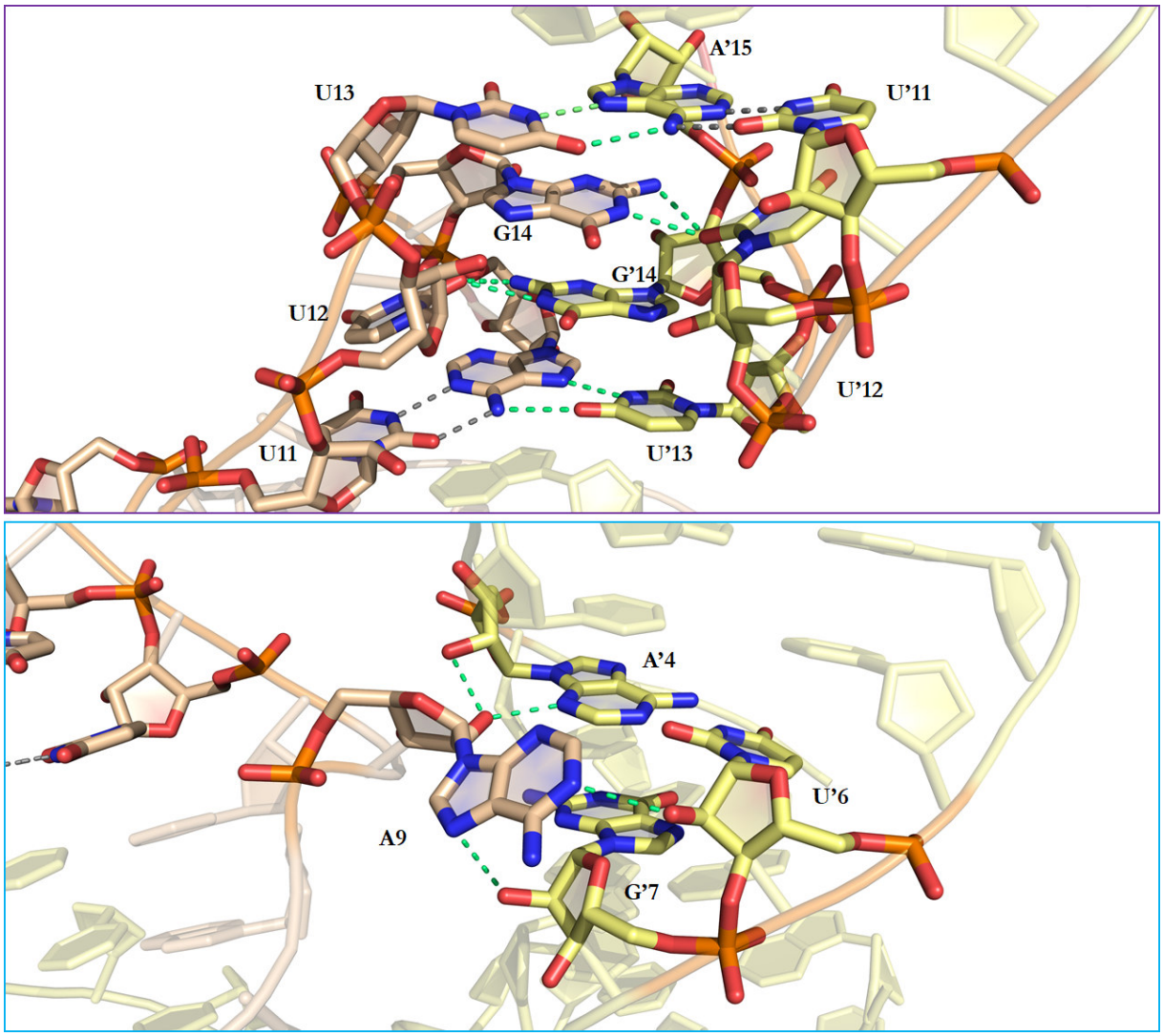


Figure 70. A close-up view for the intermolecular contacts of the new-generation VMC10 RNA substrate.

The intermolecular contacts (dashed green lines) of the kissing stem-loop region is shown in the purple box. The contacts formed by bulged A9 is shown in the blue box.

3.2 STRUCTURAL INVESTIGATION OF THE PRE-CATALYTIC STATE OF LIGATION OF 9DB1

To investigate the crystal structure of the 9DB1 in the pre-catalytic state of reaction, a mutant version of the parent deoxyribozyme was used – T29C, T7C together with two RNA substrates (Figure 71).

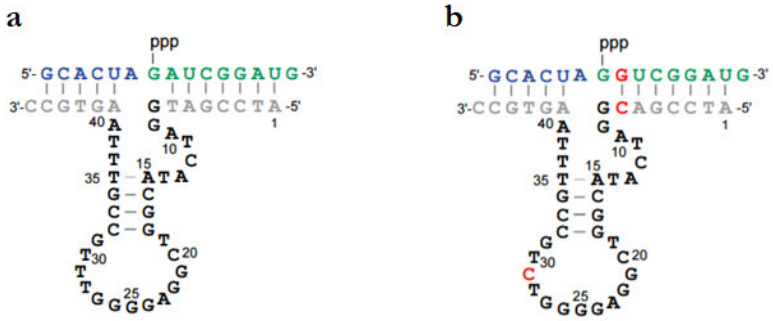


Figure 71. The RNA-ligating 9DB1 deoxyribozyme.

(a) Parent 9DB1 and (b) T29C/T7C mutant deoxyribozymes with its RNA substrates.

3.2.1 Crystallization screening

9DB1 deoxyribozyme in complex with two RNA substrate, prepared as described in section 2.2.2.1, was subjected to crystallization screening at 20 °C. Four commercially available crystallization screens – Helix and Natrix 1+2, Wizard Classic 1+2, and Wizard Classic 3+4 were used for initial screening. The crystallization drops consist of 120 nl of a 0.8 mM complex solution and 120 nl of crystallization buffer. Initial crystals appeared in condition H10 of the Wizard Classic 3+4 crystallization screen after one week.

3.2.2 Optimization of crystallization condition of the 9DB1 precatalytic-state of ligation complex

To obtain bigger crystals, optimization of crystallization conditions was performed using the hanging-drop vapor diffusion method at 20°C. 9DB1 with two RNA substrate at 1 mM concentration was mixed with solutions containing gradients of concentrations of 2-propanol (25-35%) and PEG3500 (25-35%) buffered with 0.1 M Tris pH 8.5. DNA/RNA complex and crystallization solutions were mixed in a 1:1 ratio. Drops with volumes of 2-3 μL produced crystals after around 7 days. Crystals were harvested from the drop, cryoprotected using PFPE oil, and frozen in liquid nitrogen.

3.2.3 Data collection and structure determination of the 9DB1 in the post-catalytic state

Diffraction data were collected at 100 K on the PILATUS 6M detector at the X10SA (SLS). Data were indexed, integrated, and scaled with XDS, and reduced with Pointless, Aimless, and Ctruncate within the CCP4 package. Initial phases were determined by molecular replacement using Phaser in Phenix, using the crystal structure of the 9DB1 deoxyribozyme (PDB code: 5CKK). Phaser found 1 MR solution with TFZ=16.9 and LLG=243.79. This initial model was modified by iterative rebuilding in COOT and refinement with Phenix.refine. Data collection and refinement statistics are summarized in Table 10.

Overall crystal structure of the 9DB1 T29C, T7C mutant

The structure of 9DB1 T29C, T7C deoxyribozyme revealed that the two RNA substrates were ligated during the crystallization process. The structure of the 9DB1 T29C, T7C deoxyribozyme mutant with ligated RNA product is almost identical to the original 9DB1 as the structures can be superimposed with the root mean square deviation of 0.81 Å (Figure 72).

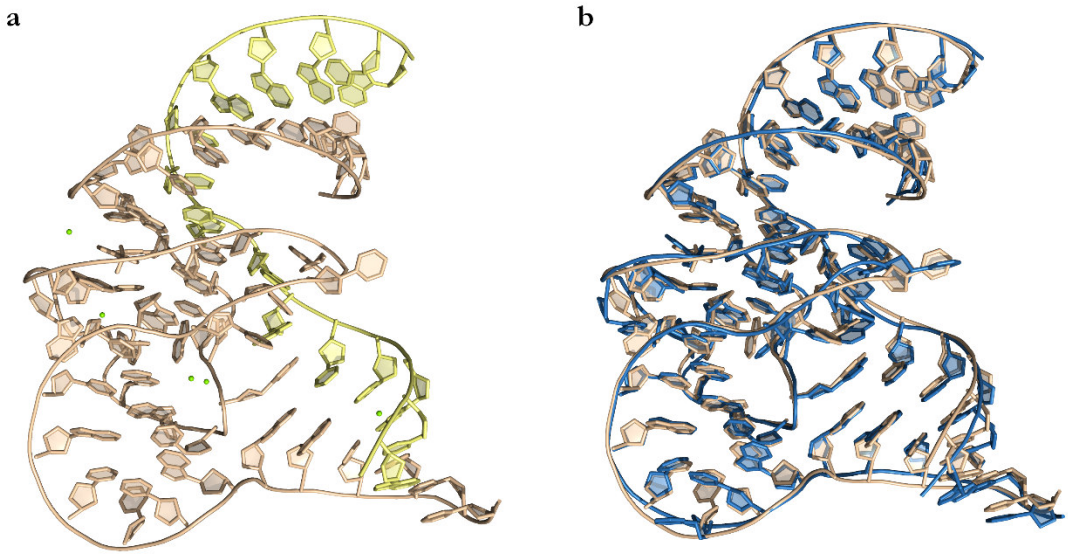


Figure 72. Crystal structure of the 9DB1 T29C/T7C mutant with ligated RNA product.
Overall crystal structure of the 9DB1 t29C T7C deoxyribozyme mutant with the ligated product. The two RNA substrates were ligated during the crystallization process. (b) Superimposed crystal structures of the T29C/T7C 9DB1 mutant and original 9DB1 (PDB code:5CKK) deoxyribozymes.

Table 10. Data collection and refinement statistics of the 9DB1 T29C, T7C mutant.

	9DB1 T29C, T7C mutant
Data collection	
Wavelength (Å)	0.999 Å
Resolution range	44.23 - 2.83 (2.99-2.83)*
No. mol/AU	1
Space group	P 43 21 2
Unit Cell	
a, b, c (Å)	78.53 78.53 53.54
α, β, γ (°)	90 90 90
R-merge	0.95 (>1.0)
Unique reflections	4315 (607)
Mean I/σI	23.4 (1.5)
Redundancy	24.6 (23.8)
Completeness (%)	99.8 (99.0)
CC1/2 (%)	100 (68.5)
Refinement	
Resolution range	44.24-2.84
R-work/R-free (%)	26.60/29.62
No. reflections	4243 (403)
Number of non-hydrogen atoms	1239
DNA/RNA	1225
Ions	14
r.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.48
Average B-factor	105.61
DNA/RNA	105.31
Ions	132.25

***Statistics for the highest-resolution shell are shown in parentheses.**

While the global architecture of the 9DB1 deoxyribozyme complexes is very similar, we can observe major structural rearrangement in the ligation junction region. The mutation at position 29 from thymine to cytidine results in the formation of the dC29 and with the G1 base pair. In the original 9DB1 structure dT29 and G1 are not co-planar and therefore their hydrogen bonding deviates from a Watson-Crick wobble base pair. As a consequence of the dC29:G1 base pair formation in the 9DB1 mutant structure, the angle between A-1 and G1 increases by around 40-45° (Figure 73).

In the 9DB1 mutant structure, the stacking of nucleotides A-1 and G1 between dA15 and dG27, as observed in the original structure, disappears. In the original structure, stacking interaction involves four nucleotides (dG27, G1, A-1, and dA15) while in the mutant structure it is divided into two stacking interactions, which occurs between nucleotides dG27/G1 and dA15/A-1.

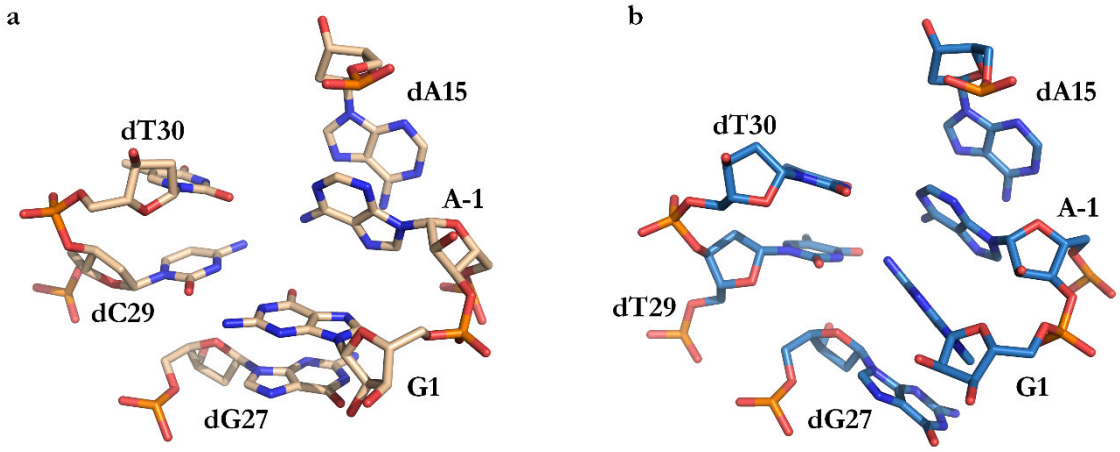


Figure 73. Comparison of the active sites.

Active site of the 9DB1 mutant T29C, T7C (a), and original 9DB1 (b)

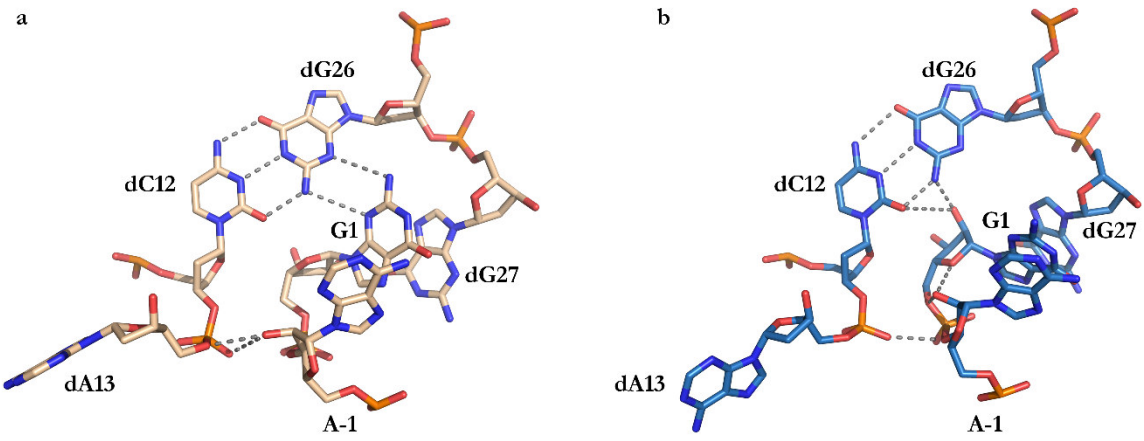


Figure 74. Hydrogen bonding interactions in the active site.

Hydrogen bonding interactions in the 9DB1 mutant T29C, T7C (a), and original 9DB1 (b) active sites.

In the mutant structure, the 2'OH of A-1 is hydrogen bonding with O5 or oxygen-phosphate-1 of the dA13 (O-O distance of 3.0 Å and 3.3 Å, respectively), while in the original structure 2'-OH is hydrogen bonding with O4 or O5 of G1 (O-O distance of 2.6 Å and 2.9 Å, respectively). Additionally, in the mutant structure, G1 is forming a triplet with the dC12:dG26 base pair via trans Sugar/Sugar (tSS) hydrogen bonding interactions, while in the original structure 2'OH group of G1 is hydrogen bonding with the sugar edge of the dC12:dG26. In both structures, O3 of A-1 is in 3.1 Å distance with the oxygen-phosphate-1 of dA13.

3.2.4 Overcoming the ligation of the RNA substrates by the 9DB1 deoxyribozyme during crystallization

The initial strategy to capture the pre-catalytic state of ligation using reactive RNA substrates did not work. To trap the 9DB1 deoxyribozyme in its pre-catalytic state for crystallization, I have generated new non-reactive RNA substrates. Two new RNA substrates were generated - (a) 6-nucleotide long left arm substrate lacks 2'OH group at the 3' end and (b) 9-nucleotide long right arm with the mono-phosphate group at the 5' end (Figure 75). Those mutations are responsible for forming incomplete reacting groups.

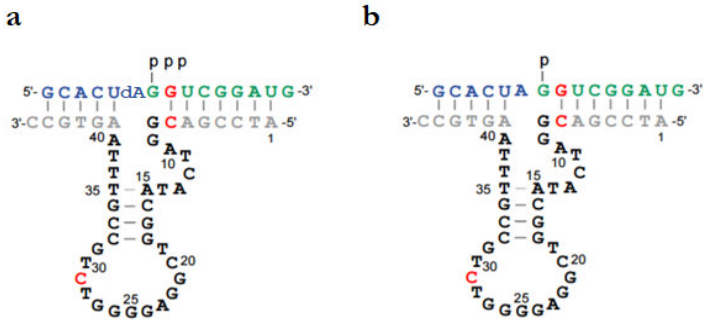


Figure 75. New generation constructs used for the investigation of the pre-catalytic state of ligation of 9DB1.

(a) The 9DB1 construct with 5'-triphosphate and a single 3'dA mutation RNA substrates (b) The 9DB1 construct with 5'-monophosphate and 3'rA RNA substrates.

Crystallization of the new generation, pre-catalytic constructs of the 9DB1

9DB1 deoxyribozyme in complex with non-reactive RNA substrates were prepared as described in section 2.2.2.1. Crystallization of the new generation, pre-catalytic constructs was performed in the same way as for the 9DB1 complex with reactive substrates. The complex with dA mutation at the end of 3' (left-arm substrate) formed the initial crystal in condition A11 of the MIDAS crystallization screen after one week (Figure 76, panel a).

For the optimization procedure, I combined the 9DB1 with two RNA substrates at 1mM concentration and mixed it with solutions containing a gradient of concentrations (30-40%) of pentaerythritol ethoxylate (5/4 PO/OH), 0.1 M HEPES pH 6.5, and 0.2 M Calcium chloride dihydrate. DNA/RNA complex and crystallization solutions were mixed in a 1:1 ratio. Drops with volumes of 2-3 μL produced microcrystals after around 7 days (Figure 76, panel b).

The crystal from the MIDAS A11 condition was subjected to a diffraction experiment on a PX Scanner X-ray diffraction system (Agilent Technologies). The diffraction image is shown in Figure 77. The obtained diffraction spots are very weak and it is hard to differentiate if it salt or macromolecule crystal.

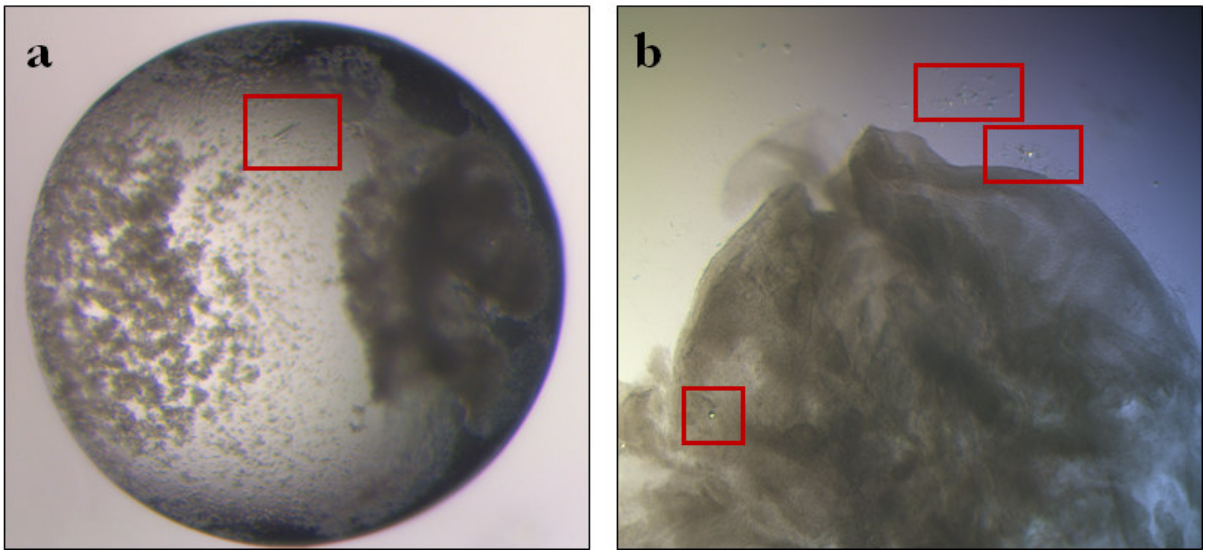


Figure 76. Initial crystals of the 9DB1 with non-reactive RNA substrates.

(a) Initial crystallization hit. Crystal has grown up in the MIDAS A11 condition (b) Microcrystals grown up during optimization of crystallization condition. Crystals are highlighted in red boxes.

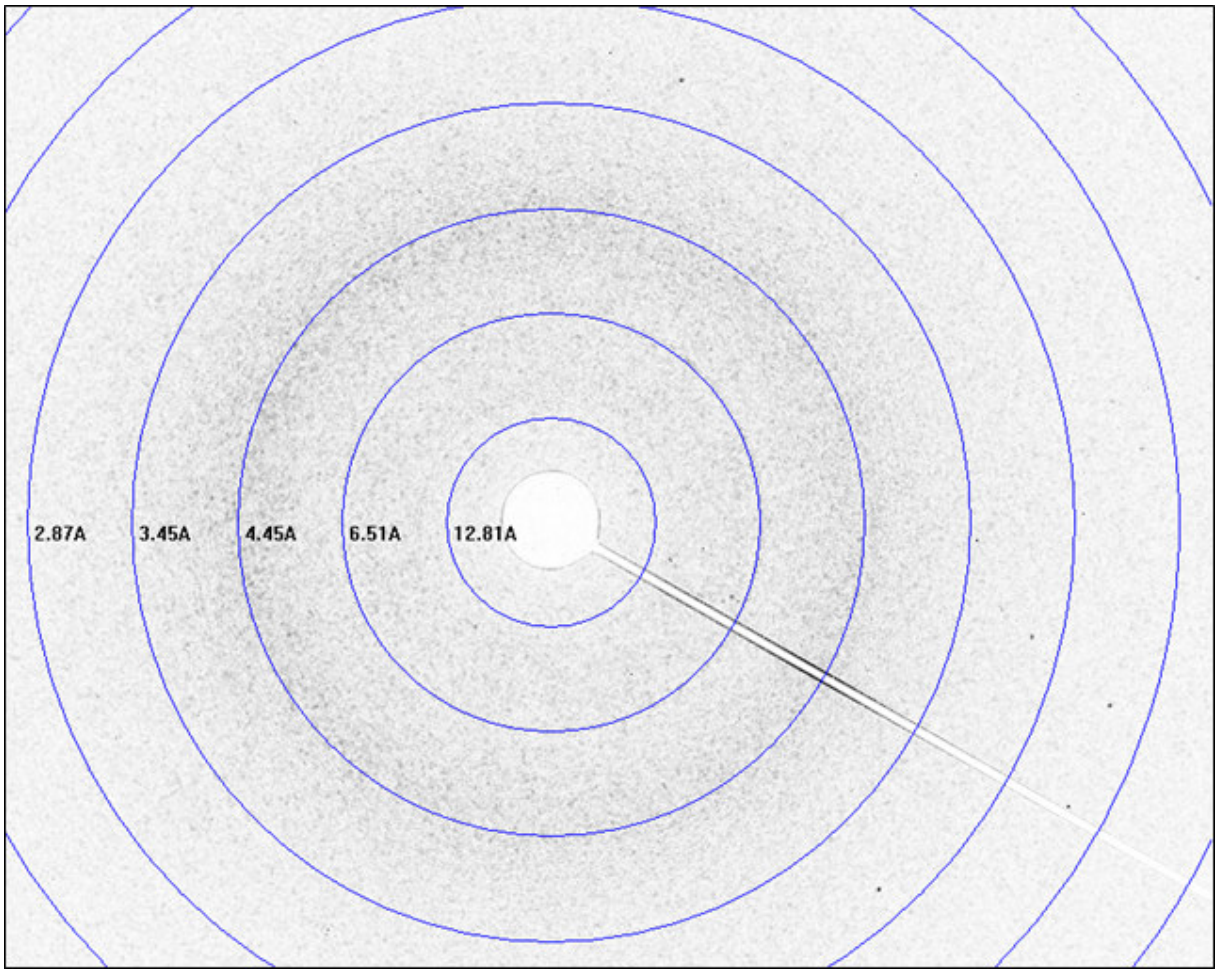


Figure 77. Diffraction pattern of the pre-catalytic 9DB1 crystal.

3.3 Structural investigation of the Chili aptamer

To reveal the fluorescence turn-on mechanism of the Chili RNA, I crystallized fluorogenic aptamer in complex with its two cognate ligands – DMHBI$^+$ and DMHBO$^+$. The 52-nucleotide long Chili RNA was produced by in vitro transcription and PAGE purified, as described in sections 2.1.4 and 2.1.5.1

3.3.1 Chili RNA aptamer crystallization

Crystallization screening

Chili DMHBI$^+$ and Chili-DMHBO$^+$ complexes, prepared as described in section 2.2.1.2, were subjected to crystallization screening at 20 °C. Two commercially available crystallization screens – Helix and Natrix 1+2, were used for initial screening. The crystallization drops consist of 120 nl of a 0.5 mM complex solution and 120 nl of crystallization buffer. Initial crystals appeared after 1 day and grow to full size in 5-7 days (Figure 78).

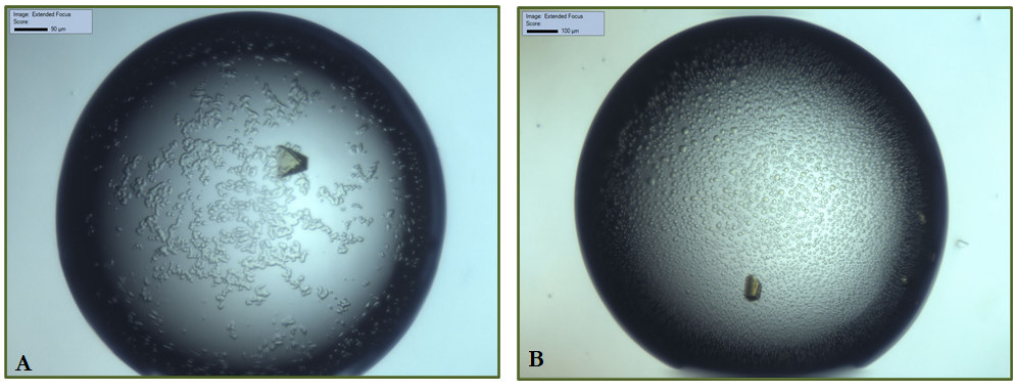


Figure 78. Crystals of CHILI aptamer in complex with DMHBI$^+$ (A) and DMHBO$^+$ (B). Crystallization conditions: 15 % PEG 400, 50 mM MES pH 7.0, Spermine tetrahydrochloride 12 mM.

Optimization of crystallization condition of the Chili RNA aptamer and heavy metal derivatization

For the crystallization optimization procedure, concentrations of the Chili–DMHBI$^+$ (Crystal I) and Chili–DMHBO$^+$ (Crystal II) complexes were reduced to ~0.1 mM. Larger single crystals were grown at 20 °C using the hanging and/or sitting drop vapor-diffusion method by mixing Chili–DMHBI$^+$ and Chili–DMHBO$^+$ complexes with solutions containing 12–17% PEG 400, 0.1 M MES pH 5.4-5.8, and 12 mM spermine tetrahydrochloride in 1:1 ratio. Drops with volumes of 0.5–2 μL produced crystals after 1–2 days and larger single crystals were grown at 20 °C using the hanging and/or sitting drop vapor-diffusion method by mixing Chili–DMHBI$^+$ and Chili–

DMHBO$^+$ complexes with solutions containing 12–17% PEG 400, 0.1 M MES pH 5.4–5.8 and 12 mM spermine tetrahydrochloride in 1:1 ratio. Drops with volumes of 0.5–2 µL produced crystals after 1–2 days and reached full size after 5-7 days.

The incorporation of heavy metal ions into the crystals was achieved by soaking and by co-crystallization. The co-crystallization with heavy metal was performed at 20 °C using the sitting drop vapor-diffusion method by mixing Chili–DMHBO$^+$ complex with solutions containing 12–17% PEG 400, 0.1 M MES pH 5.4–5.8, 12 mM spermine tetrahydrochloride, and 1 mM Iridium (III) hexamine, prepared according to (Batey and Kieft 2016) in 1:1 ratio. Before data collection, Chili–DMHBO$^+$ Iridium (III) hexamine co-crystals (Crystal III) were back-soaked in the mother liquor without Iridium (III) hexamine containing 30% glycerol and flash frozen.

Soaking of the crystal was done by transferring the native crystal to the mother liquor solution containing 20% glycerol and 0.1 mM Iridium (III) hexamine (Crystal IV). After 10 minutes, the crystal was back-soaked in mother liquor solution containing 30% glycerol and flash-frozen in liquid nitrogen.

3.3.2 Data collection and structure determination

Diffraction data were collected at 100 K on EIGER16M and PILATUS 6M detectors at the X10SA (SLS) or P11 (DESY) beamlines. Data were indexed, integrated, and scaled with XDS, and reduced with Pointless, Aimless, and Ctruncate within the CCP4 package. Structure determination of the Chili RNA aptamer was performed as described in section 2.2.3.2.1. Data collection and refinement statistics are summarized in Table 11.

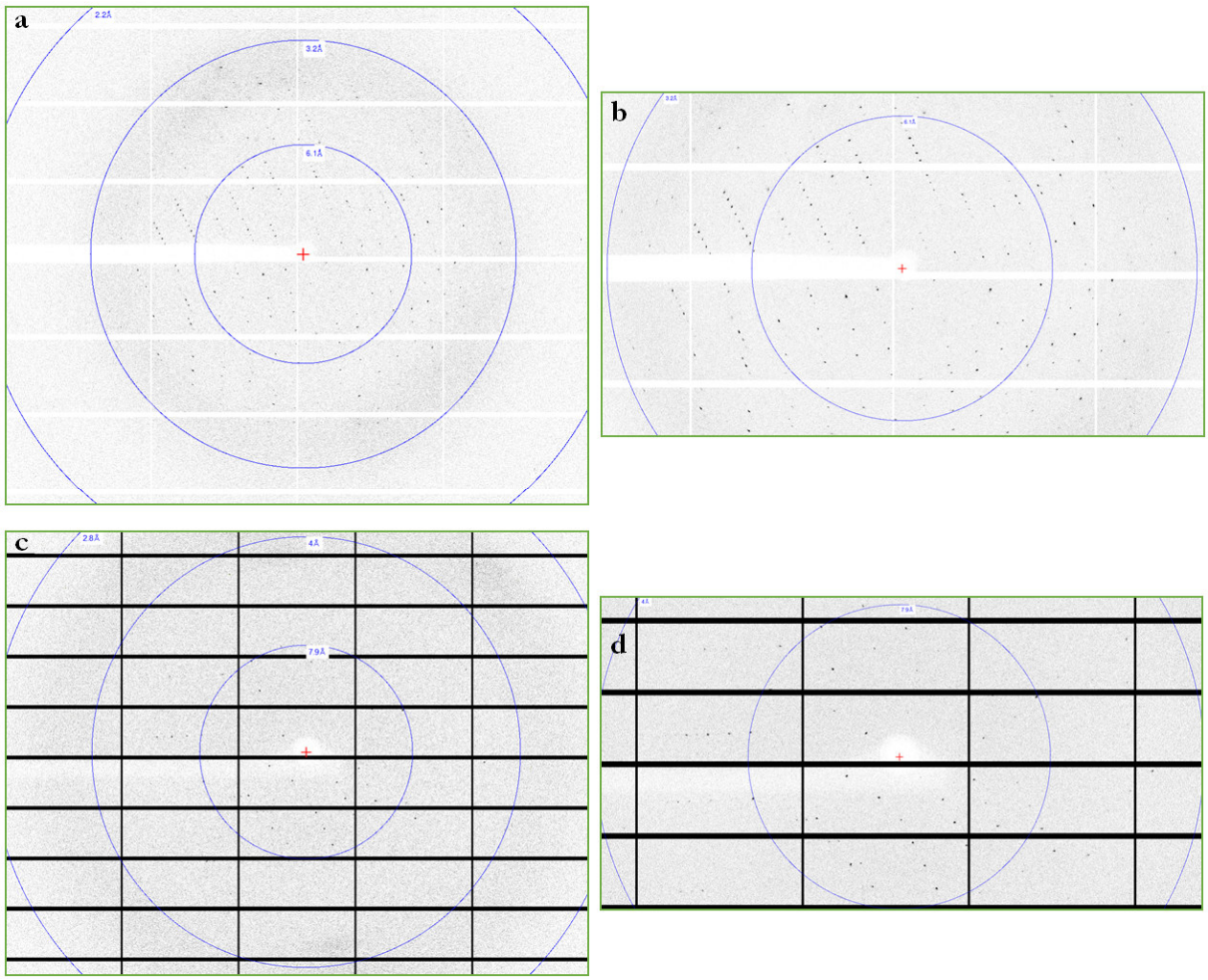


Figure 79. Diffraction pattern of the Chili RNA aptamer.

(a,b) Diffraction pattern of native Chili-DMHBO$^+$ complex and its close up view. (c,d) Diffraction pattern of Chili-DMHBO$^+$ complex co-crystallized with Iridium (III) hexamine and its close up view.

Table 11. Data collection and refinement statistics.

	Chili DMHBI$^+$ complex (Crystal I)	**Chili DMHBO$^+$ complex (Crystal II)**	**Chili DMHBO$^+$ Iridium co-crystallized form (Crystal III)**	**Chili DMHBO$^+$ Iridium soaked form (Crystal IV)**
Data collection				
Wavelength (Å)	1.0332	0.9998	1.1035	1.1048
Resolution range	49.48 - 2.89 (3.06 - 2.89)*	40.97 - 2.24 (2.34 - 2.24)	45.92 - 2.80 (2.95 - 2.80)	46.14 - 2.85 (3.00 - 2.85)
No. mol/AU	4	4	2	4
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1	I 1 2 1	P 1 2_1 1
Unit Cell				
a, b, c (Å)	59.07 100.03 113.87	59.33 100.35 113.29	70.61 54.07 87.53	57.19 106.76 63.55
α, β, γ (°)	90 90 90	90 90 90	90 95.49 90	90 116.53 90
R-merge	0.195 (>1.0)	0.040 (0.954)	0.065 (>1.0)	0.116 (>1.0)
Unique reflections	15716 (2474)	25428 (1271)	8215(1177)	15836 (2270)
Mean I/σI	8.2 (0.4)	19.3 (1.9)	13.0 (0.6)	8.6 (0.4)
Redundancy	13.0 (12.6)	6.6 (6.3)	13.4 (11.7)	7.1 (7.3)
Completeness (%) (Spherical)	99.7 (99.0)	76.7 (24.9)	99.7 (98.0)	98.7 (97.7)
Completeness (%) (Ellipsoidal)	-	94.5 (91.5)	-	-
CC1/2 (%)	99.9 (31.7)	100 (80.6)	100 (53.4)	99.8 (21.2)
Refinement				
Resolution range	46.45 - 2.95	40.98 - 2.25	45.92 - 2.80	46.14 - 2.99
R-work/R-free (%)	21.84/26.47	20.21/23.73	22.50/24.13	21.78/25.61
No. reflections	14204 (986)	25382 (706)	8166 (780)	15583 (1400)
No. of non-hydrogen atoms	4651	4773	2307	4655
Macromolecules	4378	4400	2170	4389
Ligands/Ions	273	336	137	264
Water	0	37	0	2
r.m.s deviations				
Bond lengths (Å)	0.001	0.002	0.013	0.002
Bond angles (°)	0.46	0.54	0.63	0.65
Average B-factor	103.64	68.46	155.49	151.24
RNA	102.45	68.39	153.64	149.59
Fluorophore	93.77	52.51	162.46	142.10
Ligands/ions	111.24	69.62	197.65	151.47
Water	-	53.83	-	85.15

***Statistics for the highest-resolution shell are shown in parentheses.**

3.3.3 Chili RNA aptamer structure analysis

Overall structure of the Chili RNA aptamer bound to DMHBO$^+$ and DMHBI$^+$

The crystallographic asymmetric unit (ASU) of the native dataset and Ir(III)-soaked unit cell contained four copies of RNA-ligand complexes, while the Ir(III)-co-crystallized crystal form contained only two copies (Figure 82). The structures of the individual copies of the Chili-DMHBO$^+$ complex in the ASU are very similar, as they can be superimposed with a root-mean-square-deviation of ~0.3 Å. The structural features of the DMHBO$^+$ and DMHBI$^+$ complexes are also highly comparable (Figures 83 and 84).

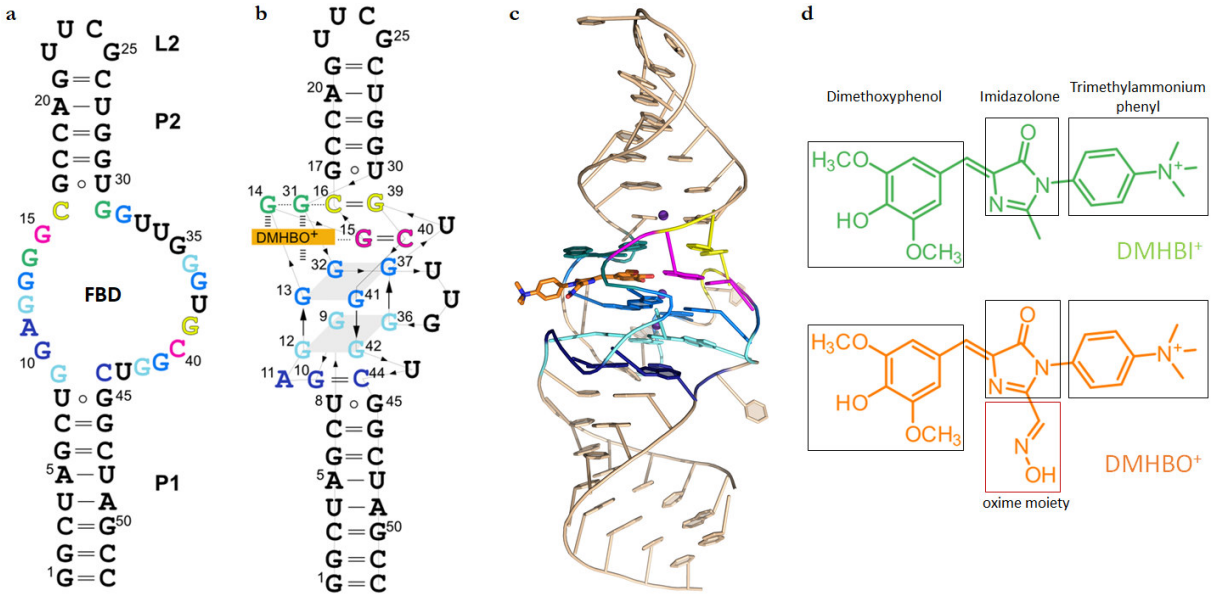


Figure 80. The overall structure of the Chili–DMHBO$^+$ and Chili–DMHBI$^+$ complexes.

(a) Sequence of the 52-nucleotide Chili RNA. The Chili stems P1 and P2 are flanking the central fluorophore binding domain (FBD) (b) Tertiary fold diagram of Chili–DMHBO$^+$ complex. Thin lines with embedded arrows indicate 5'-3' connectivity. (c) Crystal structure of the Chili-DMHBO$^+$ complex. (d) Chemical structures of DMHBO$^+$ and DMHBI$^+$. Functional groups of the fluorophores are highlighted in boxes.

The Chili RNA folds into a single coaxial helical stack with a length of approximately 70 Å that is composed of two A-form duplexes (basal 8 bp stem P1 (nucleotides 1–8/45–52), and apical 5 bp stem P2 (nucleotides 17–21/26–30) that is closed by the UUCG tetraloop L2 (nucleotides 22–25) (Figure 81). The stems P1 and P2 are separated by the central fluorophore binding domain (FBD) that accommodates the ligand (Figure 81b,d,e). P1 and P2 each contain a terminal single G:U Wobble base pair (U8:G45 and G17:U30, respectively) flanking the FBD. The P1 duplex transitions via a base triple into a two-tiered G-quadruplex, which constitutes the core of the binding site. The base triple is composed of G10:C44:A11, in which the Hoogsteen edge of A11

contacts the minor groove sugar edge of the cis Watson-Crick G10:C44 base pair, overall forming six (cWW/cSW) hydrogen bonds. The G-quadruplex is formed by T1 with guanines G9/G12/G36/G42, and T2 with guanines G13/G32/G37/G41, and is stabilized by a central K^+ ion. The trinucleotide loop U33-U34-G35 is well-resolved in one of the copies in the ASU and partially disordered in the others. The ligand is immobilized by π-stacking between an unusual G14:G31 base pair and the T2 quartet. The latter also provides the platform for stacking of the long-range canonical Watson-Crick base pair G15:C40, which is followed by C16:G39 to elongate the stem P2 (Figure 80a). The G17:U30 wobble base pair provides the binding site for an iridium hexamine complex in the major groove of the apical A-form helix. The tetraloop L2 adopts a conventional UNCG fold (Ennifar, Nikulin et al. 2000) in which G25 is in *syn* conformation and the nucleobase forms H-bonds with the ribose edge of U22.

Organization of the Chili Fluorophore Binding Domain

The architecture of the two-tiered G-quadruplex core in the Chili aptamer is characterized by three consecutive guanine steps (G12-G13, G36-G37, G42-G41) with mixed parallel and anti-parallel strand orientation, and one nonconsecutive (G9/G32) edge (Figure 80b,c). All of the guanine residues forming the G-quartets are in *anti*-conformation (torsion angle χ = -120 – -160°) except G32, which is in *high anti*-conformation (torsion angle χ = -79°). The G-quartets T1 and T2 have an opposite polarity of the Hoogsteen hydrogen bonding pattern and present a 5-5 stacking geometry featuring a partial overlap of the 5-membered imidazole rings (Figure 81c).

The G-quartets of the ligand-binding site in the Chili RNA is stabilized by a central octacoordinated K^+ ion (M_A, Figure 81c) with average K^+–guanine-O6 distances of 2.8 Å for T1 and 3.0 Å for T2. Notably, above the G-quadruplex and along the same central axis, a second K^+ ion (M_B, Figure 81d) is coordinated to T2 (average K^+ - guanine O6 distance of 2.8 Å), the phenolic hydroxy, and one methoxy group of the ligand (2.8 Å) as well as O6 of G15 (3.1 Å). The inter-K^+ distance is 3.5 Å.

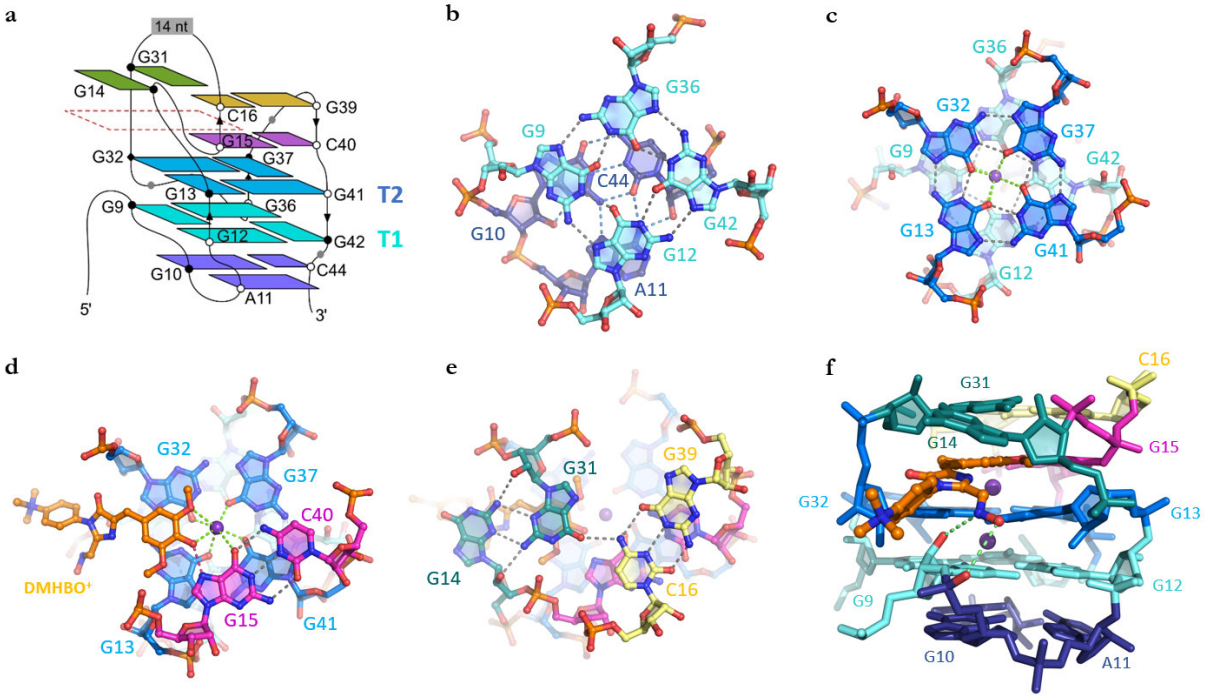


Figure 81. Organization of the Chili aptamer Fluorophore Binding Domain.

(a) Schematic representation of the connectivity and stereochemistry of the FBD. White and black circles denote C3'-*endo* and C2'-*endo* ribose puckers, respectively. Black outline denotes *anti*-conformation (all guanines in T1 and T2 are *anti*, except G32, which is *high anti*, as denoted by black diamond). The position of the ligand is outlined with the red dashed line. The color code of the nucleobases is as in Figure 80 and is also used in panels b to f. (b) G-quartet T1 and the G10:C44:A11 base triple. (c) T1 and T2 with the central potassium ion M_A, showing the opposite polarity of the quartets with partial 5-5 stacking geometry. Grey and green dashed lines represent hydrogen-bonding and inner-sphere cation coordination, respectively. (d) The DMHBO$^+$ ligand is stacked on G-quartet T2 and coordinated with potassium ion M_B and the G15:C40 base pair. The red dashed line represents a hydrogen-bond from the hydroxy group of the ligand to N7 of G15. (e) A trans Sugar/Sugar edge (tSS) base pair of G14 and G31 stabilizes DMHBO$^+$ via π-stacking interactions. O6 of G31 forms an additional hydrogen bond with the amino group of C16 in the C16:G39 base pair. (f) Stick representation of the ligand-binding site showing hydrogen bonding interactions between the oxime moiety of the ligand and the RNA backbone.

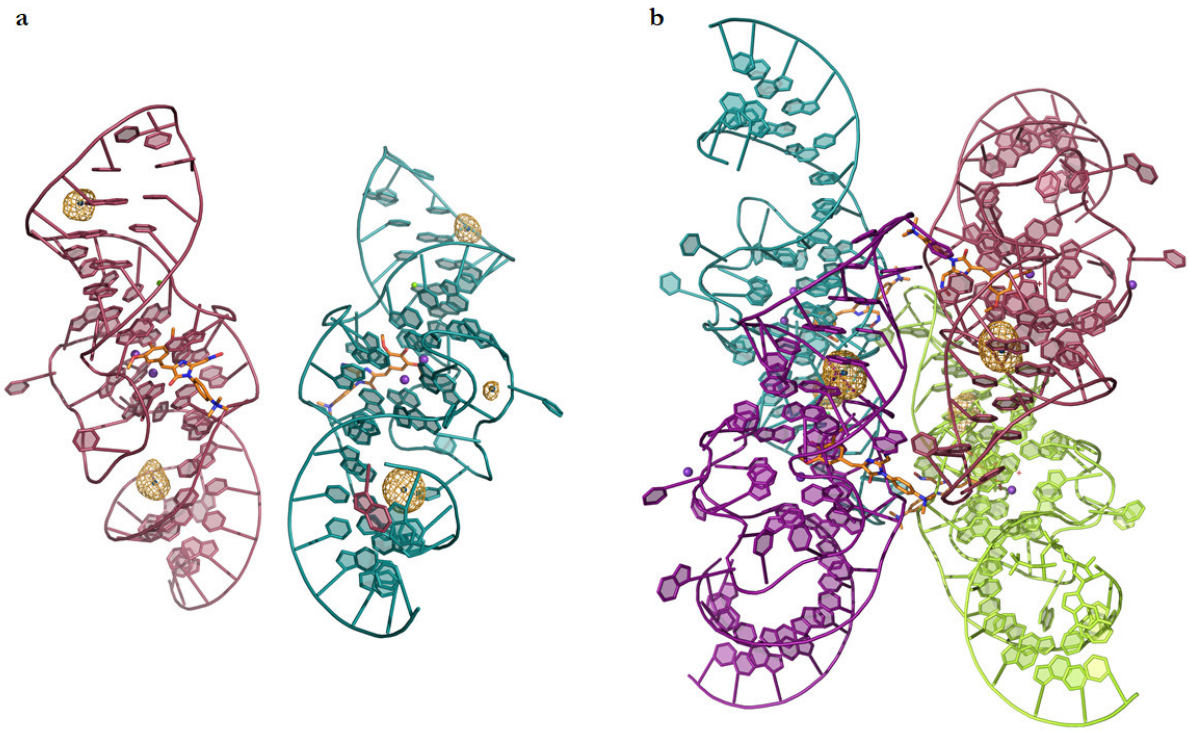


Figure 82. Overall crystal structures of the Chili-DMHBO$^+$ Iridium derivatives.

(a) Overall crystal structure co-crystallized with Iridium (III) hexamine. The AU contains two molecules of Chili-DMHBO$^+$ molecules (pink and blue) (b) Overall crystal structure soaked with Iridium (III) hexamine. The AU contains four molecules of Chili-DMHBO$^+$ molecules (violet, green, pink, and blue). Yellow mesh indicates anomalous difference Fourier map contoured at 5σ, superimposed around Ir^{3+} cations. The difference map was computed from data collected at the Ir L-III edge.

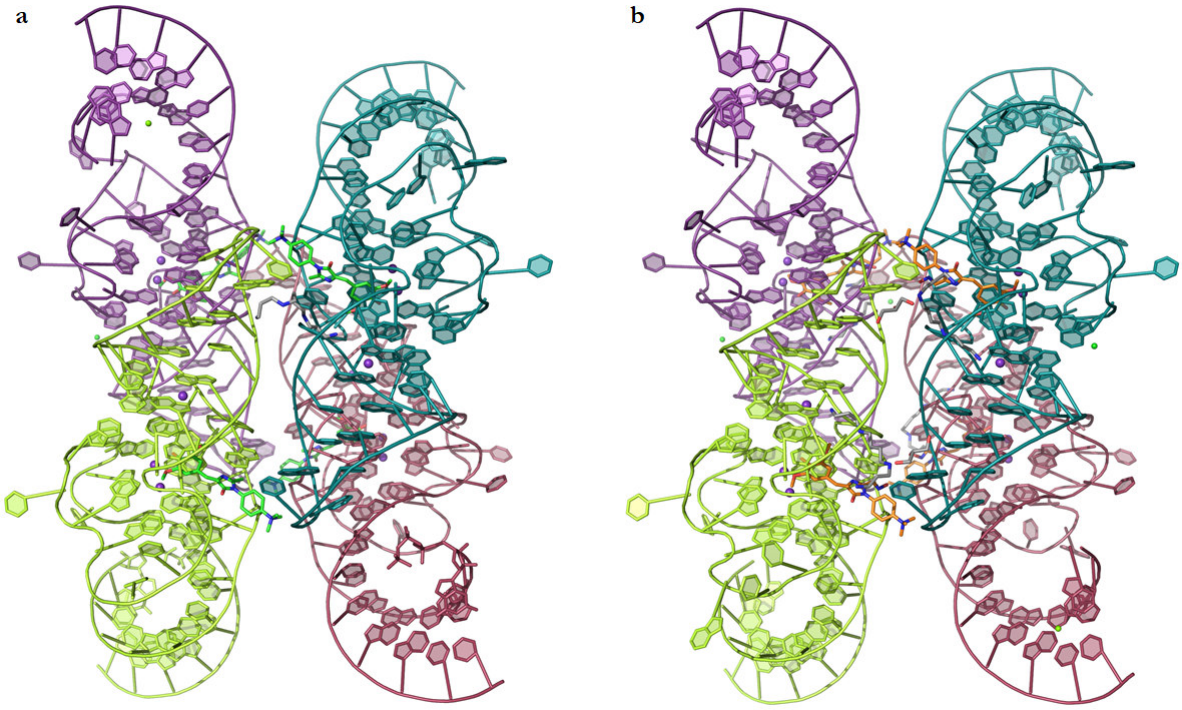


Figure 83. Overall crystal structures of the (a) Chili-DMHBO$^+$ and (b) Chili-DMHBI$^+$.

The crystallographic asymmetric unit contains four molecules of the Chili-fluorophore complex (green, violet, blue, and pink). The DMHBI$^+$ and DMHBO$^+$ fluorophores are colored in green and orange respectively.

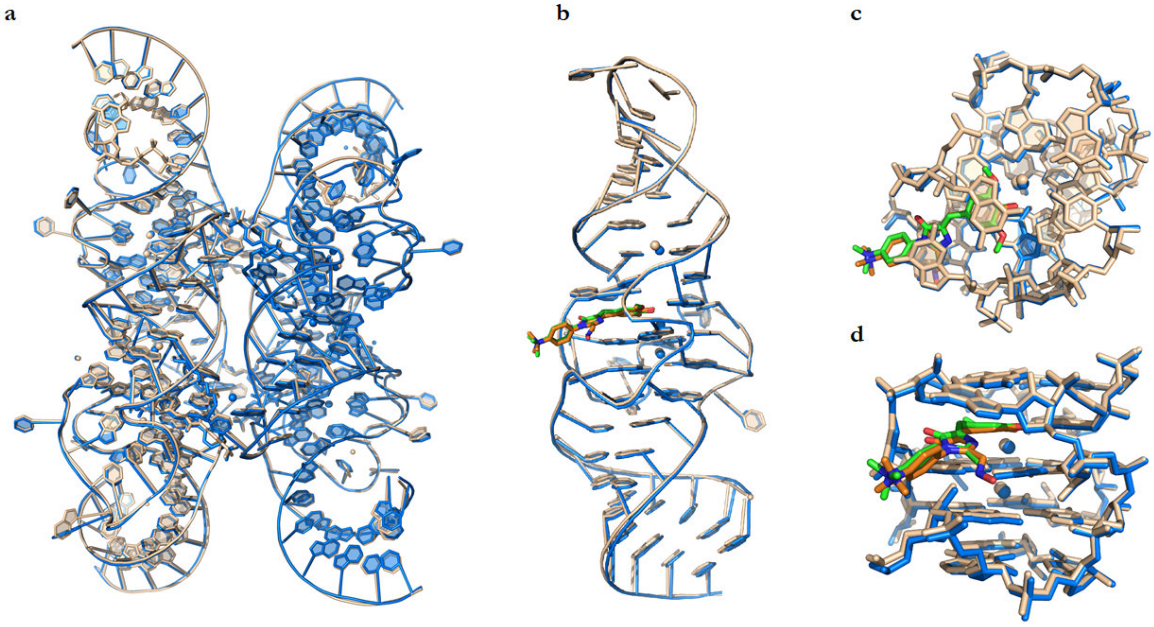


Figure 84. Comparison of Chili−DMHBO$^+$ and Chili−DMHBI$^+$ complexes.

Superimposed asymmetric unit molecules of the Chili RNA aptamer. Chili RNA is represented in blue for the DMHBO$^+$ complex and in beige for the DMHBI$^+$ complex. DMHBO$^+$ and DMHBI$^+$ fluorophores are

colored in orange and green respectively. (b) Superimposed single chains (Chain A) of Chili RNA aptamer with bound fluorophores. (c, d) Top and side view for the fluorophore binding site.

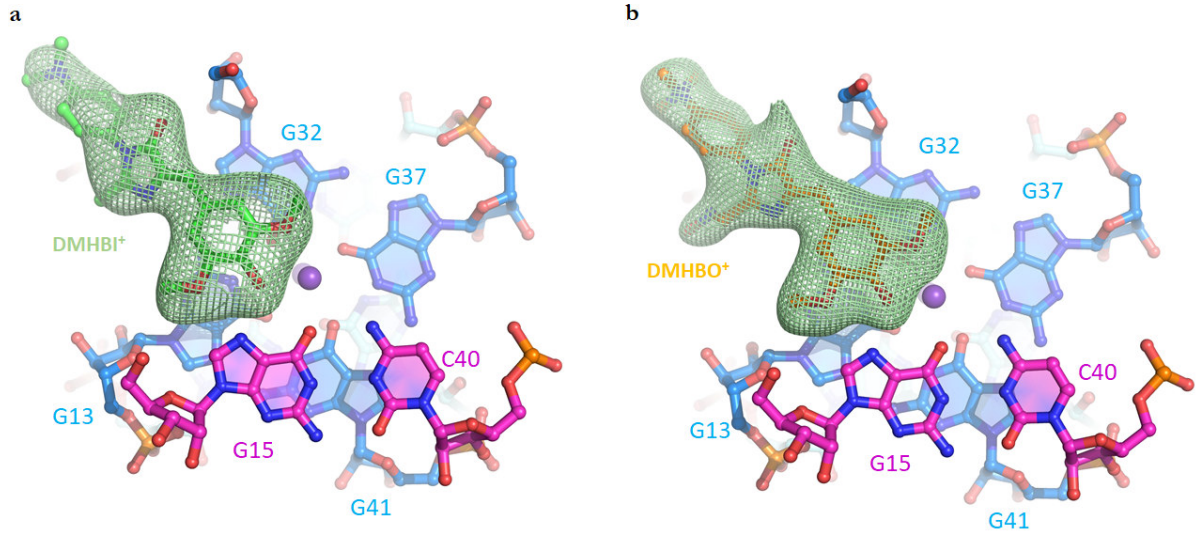


Figure 85. Comparison of the fluorophore binding site.

DMHBI+ (a) and DMHBO+ (b) binding site. The green mesh indicates a polder omit map contoured at 4σ.

Electrostatic interactions of the phosphate backbone of C24 and G25 in L2 with the positively charged trimethylammoniumphenyl side chain of a ligand bound to a neighboring Chili molecule contribute to inducing the head-to-head orientation of two copies of the RNA-ligand complex in the ASU. An intermolecular H-bond between the 2'-OH of G25 with N7 of G14 in the neighboring molecule constitutes an additional inter-subunit contact in the crystal lattice (Figure 86). An additional π-π stacking interaction between two guanosines in the P1 stems are responsible for tail-to-tail orientation of the Chili molecules in the crystal (Figure 87).

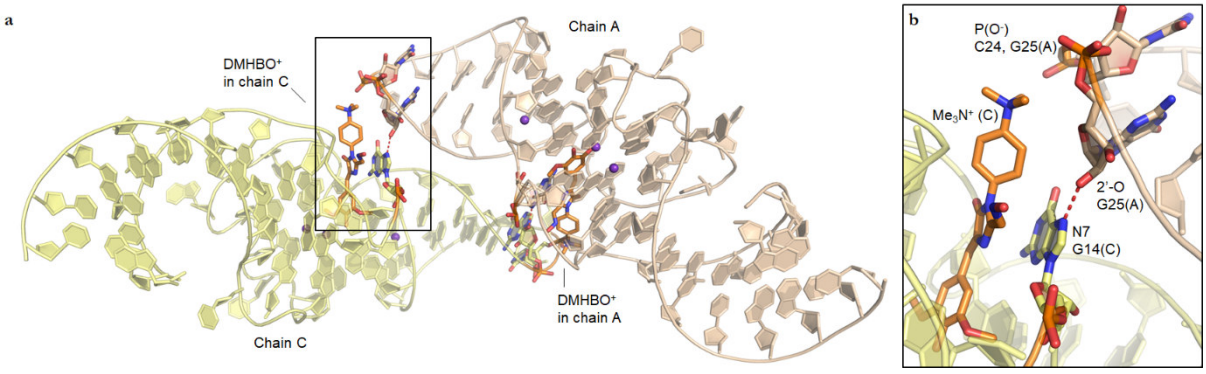


Figure 86. Intermolecular interactions between two copies of Chili RNA ligand complexes.

(a) Intermolecular contacts between two Chili RNA chains responsible for head-to-head orientation. The hydrogen bond is formed between 2'-OH of G25 with N7 of G14 (O-N distance 2.6 Å) and the positively charged side chain of a ligand attracts the negatively charged backbone in L2. (b) Close up view for the intermolecular contacts between neighboring Chili molecules.

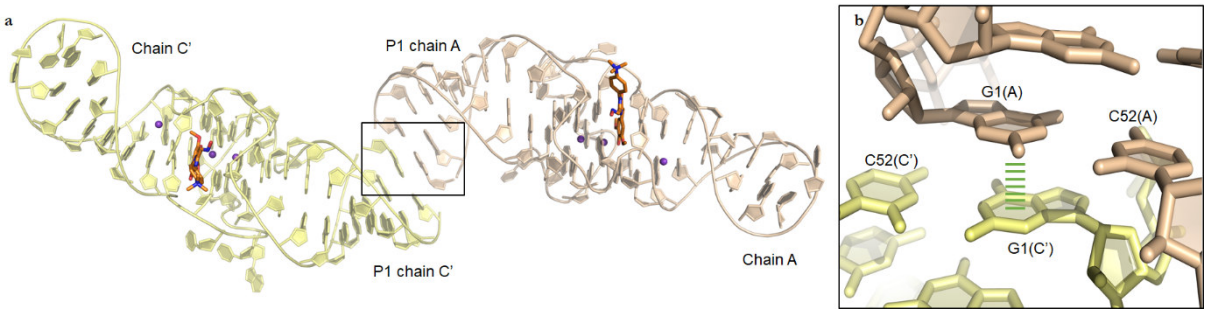


Figure 87. π-π stacking intermolecular interactions between two copies of Chili RNA.

(a) Intermolecular interactions π-π stacking between P1 stems of the Chili RNA responsible for tail-to-tail orientation. (b) Close up view for the intermolecular contacts between neighboring Chili molecules.

Due to the stacking interactions of dimethoxyphenol and imidazolone moieties with nucleobases laying in distinct planes, the $DMHBO^+$ ligand adopts a nonplanar confirmation within the binding site with twist and tilt angles of $\varphi = -31°$ and $\tau = 12°$, respectively (Figure 88a). Thus, the benzylidene moiety of the ligand stacks on top of G13:G32 of the T2 quartet and G31 in the direction of the apical stem. The imidazolone moiety projects outward from the center of the quadruplex where it is anchored by stacking with G14 (Figure 88b).

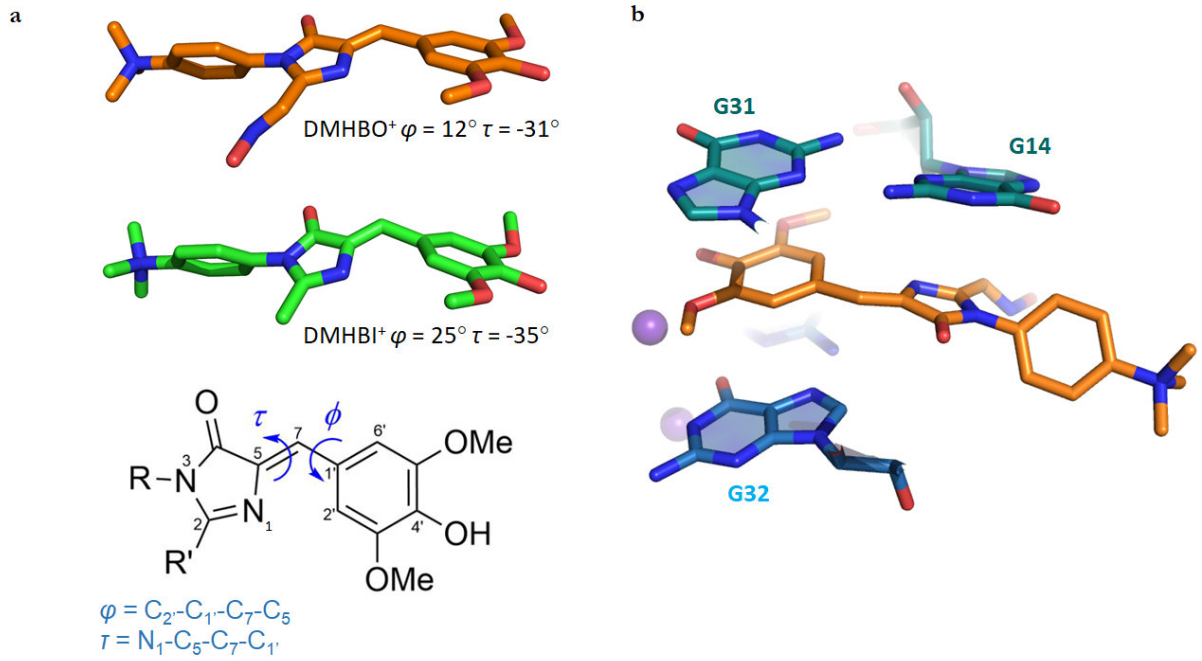


Figure 88. Conformations of $DMHBO^+$ and $DMHBI^+$ fluorophores.

(a) Torsion angles φ and τ, and representative $DMHBO^+$ / $DMHBI^+$ ligand conformations in the chain A (b) Relationship between torsion angles and stacking with nucleobases.

The G14:G31 base pair is stabilized by a trans Sugar/Sugar edge (tSS) interaction that is mediated by mutual hydrogen bonds between the NH_2 of either guanine and N3/2'-O of the other guanine

nucleotide. G31 is further held in place by a hydrogen bond between its O6 and N4 of C16, which itself is part of a long-range Watson-Crick base pair with G39 (Figure 81e).

The cationic trimethylammoniumphenyl side chain of the ligand is located outside the perimeter of the binding pocket and is twisted at an angle of 60° concerning the plane of the imidazolone ring (Figure 81d, f).

Besides stacking interactions, the imidazolone and phenolic moieties establish polar interactions with the RNA (Figure 81d). Thus, the oxime hydroxy group of the imidazolone ring forms two hydrogen bonds with the phosphate backbone at G10 and the 2'-OH group of G9 (Figure 81f). These contacts explain the difference in binding affinity between $DMHBO^+$ (K_D = 12 nM) and $DMHBI^+$ (K_D = 65 nM) (Steinmetzger, Palanisamy et al. 2019). The dimethoxyphenol moiety of the ligand is coplanar with the long-range Watson-Crick base pair G15:C40 and forms polar contacts with the metal ion M_B, via the phenolic hydroxy group and one of the methoxy groups. Most notably, the phenolic hydroxy group comes into close contact with N7 of G15 (2.5 Å O–N distance) (Figure 81d), indicative of a mechanism that governs the protonation state underlying the exceptional fluorescence properties of the Chili–$DMHBO^+$ complex.

4 DISCUSSION

4.1 STRUCTURAL INVESTIGATION OF THE RNA-CLEAVING VMC10 DEOXYRIBOZYME

The VMC10 deoxyribozyme catalyzes the site-specific cleavage of an RNA substrate between guanosine and adenosine. It was previously shown that the VMC10 deoxyribozyme is strongly inhibited by the presence of the methylated nucleoside m6A at the cleavage site. Here, the goal was to elucidate the structural basis for recognition of the methylated nucleobase by solving the crystal structure of the deoxyribozyme in complex with an uncleavable analog of the RNA substrate, containing either methylated and unmethylated adenosine. Surprisingly, the RNA substrate dissociated from the deoxyribozyme during the crystallization process. This happened for several of the investigated constructs, resulting in three different RNA-only structures. Two structures were solved for unmethylated RNAs and one m6A-containing RNA structure will be discussed.
The RNA substrates sequences from generations 1-3 were designed using the original substrate sequence used for the in vitro selection procedure (Sednev, Mykhailiuk et al. 2018). To increase the chance of successful crystallization, the RNA substrates used for the crystallization contained a single 2'-deoxyguanosine mutation at the cleavage site (dG|A) to prevent the cleavage of the substrate, resulting in a complex that mimics the pre-catalytic state of the RNA-deoxyribozyme complex. Additionally, the unmethylated RNA and the methylated RNA substrates were used in the crystallization process. Crystallization of the second and third generations of the VMC10-substrate constructs resulted in the dissociation of the RNA substrates from the deoxyribozyme. Surprisingly, despite the sequence identity, the R616 and R633 RNA substrates formed different structures. The only difference between the two RNA substrates is that R633 contains methylated adenosine at position 8. The unmethylated RNA substrate (R616) crystallized in the primitive hexagonal space group, while the methylated substrate (R633) crystallized in the primitive trigonal space group. The R616 substrate formed an RNA dimer while the R633 substrate formed a homo-tetramer. There are two main reasons why both substrates formed different structures. In the R616 structure, the adenosine 8 is forming intermolecular hydrogen bonding contact with the uridine 12. This intermolecular interaction is crucial in the orientation of the RNA duplexes in the crystal packing. The presence of the N6-methyladenosine at position 8 would result in the steric clash between N6 of adenosine and O2 of uridine, therefore intermolecular interaction between A8 and U12 cannot be formed in R633. Additionally, the R633 substrate crystallized in the low-pH (pH 4.5), which allowed the formation of the C:G*C^+ triplets, stabilizing the core of the RNA-tetramer. Despite the difference in the global architecture of R616 and R633, both structures contain a non-canonical trans Hoogsteen/sugar edge (tHS) A:dG base pair also called sheared base pair, where

the sugar edge of 2'-deoxyguanosine contacts the major groove of the adenosine (Figure 89 a,b). The sheared A:G base pair is a highly conserved element of a kink-turn motif in the RNA (Klein, Schmeing et al. 2001). Huang et al. have shown that the inclusion of the N6-methyladenine, completely disturbs the sheared A:G base pair in the k-turn, which prevails formation of the box C/D snoRNP (Huang, Ashraf et al. 2017). In the crystal structure of duplex RNA with unmodified adenine (Figure 89c; PDB: 5LR3), the sheared G:A is formed, with the hydrogen bonds between N6/N7 of adenosine and N3/N2 of guanosine respectively. In the presence of N6-methyladenosine sheared G:A base pair is disturbed, where hydrogen bond is formed between N6 of adenosine and 2'OH group of guanosine (Figure 89d; PDB: 5LR4). In the crystal structure of R633, the sheared A:G base pair is not disturbed in the presence of N6-methylation. This could be caused by the lack of the 2'-hydroxy group of guanosine which may serve as a hydrogen bond acceptor. Additionally, the presence of the 2'-hydroxyl group could produce a steric clash with the N6-methylated adenosine leading to disruption of the sheared m6A:G base pair. In another study, Ashraf et. al have shown that it is possible to place N6-methyladenine at different positions in the k-turn structures and keep the sheared m6A:G base pairs (Figure 89e,f; PDB: 6Q8U, 6Q8V) (Ashraf, Huang et al. 2019). The sheared m6A:G base pairs from structures 6Q8V and 6Q8U have the same hydrogen-bonding pattern as the tHS base pair observed in the R633, except the 6Q8V where an additional bond between N3 of guanosine and N6 of adenosine is formed. However, we can observe the difference in the guanosine sugar pucker conformation. In the R633 structure, the guanosine forming the sheared base pair adopts C3-*endo* conformation, while in 6Q8V and 6Q8U structures, guanosine adopts C2-*endo* conformations (Figure 89 b,e,f). A similar situation can be observed for the unmodified sheared A:dG base pair in comparison to A:G in 5LR3 (Figure 89 a,c)

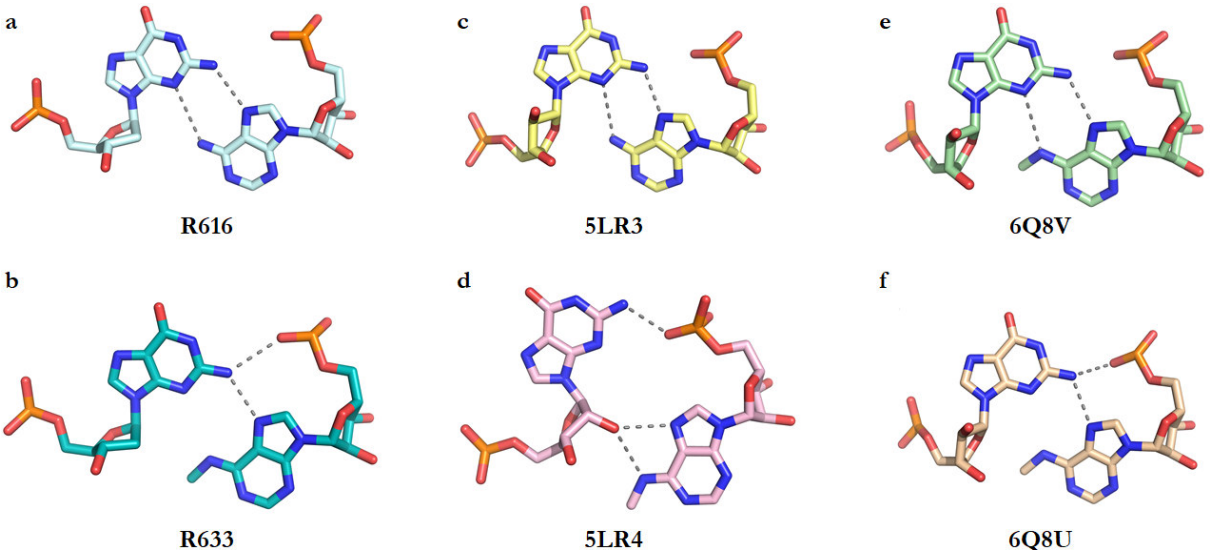


Figure 89. The sheared A/m6A:G base pairs.

The other interesting feature observed in the VMC10 methylated RNA substrate structure is the tetraloop-like motif formed in the basal and apical stems of the tetramer. The tetraloop is defined as a loop formed by four unpaired nucleotides, which caps the end of the helix. Tetraloops are one of the most abundant secondary structure motif observed in the RNA and play important role in the RNA folding, structural scaffolds, and serve as recognition sites for proteins and nucleic acids (Cheong 2010). The most common tetraloops found in the RNA (over 70%) belong to the GNRA, UNCG families (Woese, Winker et al. 1990), however many other families of RNA tetraloops have been classified. These are CUYG (Woese, Winker et al. 1990), GANC (Keating, Toor et al. 2008), (A/U)GNN (Butcher, Dieckmann et al. 1997, Wu, Yang et al. 2001), UUUM (DeJong, Marzluff et al. 2002, Zanier, Luyten et al. 2002) and UNAC (Zhao, Huang et al. 2012) tetraloop families. The use of RNA Characterization of Secondary Structure Motifs (CoSSMos) combined with Dissecting the Spatial Structure of RNA (DSSR) and a bioinformatic approach, allowed for the identification and characterization of three new tetraloop families: YGAR, UGGU, and RMSA (N=A/G/C/T, R=A/G, Y=C/T, M=A/C, S=G/C) (Richardson, Adams et al. 2019).

In the tetraloop-like motif observed in the R633 structure (Figure 90d), the sharp turn in the phosphate backbone occurs between guanosine and uridine. This RNA structural motif, which contains a sharp bend in the backbone direction between the first and second nucleotide of the UNR sequence (UGA in the R633 structure) is called the U-turn (Quigley and Rich 1976). The U-turn (Figure 90a) is stabilized by the two specific hydrogen-bonding interactions involving a 2'-hydroxyl group of uridine and N7 of the R in the UNR sequence. The second hydrogen bond is formed between the imino proton of uridine and the phosphate group of the nucleotide, immediately following the U-turn (Figure 90a,b) (Quigley and Rich 1976, Pley, Flaherty et al. 1994).

The U-turn motif was identified in the frequently occurring GNRA tetraloops (Jucker and Pardi 1995). The GNRA tetraloop (Figure 90b,e) is stabilized by the hydrogen bond interactions very similar to those observed in the U-turn motif (Figure 90a,c). In the GNRA tetraloop, the amino proton of the G1 is forming a hydrogen bond with the phosphate oxygen of the A4, while in the U-turn motif the bond is formed by the imino proton of uridine. The additional hydrogen bond interactions involved in the stabilization of GNRA tetraloop involve an amino proton of 4A and a 2'-hydroxyl group of the G1 and the trans Hoogsteen/Sugar edge base pair interaction between G1 and A4. The GNRA tetraloop is additionally stabilized by the same hydrogen bond between the N7 of R3 and 2'-hydroxyl of the G1, as observed in the U-turn motif.

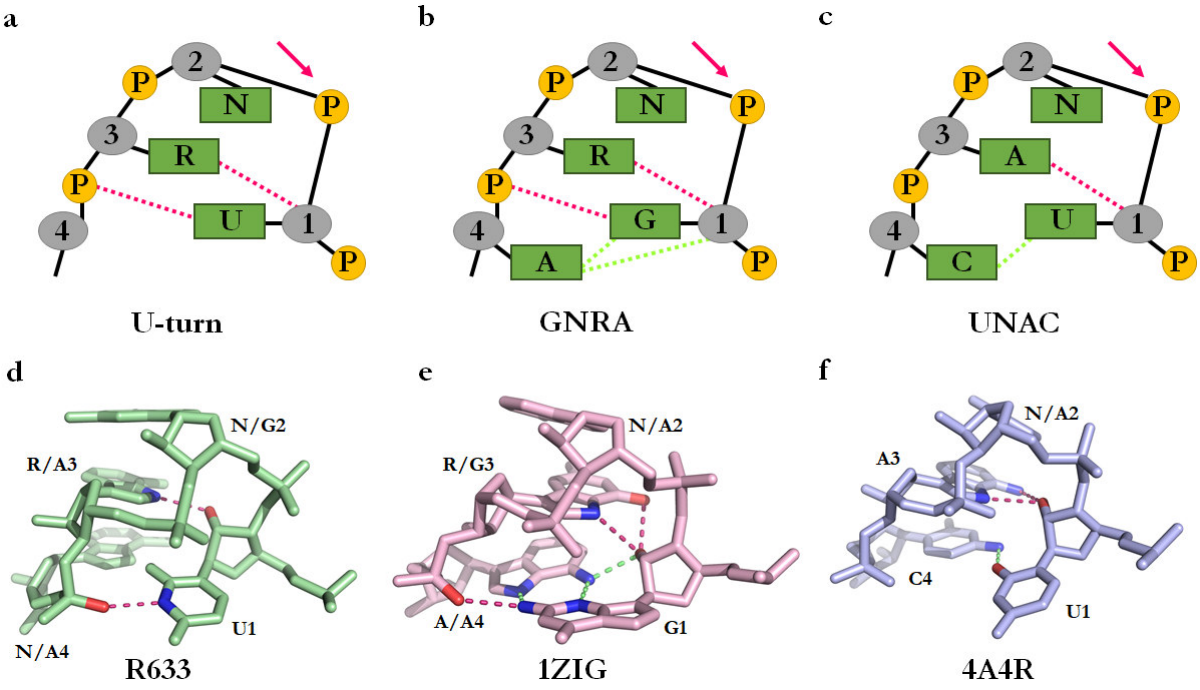


Figure 90. Comparison of different families of RNA tetraloops.

Schematic representation of (a) U-turn, (b) GNRA and (c) UNAC tetraloops. (d) The tetraloop-like motif present in the R633 structure. The NMR structures of the GAGA (d) and UNAC tetraloops with corresponding PDB codes. The red arrow indicates turning phosphate.

The UNAC tetraloops (Figure 90c,f) contain several structural features that allow them to mimic the GNRA tetraloops. Comparably to the GNRA tetraloops, the first and fourth nucleotides of the UAAC and UCAC (UMAM) tetraloops are forming at one hydrogen bond and the N7 and N6 amino proton of the third adenosine are involved in hydrogen bond interactions with a 2'-hydroxyl group of the first uridine. Additionally, the first U is stacking on the 5' of the stem and the second M is stacking on the 3' of the stem (N in the GNRA and R633 structures). In the R633 tetraloop-like structure, in contrast to the GNRA and UMAC tetraloops, nucleotides U1 and A4 are not involved in the base pair interaction. However, in all three structures, the N7 atom of a third nucleotide in the loop is forming a hydrogen bond interaction with the 2'-hydroxyl group of the first nucleotide (Figure 90d,e,f). The U-turn motif, observed in the GNRA, UMAC, and the R633 structures was also identified in the anticodon loop and the T-loop of yeast $tRNA^{Phe}$ and Hammerhead ribozyme. The unmodified $tRNA^{Phe}$ anticodon loop from the *Escherichia coli* (Byrne, Konevega et al. 2010) has an identical fold as a tetraloop-like motif observed in the R633 structure (Figure 91c).

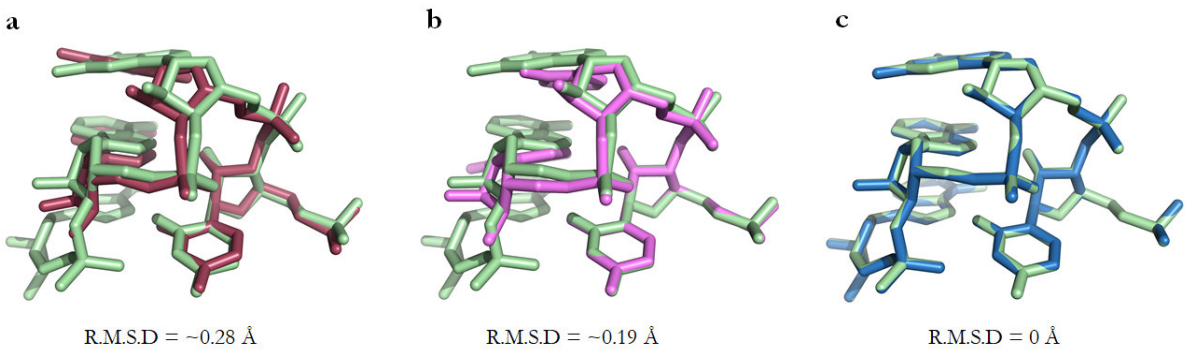


Figure 91. Comparison of the U-turns.

Superimposed R633 tetraloop-like motif against U-turn motifs from the (a) Hammerhead (HH) ribozyme (PDB:1HMH), (b) yeast $tRNA^{Phe}$ TΨC loop (PDB:1EHZ), and (c) *E.coli* $tRNA^{Phe}$ anticodon loop (PDB:3L0U). The R633 tetraloop-like motif is colored in green. The HH ribozyme U-turn motif, yeast $tRNA^{Phe,}$ and *E.coli* $tRNA^{Phe}$ anticodon loop are colored in red, pink, and blue, respectively

The selection substrate sequence used in generations 1-3 of the VMC10 deoxyribozyme constructs produced non-diffracting crystals or RNA-containing crystals. To overcome the dissociation of the RNA substrate from the deoxyribozyme during the crystallization process, the substrate sequence has been re-designed using the long noncoding RNA MALAT1 fragment. MALAT1, also known as nuclear-enriched abundant transcript 2 (NEAT2), is an 8,708 nucleotides long non-coding RNA (lncRNA) that is highly abundant in mammalian cells and it is localized to nuclear bodies (Hutchinson, Ensminger et al. 2007, Cabili, Dunagin et al. 2015). MALAT1 is a highly m6A-modified lncRNA, carrying multiple m6A modification sites (Dominissini, Moshitch-Moshkovitz et al. 2012, Meyer, Saletore et al. 2012, Coker, Wei et al. 2019). The structure of MALAT1 is largely unknown, however, structures of hairpins at nucleotides 2509-2537 (H63) and 2,556-2,586 (H64), were determined (Liu, Dai et al. 2015, Liu, Zhou et al. 2017). Hairpins H63 and H64 are forming binding sites for the HNRNPG and HNRNPC proteins, respectively.

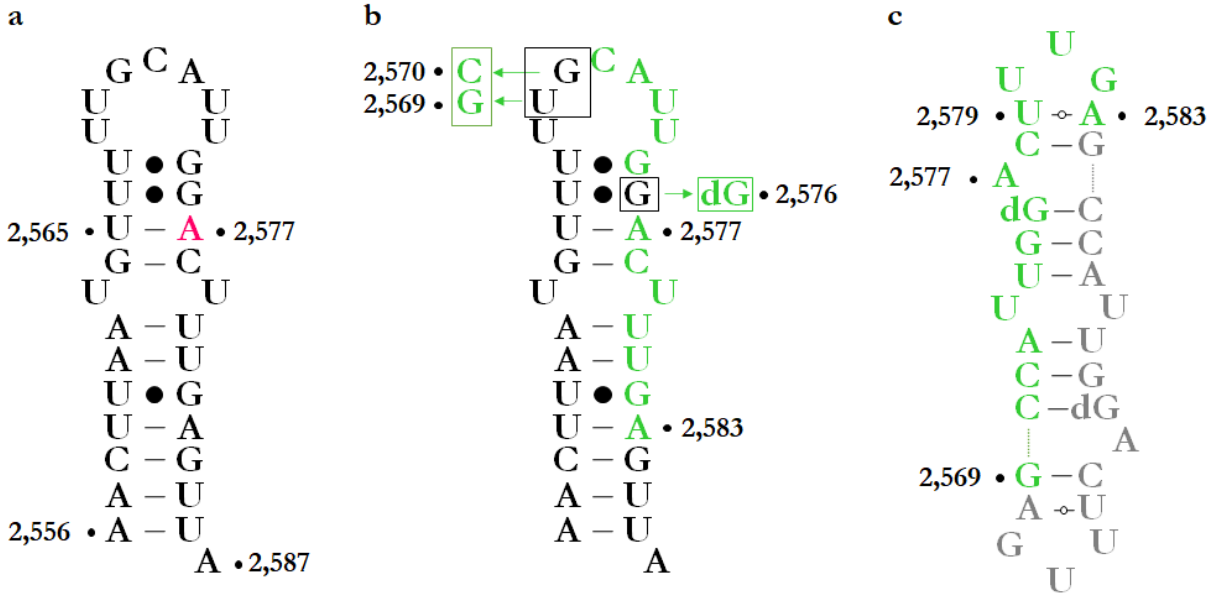


Figure 92. MALAT1 H64 hairpin stem.

(a) MALAT1 H64 hairpin stem (b) Point mutations (marked in green boxes) introduced to the MALAT1 H64 hairpin stem used for the R698 RNA (c) Structure formed by the R698. Two R698 molecules forming the A-form double-stranded RNA flanked by the two hairpin-like loops.

For the 4th generation, VMC10 DNAzyme constructs, RNA sequence coming from H64 hairpin stem was used (Figure 92a). The MALAT1 H64 contains an m6A modification site at position 2,577 (A2,577). When A2,577 is unmethylated, it is base-paired with U2,565 and therefore the HNRNPC binding domain is partially inaccessible. When the A2,577 gets methylated, HNRNPC binds a U_5-tract that is half-buried in the hairpin-stem opposing the 2,577-A/m6A site (Liu, Dai et al. 2015). The VMC10 R698 substrate sequence was obtained using MALAT1 H64 at nucleotides 2,569-2,583 (15-nucleotide long single-stranded RNA) with two point mutations at the 5' end - U2,569G, G2,570C. Additionally, the guanosine 2,576 was mutated to deoxyguanosine to prevent RNA cleavage by the VMC10 DNAzyme during crystallization (Figure 92b). Surprisingly, also in this case the RNA substrate dissociated from the deoxyribozyme, and the RNA-only structure was solved. In the R698 structure (Figure 92c), nucleotides 2,579-2,583 formed a hairpin-like loop, while in the MALAT1 H64 structure those nucleotides are forming a hairpin stem, except U2,579 forming a single-nucleotide bulge. Additionally, in the R698 crystal structure, A2,577 is forming a single-nucleotide bulge, while in the MALAT1 fragment, non-methylated A2,577 is forming a wobble base pair, stabilizing the hairpin structure. Zhou et. al, using nuclear magnetic resonance and Förster resonance energy transfer, have demonstrated that m6A selectively destabilizes the portion of the hairpin stem where the U_5-tract is located, increasing the solvent accessibility of the neighboring bases while maintaining the overall hairpin structure (Zhou, Parisien et al. 2016).

Another interesting feature observed in the R698 crystal structure is that the hairpin-like loop is forming a kissing-stem loop interaction between two molecules. A kissing stem-loop is formed when bases between two hairpin loops are forming a base pair (Nowakowski and Tinoco 1997). RNA/RNA kissing stem-loop complexes were found both in prokaryotic and eukaryotic organisms and were shown to play a variety of regulatory functions (Brunel, Marquet et al. 2002). The loop-loop interactions play important roles in the replication of retroviruses. In the human immunodeficiency virus type 1 (HIV-1), stem-loop structures known as the dimerization initiation site (DIS) cause retroviral genomic RNA dimerization by forming a transient kissing stem-loop complex between two DIS (Paillart, Skripkin et al. 1996). The crystal structures of the DIS of HIV-1 subtypes A and B have confirmed, that two molecules adopt hairpin conformation, and the kissing-stem loop complex is formed by a six base self-complementary stretch (Ennifar, Walter et al. 2001). In the Moloney murine leukemia virus, a stable homodimeric RNA kissing complex is formed through only two intermolecular G:C base pairs (Kim and Tinoco 2000, Li, Bustamante et al. 2006).

In the R698 crystal structure, the kissing stem-loop is rather unusual since the intermolecular interactions are not formed by the canonical Watson-Crick base pairs. The R698 loop-loop complex is formed by the two A:U Hoogsteen and two G:U base pairs. The G:U base pairs do not obey the traditional wobble G:U hydrogen-bonding pattern, where O6 and N1 of guanosine are forming two hydrogen bonds with N3 and O2 of uridine, respectively (Varani and McClain 2000). Instead of this, the bases are in trans configuration, and hydrogen bonds are formed between O2 of uridine and N1 and N2 of guanosine, respectively.

The structural investigation of the VMC10 deoxyribozyme did not yet provide any structural information on how VMC10 differentiates non-methylated and methylated RNA substrates and what is the fold of the catalytic core and the cleavage site. The RNA substrates used in the co-crystallization with DNAzyme, dissociated from the DNA/RNA complex and formed independent structures with different structural motifs, like bulges, loops, or base triplets. It was shown that natural RNA folds or motifs have provided the basis for an important strain of RNA nanotechnology known as tectoRNA (Jaeger and Chworos 2006, Grabow and Jaeger 2014, Geary, Chworos et al. 2017). The branched kissing loops can be also used for the construction of various RNA homooligomeric nanostructures (Liu, Geary et al. 2020). The structural motifs from the VMC10 RNA substrates could also be used to create particular shapes/organize RNA molecules or to improve the 3D-RNA prediction software. Apart from the crystals formed only by the RNA, a crystallization hit for the VMC10 DNAzyme in complex with RNA substrate was obtained. The

presence of both DNA and RNA strands was confirmed by the HPLC analysis of the dissolved crystals. The D1966/R665 complex crystal, after the post-crystallization treatment, was diffracting to ~3.85 Å and it is a promising candidate for further investigation.

4.2 STRUCTURAL INVESTIGATION OF THE PRE-CATALYTIC STATE OF THE REACTION OF 9DB1

The 9DB1 deoxyribozyme catalyzes the formation of a native 3'-5'-phosphodiester bond between two RNA substrates in the presence of the divalent metal cations. The previously described crystal structure of the 9DB1 could not explain the role of magnesium cations as cofactors for accelerating RNA ligation. Here, the goal was to elucidate the three-dimensional structure of the 9DB1 DNAzyme in the pre-catalytic state of reaction to describe the ligation mechanism. A trimolecular complex with two reactive RNA fragments and the deoxyribozyme was crystallized. Surprisingly, the RNA substrates were ligated during the crystallization process. As a result, the 9DB1 deoxyribozyme formed a crystal in complex with its RNA product, thus representing a new structure of the post-catalytic state of the reaction. The structure of the T29C T7C 9DB1 mutant in complex with its RNA ligation product will be discussed.

The T29C T7C 9DB1 mutant with its product has an almost identical fold as parent 9DB1 (PDB:5CKK) as it can be superimposed with the root mean square deviation of 0.81 Å. The major difference between the two structures can be observed in the ligation junction. In the parent 9DB1 crystal structure, nucleotides dT29 and G1 are not co-planar and therefore their hydrogen bonding deviates from a Watson-Crick wobble base pair pattern. The mutation at position 29 from thymine to cytidine results in the base pair formation between the dC29 and the G1. The T29C mutation allowed to overcome the initially identified 9DB1 substrate requirements D|RA (D=A, G, T/U, R=A, or G) (Purtha, Coppins et al. 2005). This mutation that forms the base pair with a nucleotide at position 1 of the RNA substrate, results in the acceleration of the ligation reaction. Additionally, mutation at position 2 of the donor RNA was also examined. Originally, the 9DB1 was reported to require adenosine at position 2 of the donor RNA substrate (Purtha, Coppins et al. 2005). In the parent 9DB1 crystal structure, adenosine at position 2 is engaged in a Watson Crick base pair with the thymidine in position 7 in the DNA strand. It is hard to tell from the crystal structure, why the 9DB1 favors adenosine at position 2 since any other nucleotide at position 2 could be substituted to form a base pair with the opposite nucleotide at position 7 in the DNA strand. In the T29, T7C 9DB1 mutant crystal structure, the mutated thymidine to deoxycytidine at position 7 is engaged in a Watson-Crick base pair with guanosine at position 2 of the RNA donor RNA. This result tells us that a modified version of 9DB1 can accept any nucleotide at position 2, using

the original A|G ligation junction, even if the original 9DB1 was originally reported to accept only substrates with adenosine at position 2 (Purtha, Coppins et al. 2005).

The crystal structure of the T29, T7C 9DB1 supports the previously reported crystal structure of DNAzyme in the post-catalytic product-bound state of reaction and it was unable to explain the role of magnesium ions in the catalysis. It also confirms that phosphate oxygen of dA13 is crucial in activating the 3'-OH group to promote RNA ligation (Ponce-Salvatierra, Wawrzyniak-Turek et al. 2016).

Three years after the parent 9DB1 crystal structure was revealed, Aranda et al. have reported a comprehensive characterization of the reaction mechanism for 9DB1 DNAzyme using classical molecular dynamics (MD) combined with high-level hybrid quantum mechanics/molecular mechanics (QM/MM) free-energy calculations (Aranda, Terrazas et al. 2019). They were able to locate the Mg^{2+} binding sites, which are crucial for positioning the reactive groups and stabilizing the transition state. In the proposed mechanism, the O3' atom from A-1 attacks the α-phosphate group of the G1, and the leftover proton from the O3' atom is then transferred to the DNA (catalytic dA13 *pro*-Sp phosphate oxygen). The mechanism involves the active participation of two Mg^{2+} ions, not observed in the original crystal structure.

The other group, using density functional theory (DFT) calculations, have reported a two-step reaction mechanism for 9DB1, similar to the mechanism proposed by Aranda et al., which involves a phosphorane intermediate, where the 3'-hydroxyl group (nucleophile) of the acceptor RNA fragment attacks the 5'-triphosphate of the donor RNA fragment (Mattioli, Bottoni et al. 2019). However, in this mechanism, the presence of a divalent metal is not required for the activity of the 9DB1. In the parent 9DB1 and T29C, T7C mutant crystal structures, the magnesium cations play only a structural role, compensating for the negative charge of the phosphate backbone.

The two proposed reaction mechanisms described above are contradictory. One of them concludes that magnesium ions are crucial in the ligation reaction, while the second one concludes the opposite. Therefore, experimentally obtained structural models in the pre-catalytic and/or transition-state with bound metal ions and the pyrophosphate leaving group are necessary to properly describe the ligation mechanism of the 9DB1. My initial trials to obtain 9DB1 in complex with its RNA substrates resulted in the ligation of two RNA strands during the crystallization process and revealed the 9DB1 in a post-catalytic state. To capture the 9DB1 in the pre-catalytic state, I used two additional constructs with mutations in the RNA substrates to prevent ligation. The results indicate that 9DB1 in complex with two RNA substrates do not form any high-quality

crystals. The reason could be that the two DNA/RNA arms are too flexible to form a semi-continuous helix observed in the post-catalytic crystals. To overcome that problem it may be interesting to use Fab chaperone-assisted crystallography (Sherman, Archer et al. 2016). This could help to overcome the flexibility problem which is counterproductive to crystal packing. The Fabs were successfully used for the crystallization of the Spinach (Koirala, Shelke et al. 2018) and DIR2 (Shelke, Shao et al. 2018) aptamers.

4.3 Structural investigation of the Chili RNA aptamer

The Chili aptamer is a miniaturized version of the previously reported 13-2 aptamer and exhibits large Stoke shifts (over 130 nm). This suggests incorporating excited-state proton transfer (ESPT) in the activation mechanism as observed in Large Stoke shift fluorescent proteins such as LSSmOrange and LSSmKate. Here, the goal was to elucidate the three-dimensional structure of the Chili aptamer with its cognate ligands to answer the question of how Chili mimics LSS FP. Two structures of Chili in complex with $DMHBI^+$ and $DMHBO^+$ were solved and will be discussed.

The crystal structure of the Chili RNA has revealed that the fluorophore binding domain comprises a G-quadruplex. The binding domains of Spinach and Corn aptamers, the most structurally characterized fluorogenic aptamers so far, also comprises the G-quadruplex motif (Neubacher and Hennig 2019, Trachman and Ferre-D'Amare 2019). The only example of the fluorogenic RNA aptamer lacking a G-quadruplex motif is DIR2 (Shelke, Shao et al. 2018). DIR2 was selected to activate the fluorescence of cyanine dyes, dimethylindole red (DIR), and oxazole thiazole blue (OTB), allowing detection of two well-resolved emission colors. The G-quadruplex provides a large surface area for stacking with the ligand, stabilizing a rigid conformation and thereby minimizes non-radiative deactivation of the excited state. In Spinach and Corn, fluorophores are bound to the aptamer in the co-planar conformation, while in the Chili, ligands are bound in the non-planar conformation. The non-planar conformation is dictated by the distinctive π-π stacking interactions and by the hydrogen bonds formed between the oxime moiety and the RNA backbone. The Chili bound fluorophores $DMHBI^+$ ($\varphi = -35°$, $\tau = 25°$) and $DMHBO^+$ ($\varphi = -31°$, $\tau = 12°$) in its non-planar conformation resemble some fluorophores found in the large Stokes shift fluorescent proteins like LSSmKate1 ($\varphi = -21°$ and $\tau = 12°$; PDB: 3NT9) (Piatkevich, Malashkevich et al. 2010) or mCRISPRed ($\varphi = -32°$ and $\tau = 24°$, PDB: 6XWY) (Erdogan, Fabritius et al. 2020).

DFHBI and DFHO, the Spinach and Corn fluorophores, are bound to the aptamer in their deprotonated (phenolate) form and adopt a co-planar conformation to improve their fluorescence

(Figure 93a,b). In contrast to Spinach and Corn, Chili RNA aptamer binds and activates its fluorophores in the protonated (phenol) form (Steinmetzger, Palanisamy et al. 2019) and incorporates an excited-state proton transfer (ESPT) turn-on mechanism (Figure 93c). In the ESPT, the phenolic proton is extracted after photoexcitation of the bound ligand, and fluorescence is emitted from the phenolate form. The ESPT turn-on mechanism also results in large Stokes shifts in LSSmOrange and LSSmKate1 and LSSmKate2. In the LSSmOrange (Figure 93d) serine 146 (S146) forms two hydrogen bonds, one with the fluorophore and another with aspartic acid 161 (D161), and acts as a mediator for ESPT from the chromophore to the carboxyl group of aspartic acid 161 (Pletnev, Shcherbakova et al. 2014). In the LSSmKate1 (Figure 93e), a hydrogen bond between the chromophore hydroxyl and glutamic acid 160 (E160) allows ESPT, while in LSSmKate2 (Figure 93f), a proton transfer from chromophore through serine 158 (S158) to the aspartic acid 160 (D160) serves as a possible ESPT pathway (Piatkevich, Malashkevich et al. 2010).

In the crystal structure of the Chili RNA aptamer, the N7 of G15 forms a hydrogen bond with the hydroxy group of $DMHBI^+$ and $DMHBO^+$ and serves as a possible proton acceptor and directly controls the protonation state of the ligand in response to photoexcitation (Figure 94). This mechanism leads to large Stokes shifts (over 130 nm) observed in Chili and thus mimics LSS fluorescent proteins.

In summary, the structures of the Chili aptamer in complex with the green-emitting $DMHBI^+$ and the red-emitting $DMHBO^+$ ligands identified a new mechanism for ligand binding and fluorescence activation. The Spinach and Corn fluorophores are bound in their phenolate form in co-planar conformation to maximize their fluorescence. The Chili binds its ligands in the protonated phenol form and incorporates the ESPT turn-on mechanism. The hydrogen bond interaction between the hydroxyphenyl group and the N7 guanine nitrogen together with the coordination of a potassium ion establishes the Hoogsteen side of a G:C base pair on top of the guanine quartet as a new motif to facilitate excited-state proton transfer. Moreover, the crystal structures of Chili RNA aptamer further strengthen the view of G-quadruplexes flanked by canonical A-form duplexes as versatile and privileged architectures for fluorescence activation of HBI chromophores.

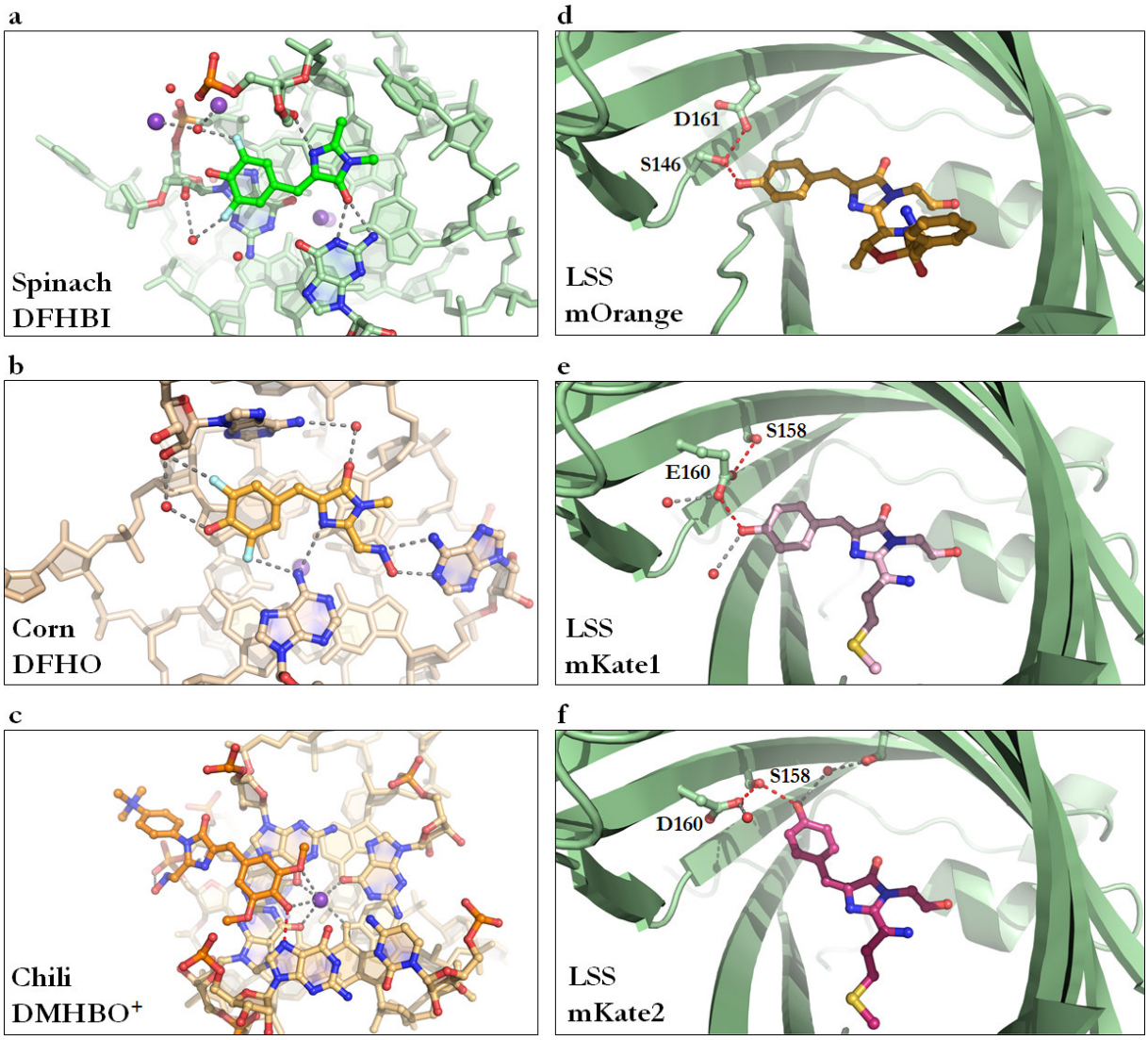


Figure 93. Comparison of fluorogenic aptamers and Large Stokes shift fluorescent proteins.

The Spinach and Corn aptamers bind their fluorophores in a deprotonated form in a co-planar conformation to maximize their fluorescence. The Chili binds its fluorophore in the protonated form and allows the incorporation of an excited-state proton transfer turn-on mechanism, observed in LSSmOrange LSSmKate1 and LSSmKate2 fluorescent proteins. The red dashed lines represent the ESPT pathway. The potassium and water molecules are represented as purple and red spheres respectively.

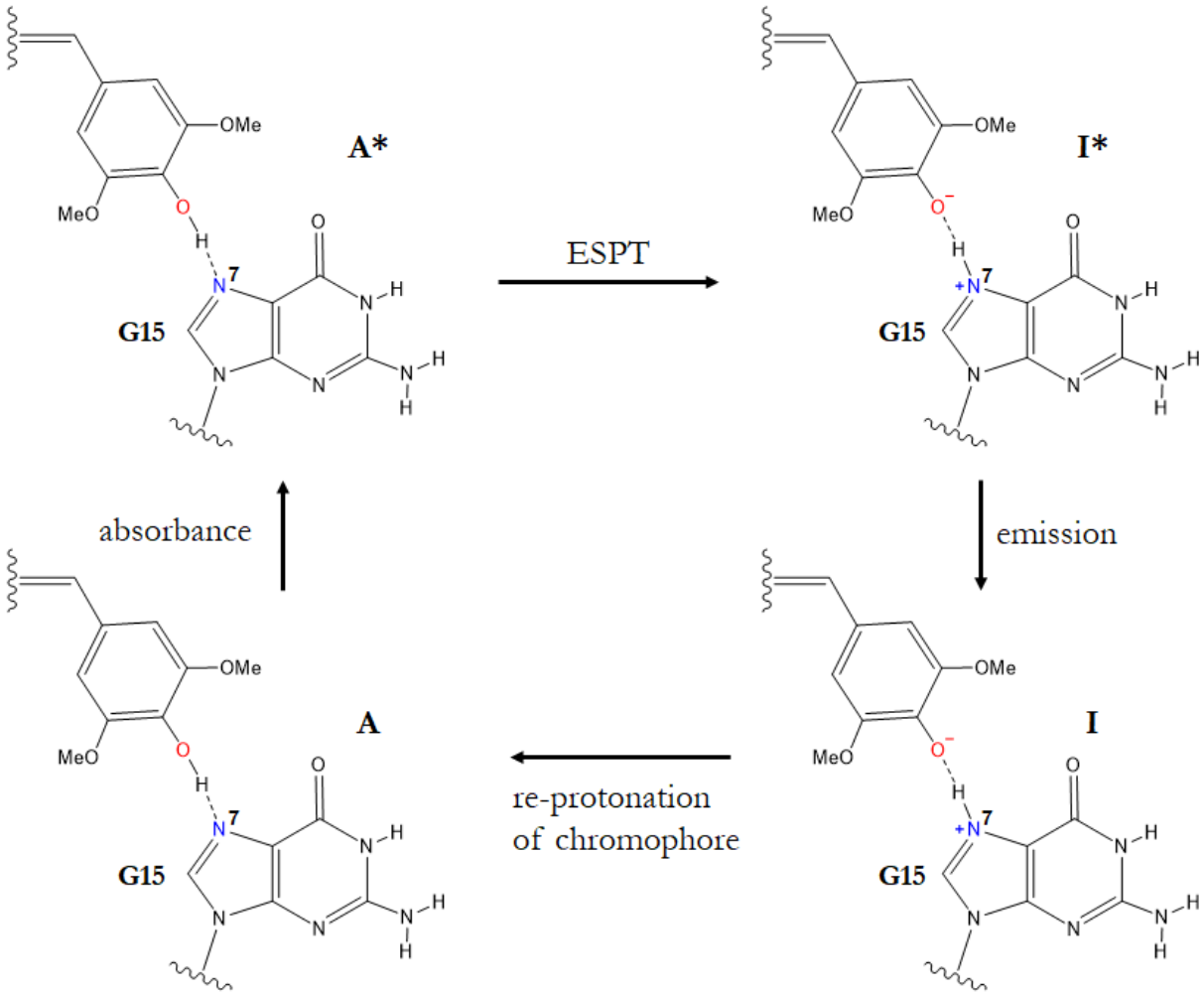


Figure 94. Proposed fluorescence turn-on mechanism of the Chili aptamer.

Chili RNA aptamer binds fluorophore in the protonated form. The fluorophore ground state (compound A) gets photo-excited (compound A*) and after ESPT is transformed into the intermediate excited state (compound I*). After fluorescence emission, the proton transfer from the N7 of guanosine to the fluorophore regenerates the ground state of the protonated fluorophore.

5 BIBLIOGRAPHY

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallogr D Biol Crystallogr **68**(Pt 4): 352-367.

Altona, C. and M. Sundaralingam (1972). "Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation." J Am Chem Soc **94**(23): 8205-8212.

Appaiahgari, M. B. and S. Vrati (2007). "DNAzyme-mediated inhibition of Japanese encephalitis virus replication in mouse brain." Mol Ther **15**(9): 1593-1599.

Aranda, J., M. Terrazas, H. Gómez, N. Villegas and M. Orozco (2019). "An artificial DNAzyme RNA ligase shows a reaction mechanism resembling that of cellular polymerases." Nature Catalysis **2**(6): 544-552.

Ashraf, S., L. Huang and D. M. J. Lilley (2019). "Effect of methylation of adenine N(6) on kink turn structure depends on location." RNA Biol **16**(10): 1377-1385.

Babendure, J. R., S. R. Adams and R. Y. Tsien (2003). "Aptamers switch on fluorescence of triphenylmethane dyes." J Am Chem Soc **125**(48): 14716-14717.

Bang, I. (1910). "Untersuchungen über die Guanylsäure." Biochem. Z. **26 (in German)**: 293-311.

Basu, S., B. Sriram, R. Goila and A. C. Banerjea (2000). "Targeted cleavage of HIV-1 coreceptor-CXCR-4 by RNA-cleaving DNA-enzyme: inhibition of coreceptor function." Antiviral Res **46**(2): 125-134.

Batey, R. T. and J. S. Kieft (2016). "Soaking Hexammine Cations into RNA Crystals to Obtain Derivatives for Phasing Diffraction Data." Methods Mol Biol **1320**: 219-232.

Bing, T., X. Yang, H. Mei, Z. Cao and D. Shangguan (2010). "Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories." Bioorganic & Medicinal Chemistry **18**(5): 1798-1805.

Bouhedda, F., K. T. Fam, M. Collot, A. Autour, S. Marzi, A. Klymchenko and M. Ryckelynck (2020). "A dimerization-based fluorogenic dye-aptamer module for RNA imaging in live cells." Nat Chem Biol **16**(1): 69-76.

Breaker, R. R. (2012). "Riboswitches and the RNA world." Cold Spring Harb Perspect Biol **4**(2).

Breaker, R. R. and G. F. Joyce (1994). "A DNA enzyme that cleaves RNA." Chemistry & Biology **1**(4): 223-229.

Breaker, R. R. and G. F. Joyce (2014). "The expanding view of RNA and DNA function." Chem Biol **21**(9): 1059-1065.

Brunel, C., R. Marquet, P. Romby and C. Ehresmann (2002). "RNA loop–loop interactions as dynamic functional motifs." Biochimie **84**(9): 925-944.

Buchhaupt, M., C. Peifer and K. D. Entian (2007). "Analysis of 2'-O-methylated nucleosides and pseudouridines in ribosomal RNAs using DNAzymes." Anal Biochem **361**(1): 102-108.

Bujnowska, M., J. Zhang, Q. Dai, E. M. Heideman and J. Fei (2020). "Deoxyribozyme-based method for absolute quantification of N (6)-methyladenosine fractions at specific sites of RNA." J Biol Chem **295**(20): 6992-7000.

Burge, S., G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle (2006). "Quadruplex DNA: sequence, topology and structure." Nucleic Acids Research **34**(19): 5402-5415.

Burmeister, J., G. von Kiedrowski and A. D. Ellington (1997). "Cofactor-Assisted Self-Cleavage in DNA Libraries with a 3'–5'-Phosphoramidate Bond." Angewandte Chemie International Edition in English **36**(12): 1321-1324.

Butcher, S. E., T. Dieckmann and J. Feigon (1997). "Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop." J Mol Biol **268**(2): 348-358.

Byrne, R. T., A. L. Konevega, M. V. Rodnina and A. A. Antson (2010). "The crystal structure of unmodified tRNA Phe from Escherichia coli." Nucleic Acids Research **38**(12): 4154-4162.

Cabili, M. N., M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, J. L. Rinn and A. Raj (2015). "Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution." Genome Biology **16**(1): 20.

Cang, X., J. Šponer and T. E. Cheatham, 3rd (2011). "Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis." J Am Chem Soc **133**(36): 14270-14279.

Carlsson, C., A. Larsson, M. Jonsson, B. Albinsson and B. Norden (1994). "Optical and Photophysical Properties of the Oxazole Yellow DNA Probes YO and YOYO." The Journal of Physical Chemistry **98**(40): 10313-10321.

Chandra, M. and S. K. Silverman (2008). "DNA and RNA can be equally efficient catalysts for carbon-carbon bond formation." J Am Chem Soc **130**(10): 2936-2937.

Chen, F., Z. Li, R. Wang, B. Liu, Z. Zeng, H. Zhang and J. Zhang (2004). "Inhibition of ampicillin-resistant bacteria by novel mono-DNAzymes and di-DNAzyme targeted to beta-lactamase mRNA." Oligonucleotides **14**(2): 80-89.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Chen, X., N. Li and A. D. Ellington (2007). "Ribozyme catalysis of metabolism in the RNA world." Chem Biodivers **4**(4): 633-655.

Chen, X., D. Zhang, N. Su, B. Bao, X. Xie, F. Zuo, L. Yang, H. Wang, L. Jiang, Q. Lin, M. Fang, N. Li, X. Hua, Z. Chen, C. Bao, J. Xu, W. Du, L. Zhang, Y. Zhao, L. Zhu, J. Loscalzo and Y. Yang (2019). "Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs." Nat Biotechnol **37**(11): 1287-1293.

Cheong, C. C., H.-K. (2010). RNA Structure: Tetraloops. eLS.

Chinnapen, D. J.-F. and D. Sen (2004). "A deoxyribozyme that harnesses light to repair thymine dimers in DNA." Proceedings of the National Academy of Sciences **101**(1): 65-69.

Coker, H., G. Wei and N. Brockdorff (2019). "m6A modification of non-coding RNA and the control of mammalian gene expression." Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1862**(3): 310-318.

Coppins, R. L. and S. K. Silverman (2004). "A DNA enzyme that mimics the first step of RNA splicing." Nat Struct Mol Biol **11**(3): 270-274.

Coppins, R. L. and S. K. Silverman (2005). "A deoxyribozyme that forms a three-helix-junction complex with its RNA substrates and has general RNA branch-forming activity." J Am Chem Soc **127**(9): 2900-2907.

Cruz, R. P., J. B. Withers and Y. Li (2004). "Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme." Chem Biol **11**(1): 57-67.

Cuenoud, B. and J. W. Szostak (1995). "A DNA metalloenzyme with DNA ligase activity." Nature **375**(6532): 611-614.

Dash, B. C., T. A. Harikrishnan, R. Goila, S. Shahi, H. Unwalla, S. Husain and A. C. Banerjea (1998). "Targeted cleavage of HIV-1 envelope gene by a DNA enzyme and inhibition of HIV-1 envelope-CD4 mediated cell fusion." FEBS Lett **431**(3): 395-399.

Day, N. R. D., M.W. (2014). The fluorescent protein revolution, Boca Raton, Florida : CRC Press/Taylor and Francis.

Day, R. N. and M. W. Davidson (2009). "The fluorescent protein palette: tools for cellular imaging." Chemical Society Reviews **38**(10): 2887-2921.

DeJong, E. S., W. F. Marzluff and E. P. Nikonowicz (2002). "NMR structure and dynamics of the RNA-binding site for the histone mRNA stem-loop binding protein." RNA **8**(1): 83-96.

Dolgosheina, E. V., S. C. Jeng, S. S. Panchapakesan, R. Cojocaru, P. S. Chen, P. D. Wilson, N. Hawkins, P. A. Wiggins and P. J. Unrau (2014). "RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking." ACS Chem Biol **9**(10): 2412-2420.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek and G. Rechavi (2012). "Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq." Nature **485**(7397): 201-206.

Ellington, A. D. and J. W. Szostak (1990). "In vitro selection of RNA molecules that bind specific ligands." Nature **346**(6287): 818-822.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Ennifar, E., A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov and P. Dumas (2000). "The crystal structure of UUCG tetraloop." J Mol Biol **304**(1): 35-42.

Ennifar, E., P. Walter, B. Ehresmann, C. Ehresmann and P. Dumas (2001). "Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site." Nat Struct Biol **8**(12): 1064-1068.

Erdogan, M., A. Fabritius, J. Basquin and O. Griesbeck (2020). "Targeted In Situ Protein Diversification and Intra-organelle Validation in Mammalian Cells." Cell Chemical Biology **27**(5): 610-621.e615.

Evans, P. R. (2011). "An introduction to data reduction: space-group determination, scaling and intensity statistics." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 282-292.

Famulok, M. and G. Mayer (2014). "Aptamers and SELEX in Chemistry & Biology." Chemistry & Biology **21**(9): 1055-1058.

Faulhammer, D. and M. Famulok (1997). "Characterization and divalent metal-ion dependence of in vitro selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides." J Mol Biol **269**(2): 188-202.

Filonov, G. S., J. D. Moon, N. Svensen and S. R. Jaffrey (2014). "Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution." J Am Chem Soc **136**(46): 16299-16308.

Fokina, A. A., D. A. Stetsenko and J. C. François (2015). "DNA enzymes as potential therapeutics: towards clinical application of 10-23 DNAzymes." Expert Opin Biol Ther **15**(5): 689-711.

Fyfe, A. C., P. W. Dunten, M. M. Martick and W. G. Scott (2015). "Structural Variations and Solvent Structure of r(UGGGGU) Quadruplexes Stabilized by Sr(2+) Ions." Journal of molecular biology **427**(12): 2205-2219.

Garman, E. F. (2010). "Radiation damage in macromolecular crystallography: what is it and why should we care?" Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 339-351.

Geary, C., A. Chworos, E. Verzemnieks, N. R. Voss and L. Jaeger (2017). "Composing RNA Nanostructures from a Syntax of RNA Structural Modules." Nano Letters **17**(11): 7095-7101.

Gellert, M., M. N. Lipsett and D. R. Davies (1962). "Helix formation by guanylic acid." Proc Natl Acad Sci U S A **48**: 2013-2018.

Geyer, C. R. and D. Sen (1997). "Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme." Chem Biol **4**(8): 579-593.

Gill, M. L., S. A. Strobel and J. P. Loria (2005). "205Tl NMR methods for the characterization of monovalent cation binding to nucleic acids." J Am Chem Soc **127**(47): 16723-16732.

Gill, M. L., S. A. Strobel and J. P. Loria (2006). "Crystallization and characterization of the thallium form of the Oxytricha nova G-quadruplex." Nucleic Acids Res **34**(16): 4506-4514.

Goila, R. and A. C. Banerjea (1998). "Sequence specific cleavage of the HIV-1 coreceptor CCR5 gene by a hammer-head ribozyme and a DNA-enzyme: inhibition of the coreceptor function by DNA-enzyme." FEBS Lett **436**(2): 233-238.

Goila, R. and A. C. Banerjea (2001). "Inhibition of hepatitis B virus X gene expression by novel DNA enzymes." Biochem J **353**(Pt 3): 701-708.

Gong, L., Z. Zhao, Y.-F. Lv, S.-Y. Huan, T. Fu, X.-B. Zhang, G.-L. Shen and R.-Q. Yu (2015). "DNAzyme-based biosensors and nanodevices." Chemical Communications **51**(6): 979-995.

Grabow, W. W. and L. Jaeger (2014). "RNA Self-Assembly and RNA Nanotechnology." Accounts of Chemical Research **47**(6): 1871-1880.

Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace and S. Altman (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3 Pt 2): 849-857.

Han, K. Y., B. J. Leslie, J. Fei, J. Zhang and T. Ha (2013). "Understanding the Photophysics of the Spinach–DFHBI RNA Aptamer–Fluorogen Complex To Improve Live-Cell RNA Imaging." Journal of the American Chemical Society **135**(50): 19033-19038.

Hardin, C. C., A. G. Perry and K. White (2000). "Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids." Biopolymers **56**(3): 147-194.

Hoadley, K. A., W. E. Purtha, A. C. Wolf, A. Flynn-Charlebois and S. K. Silverman (2005). "Zn2+-dependent deoxyribozymes that form natural and unnatural RNA linkages." Biochemistry **44**(25): 9217-9231.

Höbartner, C. (2019). "How DNA catalyses RNA ligation." Nature Catalysis **2**(6): 483-484.

Höbartner, C., P. I. Pradeepkumar and S. K. Silverman (2007). "Site-selective depurination by a periodate-dependent deoxyribozyme." Chemical Communications(22): 2255-2257.

Hollenstein, M. (2015). "DNA Catalysis: The Chemical Repertoire of DNAzymes." Molecules **20**(11): 20777-20804.

Hou, W., Q. Ni, J. Wo, M. Li, K. Liu, L. Chen, Z. Hu, R. Liu and M. Hu (2006). "Inhibition of hepatitis B virus X gene expression by 10-23 DNAzymes." Antiviral Res **72**(3): 190-196.

Hou, W., J. E. Wo, M. W. Li and K. Z. Liu (2005). "In vitro cleavage of hepatitis B virus C mRNA by 10-23 DNA enzyme." Hepatobiliary Pancreat Dis Int **4**(4): 573-576.

Hou, Z., J. R. Meng, C. Niu, H. F. Wang, J. Liu, B. Q. Hu, M. Jia and X. X. Luo (2007). "Restoration of antibiotic susceptibility in methicillin-resistant Staphylococcus aureus by targeting mecR1 with a phosphorothioate deoxyribozyme." Clin Exp Pharmacol Physiol **34**(11): 1160-1164.

Hou, Z., J. R. Meng, J. R. Zhao, B. Q. Hu, J. Liu, X. J. Yan, M. Jia and X. X. Luo (2007). "Inhibition of beta-lactamase-mediated oxacillin resistance in Staphylococcus aureus by a deoxyribozyme." Acta Pharmacol Sin **28**(11): 1775-1782.

Huang, H., N. B. Suslov, N. S. Li, S. A. Shelke, M. E. Evans, Y. Koldobskaya, P. A. Rice and J. A. Piccirilli (2014). "A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore." Nat Chem Biol **10**(8): 686-691.

Huang, L., S. Ashraf, J. Wang and D. M. Lilley (2017). "Control of box C/D snoRNP assembly by N(6)-methylation of adenine." EMBO Rep **18**(9): 1631-1645.

Hutchinson, J. N., A. W. Ensminger, C. M. Clemson, C. R. Lynch, J. B. Lawrence and A. Chess (2007). "A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains." BMC Genomics **8**: 39.