# Non-canonical nucleic acid structures induced by specific binding of metal ions

1. *Name*

**Witold Jacek Andrałojć**

2. *Diplomas, degrees conferred in specific areas of science or arts, including the name of the institution which conferred the degree, year of degree conferment, title of the PhD dissertation.*

2016 **PhD in structural biology**

International doctorate program at European Magnetic Resonance Center (CERM – *wł.* „Centro Europeo di Risonanze Magnetiche"), (https://www.cerm.unifi.it/training/international-doctorate)
(within the scope of a Marie Curie Actions ITN pNMR)
University of Florence, Italy
Supervisor: prof. Giacomo Parigi
Thesis title: „Conformational heterogeneity in multidomain biological systems studied through averaged NMR restraints"

2013 **Master thesis in chemistry and material sciences**

Interantional Master program in physical chemistry SERP-chem (www.serp-chem.eu) co-organized by:
Université Paris-Sud, Paris, France
Adam Mickiewicz University, Poznań, Poland
Università Degli Studi di Genova, Genova, Italy
Supervisor: prof. Catalin Miron
Thesis title: „On the nature of Rotational Doppler Broadening in photoelectron spectroscopy – an experimental study"

3. *Information on employment in research institutes or faculties/departments or school of arts.*

2024 Institute of Bioorganic Chemistry Polish Academy of Sciences:
Head of Department of Biomolecular NMR

2018 Institute of Bioorganic Chemistry Polish Academy of Sciences: Assistant professor

2017 -2018 Institute of Bioorganic Chemistry Polish Academy of Sciences: Research assistant

2013 -2016 CERM („Centro Europeo di Risonanze Magnetiche"), University of Florence, PhD student

4. *Description of the achievements, set out in art. 219 para 1 point 2 of the Act.*

## List of publications constituting the scientific achievement

**[H1]** **W. Andrałojć**, M. Małgowska, J. Sarzyńska, K. Pasternak, K. Szpotkowski, R. Kierzek , Z. Gdaniec*, *Unraveling the structural basis for the exceptional stability of RNA G-quadruplexes capped by a uridine tetrad at the 3′ terminus*, RNA, 2019, 25, 131-134

**[H2]** **W. Andrałojć***, K. Pasternak, J. Sarzyńska, K. Zielińska, R. Kierzek, Z. Gdaniec*, *The origin of the high stability of 3′-terminal uridine tetrads: contributions of hydrogen bonding, stacking interactions, and steric factors evaluated using modified oligonucleotide analog*s, RNA 2020, 26, 2000-2016

**[H3]** **W. Andrałojć***, J. Wieruszewska , K. Pasternak, Z. Gdaniec, *Solution Structure of a Lanthanide-binding DNA Aptamer Determined Using High Quality pseudocontact shift restraints*, Chem. Eur. J., 2022, 28, e202202114

**[H4]** J. Wieruszewska, A. Pawłowicz, E. Połomska, K. Pasternak, Z. Gdaniec, **W. Andrałojć***, *The 8-17 DNAzyme can operate in a single active structure regardless of metal ion cofactor*, Nat. Commun., 2024, 15, 4218

## Introduction

My main research goal is to improve our understanding of the proficiency of nucleic acids – DNA and RNA – for strong and site-specific binding of metal ions, as well as of the potential of such interactions to shape the structures and activities of these biomolecules. Within the scientific achievement presented in this habilitation I have determined the high-resolution structures of three nucleic acid molecules and characterized the metal-binding sites that they form. The studied systems include:

1) A four-stranded RNA G-quadruplex $r(UGGUGGU)_4$ stabilized by potassium $K^+$ ions (*papers* **H1** and **H2**)

2) A lanthanide-binding DNA aptamer in complex with its molecular target
(*paper* **H3**)

3) The catalytically proficient structure of the 8-17 DNAzyme, induced by zinc $Zn^{2+}$ binding
(*paper* **H4**)

The following dissertation is divided into three chapters. In the first one I describe the physical properties of typical metal-binding sites present in nucleic acids, as well as the structural biology methods available for their experimental characterization, concentrating on NMR spectroscopy that I employ in my research work. This chapter also serves to situate the presented experimental papers **H1-H4** within the broader context of the study of nucleic acid interactions with metal ions, preceding an in-depth discussion of each of them in Chapter 2. In turn, Chapter 3 provides a summary of similarities and differences between the studied systems, followed by a reflection how these properties influenced the research methods I applied and, in consequence, the results I obtained. It also includes an attempt at summarizing the added value brought by the presented publications into the research field concerned.

# Chapter 1

# Characteristics, importance, and methods of studying nucleic acid interactions with metal ions

## 1.1 Properties of nucleic acid interactions with metal ions and their role in shaping the structures of these macromolecules

As strongly negatively charged biopolymers – owing to the presence of phosphate groups within their chemical structures – nucleic acids in solution are always surrounded by an ensemble of neutralizing positive ions, often simple metal cations $M^{n+}$.[1] However, a major part of these counter-ions interacts with DNA/RNA in a non-specific manner exclusively through electrostatic attraction and remaining in constant motion forms around the biomolecule a, so called, ion atmosphere.[2,3] Even this type of interactions already has an important influence on the conformational preferences of nucleic acid molecules, stabilizing their compact spatial structures by shielding the electrostatic repulsion between different phosphate groups.[2,3] Less numerous, as usually responsible for around 10 % of the charge neutralizing an RNA molecule,[1,2] yet even more important for its structural preferences, are metal ions binding to it in a site-specific manner, that is interacting with selected elements within its three-dimensional structure. Such interactions may involve direct coordination of the metal ion by functional groups belonging to the nucleic acid, however the amount of such direct contacts is surprisingly low. In most cases the ion remains largely hydrated forming mainly outer-sphere interactions (hydrogen bonds between metal-bound water molecules and DNA/RNA ligands) with still

an important role of the electrostatic attraction in the interactions. This behavior is well illustrated by a statistical analysis of magnesium ion $Mg^{2+}$ (with preferred coordination number N=6) binding sites observed in crystallographic structures of RNA.[4] Among over ten thousand specific interaction sites analyzed around a third did not contain any direct $Mg^{2+}$ contacts with RNA functional groups, over a half included one or two such contacts and only below 10 % of sites featured three or four direct interactions. This situation remain is stark contrast to the picture observed for metal ions interacting with protein enzymes, for which protein ligands routinely fill the majority of the first coordination sphere. This difference is also reflected in incomparably lower binding constants for metal ion interactions with nucleic acids, as compared to proteins.[5] Already for metal ions tolerated by living cells in comparably high concentrations, such as $Ca^{2+}$, it is normal for proteins to display dissociation constants ($K_d$) in the nanomolar range,[6] while for those ions whose metabolism is tightly regulated within the cell protein systems can achieve attomolar affinities,[7] comparable to those of small molecule chelates such as EDTA. In contrast, a large portion of biologically significant $Mg^{2+}$ binding sites display millimolar $K_d$ values,[8–11] with less abundant examples of sites for which the interaction is shifted towards the micromolar range ($K_d \approx 10^{-5}$-$10^{-6}$ M).[12] Moreover, there is data suggesting that these values are already approaching the physical limit achievable for affinities in metal ion interactions with nucleic acids. Namely numerous SELEX experiments seeking nucleic acid structural motifs most capable of forming strong interactions with various metal ions including $Zn^{2+}$, $Ni^{2+}$, $Cu^{2+}$ and $Hg^{2+}$ have always yielded systems that bind the target ion with $K_d$ values not surpassing $10^{-6}$ M.[13–15] Next to the lower free enthalpies of binding, the second common characteristic of metal binding sites formed by nucleic acid molecules is their limited selectivity towards the type of metal ion accepted. This property may be related to the abovementioned increased importance of electrostatic and hydrogen bond interactions with respect to direct coordination for many DNA/RNA-metal complexes. This situation also pertains to many binding sites of structural and functional importance. An example is provided by numerous divalent metal binding sites present in biologically important ribozymes – that is RNA molecules capable of self-hydrolysis in well-defined points of their nucleotide sequences. Metal binding to these sites is mandatory for the ribozyme to fold into compact, catalytically active structures. Direct metal involvement in the catalytic process is also possible, although for many catalytic RNAs a consensus was not yet reached regarding this issue.[5,16,17] In nature ribozyme molecules utilize magnesium $Mg^{2+}$ ions as cofactors, however numerous *in vitro* studies have shown that these molecules can also interact with a wide variety of other di- and trivalent ion such as $Ca^{2+}$, $Sr^{2+}$, $Mn^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Zn^{2+}$, $Cd^{2+}$, $Pb^{2+}$, $Co^{3+}$ or lanthanides $Ln^{3+}$.[1,5,16] In many cases ribozymes interacting with these alternative cofactors remain catalytically proficient and sometimes their activity is even improved.[1] Available crystallographic structures demonstrate that different metal ions can interact with the same binding site assuming somewhat distinct geometries,[1] differentiated by for example the number of direct RNA-metal contacts dependent on factors such as ionic radius or its hydration energy. The capability to promote RNA hydrolysis in the presence of a wide variety of

different metal ions is also a characteristic of many among the ribozyme's DNA counterparts, known as DNAzymes. However, in the case of DNAzymes the amount of available high-resolution structural information is very limited, as only a handful of spatial structures of such systems are available.[18–20] One among them is the three-dimensional structure of the 8-17 DNAzyme induced by zinc $Zn^{2+}$ ion binding, determined by me, and described in paper **H4**.

Another example of limited selectivity of metal-binding sites formed by nucleic acid systems is provided by the structures of DNA and RNA G-quadruplexes. Their architectures, described in more detail in Chapter 2.1, always include a central channel, inside which metal ions bind through direct coordination to up to eight carbonyl groups belonging to guanosine nucleobases. Interactions inside G-quadruplex channels constitute a rare example of metal ion binding occurring with their full dehydration. Despite this the discussed binding sites do not display a significant selectivity and can accept a broad range of mono- and divalent ions such as $K^+$, $Na^+$, $Cs^+$, $Ba^{2+}$, $Sr^{2+}$, or $Pb^{2+}$.[21] As a part of the presented habilitation achievement – papers **H1** and **H2** – I have determined the three-dimensional structure of a four-stranded RNA G-quadruplex formed through $K^+$ binding and demonstrated that the same G-quadruplex structure forms also in the presence of $Na^+$ ions.

Despite the above-described generally low selectivity of metal binding sites in nucleic acid molecules, a series of relatively rarer metal-selective or even metal-specific sites was also identified over the years. The most well-known examples of such sites are metal-stabilized pyrimidine-pyrimidine base pairs. A thymidine-thymidine mismatch (T:T) is capable of interacting with a $Hg^{2+}$ ion, which leads to the formation of a T-Hg-T base pair with stability comparable to that of canonical Watson-Crick pairings.[22] This interaction is based upon direct coordination of the mercury ion by N3 nitrogen atoms of both thymidine bases. Similarly, the cytidine-cytidine mismatch (C:C) forms analogous interactions, this time with a silver ion $Ag^+$.[23] DNA molecules also form a relatively large set of binding sites potentially selective towards lanthanide ions $Ln^{3+}$. Such conclusion can be drawn based on the identification of a series of DNAzyme systems catalytically active only in the presence of $Ln^{3+}$ ions as cofactors.[24] However, these molecules were never deeply structurally characterized and the lack of observable catalytic activity in the presence of other ions does not automatically signify the lack of interactions. Examples of an opposite are known in the literature, like the fact that $Ln^{3+}$ ions bind to the hammerhead ribozyme, usually activated by various divalent ions, yet the complex forming is catalytically inactive.[25] On the other hand, a confirmed example of selective lanthanide ion binding by DNA is provided by an aptamer discovered in 2016.[26] Lanthanide ion binding to this system is accompanied by a large-scale change in the tertiary DNA architecture, observed in the original paper through fluorescence-based methods. A similar conformational transition was not observed when the aptamer was titrated with a wide array of other metal ions, which allowed to conclude that the binding site it contains is indeed selective towards lanthanides. However, in the absence of high-resolution structural data, the source of this selectivity remained unclear. Uncovering the molecular basis for

selective lanthanide binding by this aptamer was one of the motivations that led me to determining its three-dimensional structure as described in paper **H3**.

## 1.2 Structural biology methods for the study of nucleic acid interactions with metal ions

### *Types of available high-resolution methods*

To obtain precise information regarding the location of metal binding sites in DNA/RNA structures, as well as regarding the geometric details of these interactions it is necessary to apply one of the high-resolution structural biology methods: X-ray crystallography, cryo electron microscopy (cryo-EM) or nuclear magnetic resonance (NMR) spectroscopy. In case of binding sites possessing a well-defined interaction geometry the first two methods proved the richest sources of information, with a leading role of X-ray crystallography due to both much longer application period of this method, as well as generally higher resolution of obtained structures, which facilitates precise modeling of metal ions within the biomolecule. Crystallographic structures of sufficient resolutions allow not only for unambiguous identification of atoms directly coordinating the metal ion, but also provide precise information regarding the geometry of the formed complex, including metal-ligand bond lengths etc. For hydrated ions, directly bound water molecules can also sometimes be located, given that they assume well-defined positions with respect to the biomolecule, thanks to for example hydrogen bond formation. Among the limitations of crystallographic methods in this context one should list the inability to observe binding sites with less defined interaction geometries (even if these interactions are strong and specific), the lack of information regarding the affinity of interactions with different sites (necessary for example to distinguish the strong micromolar site from numerous sites with millimolar $K_d$ values) and the potential influence of crystal packing forces on both the DNA/RNA structures and their interactions with metal ions. In addition, the fact that crystallization of more complex nucleic acid systems is often very difficult is also significant.

### *NMR spectroscopy in DNA/RNA-metal interaction studies*

The way in which the third high-resolution structural biology method – NMR spectroscopy, that is the main tool in my research work – can be applied in the context of the study of metal ion interaction sites on nucleic acids differs significantly from the previously described two. On one hand this technique is usually unable to provide similarly precise information, as a large portion of biologically important ions cannot be directly observed in NMR spectra (their nuclei are either magnetically inactive or very insensitive). Their presence is thus inferred indirectly, through the effect it has on NMR parameters of other atomic nuclei. On the other hand, however, NMR spectroscopy ideally complements the limitations of crystallographic techniques listed above, as it:

1) studies DNA/RNA-metal complexes directly in solutions,

2) is sensitive to the presence of interactions regardless of their degree of spatial definition,

3) allows to directly link the presence of a metal ion to the structural changes occurring in the biomolecule,

4) allows for the estimation of $K_d$ values for each of the observed interaction sites

The multitude of NMR spectral parameters and techniques available for the study of nucleic acid metal ion interactions, on very different accuracy levels, requires a systematic summarization here.

***NMR structure determination for nucleic acids – classical methods (NOE and J-coupling restraints)***

In the case of systems for which the formation of a specific three-dimensional structure is itself induced by metal binding – as is the case for all nucleic acid molecules studied in papers **H1-H4** – the first key element in the study of this interaction is the characterization of the conformation of the DNA/RNA component of the observed complex. The set of NMR methods available for this task is to a large degree similar to the NMR toolset used to study nucleic acids structures in general,[27] however when the bound metal ion displays paramagnetic properties its presence can be used to obtain additional structural information.[28] The first step in NMR structural analysis always consists of the assignment of resonances observed in the NMR spectra to specific atoms in the studied biomolecule, using a set of correlation spectra.[27] This analysis also yields a clear confirmation whether the studied biomolecule assumes a single stable structural form in solution or does it feature a conformational equilibrium between several slowly exchanging species. For nucleic acid complexes with metal ions already on this early stage one can ascertain whether ion binding significantly affects the structure of the DNA/RNA component. Appearance of a series of new imino proton NMR resonances during titration with the appropriate metal ion, signifies that metal binding induces the formation of new secondary structure elements and/or the reorganization of previously existing ones into a new tertiary architecture (situation encountered in paper **H3**). One the other hand, if a similar titration leads to a disappearance of a series of spectral resonances accompanied by strengthening and sometimes sharpening of the remaining ones one can conclude that the presence of the metal affects the conformational equilibrium present in the system by stabilizing the structural forms capable of metal binding (see paper **H4**). This kind of experiment provides an early confirmation that the DNA/RNA structural form to be studied is indeed the complex with given metal ion.

To determine the high-resolution spatial structure of the given DNA/RNA molecule one has to extract from the NMR spectra the broadest possible set of spectral information directly interpretable in terms of structural parameters, the so called “NMR structural restraints”. Subsequently using

molecular dynamics simulations, restrained with the obtained set of structural restraints, one attempts to identify a spatial structure consistent with all available NMR spectral data. Such simulations are then repeated multiple times in order to verify how well the obtained restraints define different aspects of the molecular structure. The quality (resolution) of the obtained structure directly depends on the number and quality of the structural restraints applied. Classically the most important type of NMR restraints are distance restraints, derived from the observation of Nuclear Overhauser Effect (NOE) between pairs of hydrogen atoms. This effect allows for a semi-quantitative estimation of internuclear distances of up to 5-6 Å, thus these restraints have a rather limited range. NOE structural information is usually complemented by dihedral angle restraints between atoms three chemical bonds apart, derived from the values of NMR scalar couplings ("J-couplings"). This type of restraints provides information about for example the conformations assumed by the sugar-phosphate backbone in nucleic acids, including the values of the sugar pucker in each nucleotide unit. These two types of structural restraints remain also today the most used information sources in the process of NMR structure determination, including the structures determined by me in papers **H1**, **H3** and **H4**.

***NMR structure determination for nucleic acids – long-range restraints (RDC, PCS, PRE)***

Given the local nature of classical NMR restraints, some global features of more complex nucleic acid molecules, such as exact values of interhelical angles or the degree of helix bending, are often determined in NMR structures with a lower degree of accuracy.[29] These limitations can be eliminated through the application of newer types of NMR restraints which provide more global structural information. However, due to experimental difficulties related to their observation, these restraint types still remain rarely used for nucleic acid systems. Most commonly applied among them are residual dipolar couplings (RDC) which provide information about the orientations of vectors connecting pairs of atoms in the reference frame of the entire biomolecule.[28,30] However, RDC measurement requires the studied biomolecule to become partially oriented with respect to the external magnetic field, which is most commonly achieved through the use of so-called alignment media (containing for example lipid bicelles or entire bacteriophage nucleocapsids).[30] Isotope labelling with $^{13}C$ and/or $^{15}N$ is also often required. An alternative method to induce RDCs consists of self-orientation of the biomolecule, by the introduction of unpaired electrons with non-zero magnetic susceptibility anisotropy $\Delta\chi$ into its structure.[28] A great advantage of this method resides in the fact that the presence of unpaired electrons in the biomolecule also allows for the measurement of two additional types of long-range NMR restraints:

1) pseudo-contact shift (PCS) which also requires the unpaired electrons to possess $\Delta\chi \neq 0$

2) paramagnetic relaxation enhancement (PRE) which requires only the presence of unpaired electrons, regardless of their $\Delta\chi$ values

Both these effects are dependent on the position assumed by the unpaired electron with respect to each atom within the studied biomolecule and are observable in the NMR spectra as resonance line shifting and broadening, respectively. All these restraint types allow for more precise and accurate determination of structures of more complex biomolecules and can also be used to study their dynamic properties and intermolecular interactions. These effects, known under the common name of „paramagnetic NMR restraints", have already demonstrated their great value in structural biology of protein systems,[28] however due to technical difficulties their application to nucleic acid molecules remains very limited. Unpaired electrons with $\Delta\chi \neq 0$ can be introduced into biomolecules only using some among paramagnetic metal ions – most notably lanthanide ions $Ln^{3+}$ and selected d-block ions, such as $Co^{2+}$. The observed $\Delta\chi$ values are also crucially dependent on the first coordination sphere of the given ion in complex with biomolecule, as well as on how well-defined its position is in the reference frame of the biomolecule (the ion has to be rigidly anchored to the host biomolecule). For protein systems significant efforts were undertaken to develop a wide spectrum of methods to introduce such metal binding sites into their structures, through the development of so-called paramagnetic tags.[31] Many among these methods rely on the covalent attachment of organic metal ion complexes to the protein, while others are based on mutations to the protein aminoacid sequence, that introduce a metal binding site directly into its primary structure. Unfortunately, despite continued attempts,[32,33] practically useful counterparts of these methods are not yet available for nucleic acid systems. Only paramagnetic tags based on stable organic radicals with $\Delta\chi = 0$ remain available, that produce only the PRE effect.[34] An important motivation behind my decision to study the structure of the lanthanide binding aptamer in paper **H3**, was to determine whether paramagnetic lanthanide ions bound to this system will exhibit $\Delta\chi \neq 0$, which would open a perspective to use this system as a paramagnetic tag for nucleic acid molecules. NMR data that I collected in paper **H3** demonstrated that indeed for this system the whole set on paramagnetic effects was observable. Thanks to this fact, when solving the structure of this system, I was able to complement the classic NOE restraints with a rich set of PCS restraints. As described in more detail in Chapter 2.2, this allowed for a much more precise determination of the DNA spatial structure, as well as – for an accurate determination of the position assumed by the metal ion – what remains very rare in NMR structure determination for nucleic acids.

***Characterization of metal binding sites in NMR structures – MD simulations and CSP measurements***

In NMR studies of nucleic acid complexes with metal ions, the determination of the spatial structure of the DNA/RNA component using methods described in the previous sections does not usually provide any information about the position of the bound metal ion. To describe its properties, it is thus necessary to either conduct additional NMR experiments or to employ computational methods. Within this second approach one can perform molecular dynamics (MD) simulations taking as the

starting point the determined NMR structure. The simulations are performed in the explicit solvent mode, with water molecules and metal ions being directly represented throughout the calculations. Through the analysis of the obtained molecular dynamics trajectories the regions on the DNA/RNA surface in which metal ions preferentially reside can be identified. The occupancy and interaction geometry of each of these sites can also be described. I have applied this approach to the study of G-quadruplex r(UGGUGGU)$_4$ interactions with $K^+$ ions (papers **H1** and **H2**). The main limitation of this group of methods is the fact that – due to necessary approximations in the description of physical properties of the studied complex – observation of a specific mode of interactions in the simulations does not guarantee their presence for the actual physical system.

For this reason, the capability to even approximately locate metal binding sites using NMR is greatly desired. Such a goal can be achieved by the measurement of chemical shift perturbations (CSP) of various atoms in the DNA/RNA molecule induced by metal binding.[35] CSP values are obtained through performing metal ion titration of the studied DNA/RNA sample, obtaining resonance frequency assignment at different point throughout the titration and calculating the difference in chemical shift experienced by each atom. Measured CSPs are then mapped onto the previously solved spatial structure of the DNA/RNA molecule, which allows for the identification of regions in the structure whose chemical environment is sensitive to the presence of the metal ion. In an optimal scenario the information derived from CSP measurements can be used in tandem with the results of computational MD studies. In such a case CSPs can be used for a "low-resolution" identification of the binding site, while the subsequent simulations can provide a more in-depth geometric picture of the occurring interactions.[35] Alternatively, MD simulations can be performed first and then CSP values applied to verify which among the *in silico* observed binding sites are actually present in the real system.[36] A significant limitation of methods based on CSP derived from the fact that not all of the measured perturbations have to be caused by the direct influence of the bound metal ion on the electron density of atoms in its nearest neighborhood – which was a silent assumption in the presented reasoning. Such an assumption is true only when the complex formation does not significantly alter the spatial structure of the DNA/RNA itself. In the opposite case a significant portion of the measured CSPs will be induced by the conformational transition occurring alongside metal binding which will preclude their correlation with direct proximity to the bound ion. In my research work I concentrate explicitly on systems whose tertiary structures fold in response to metal ion binding and thus the application of CSP-based methods in my work is significantly hampered. Nevertheless, I was able to apply this technique for the identification of $Zn^{2+}$ binding site to the 8-17 DNAzyme, thanks to being able to pre-structure this system using high concentrations of $Na^+$ ions exhibiting a much lower affinity towards it (paper **H4**).

***Characterization of metal binding sites in NMR structures – NOE, PRE and PCS***

The measurement of CSPs is a method that can be applied for the study of DNA/RNA interactions with a broad spectrum of metal ions, which however can deliver only approximate information about the binding site geometry. Luckily, for selected metal ions additional, higher resolution information can be obtained thanks to the application of methods based either on NOE or paramagnetic effects – PRE and PCS. The most important metal ion interacting with RNA molecules in vivo – $Mg^{2+}$ – cannot be directly observed using NMR methods. However, already in the 1990s it was noticed that an inert complex of cobalt(III), $Co(NH_3)_6^{3+}$, interacts with very similar sites on RNA surface as a fully-hydrated magnesium ion $Mg(H_2O)_6^{2+}$.[37] The protons of $NH_3$ molecules in this cobalt complex are directly observable in the NMR spectra and NOE restraints can be measured between them and atoms belonging to the RNA molecule. This led to a broad usage of $Co(NH_3)_6^{3+}$ as a $Mg^{2+}$ substitute in RNA-metal interaction studies. Paramagnetic transition metal ions – $Mn^{2+}$ and $Ni^{2+}$ (with $\Delta\chi = 0$ in both cases) – were also utilized in a similar capacity, this time taking advantage of the fact that they generate paramagnetic relaxation enhancements (PRE) in their direct vicinity.[36–38] The PRE effect was itself used in two different ways in this context. In some papers it was employed purely qualitatively, by listing atoms on the RNA surface for which the broadening was observable and in consequence obtaining information similar in nature to the CSP.[36,37] In others, the strength of the PRE effect was analyzed quantitatively and converted into distance restraints, similar in nature to NOE restraints, yet observable at a greater range.[38] This second approach allows in theory to very precisely locate the positions assumed by metal ions in complexes with DNA/RNA, however in practice a precise extraction of the PRE values from the spectra can be difficult,[39] while their quantitative interpretation remains additional assumptions about parameters describing the relaxation process.[40] Similar problems aren't encountered for the measurement and interpretation of another type of paramagnetic NMR restraint – the pseudo-contact shift (PCS). Unfortunately, in this case the requirement for the bound metal ion to exhibit $\Delta\chi \neq 0$ have thus far significantly limited the number of DNA/RNA systems for which this effect could be observed. Indeed, only two structures of nucleic acid systems were thus far deposited in the PDB for which PCS restraints were used in the structure determination process. First of them is the structure of a DNA duplex in complex with the antibiotic chromomycin A3, which binds in the narrow groove of the double helix and in turn interacts with a paramagnetic $Co^{2+}$ ion.[41] The second one is the structure of the lanthanide binding aptamer that I solved in paper **H3**. This second structure constitutes the first system for which PCS restraints induced by direct interaction between a nucleic acid molecule and a paramagnetic metal ion were measured. This discovery provides an avenue towards a much broader application of paramagnetic restraints in NMR studies of nucleic acids in the future, thanks to the possibility of transplanting the identified lanthanide binding site to other DNA/RNA systems (Chapter 3.2).

***DNA/RNA-metal interaction and their NMR studies – a summary***

Nucleic acids interact with metal ions in a significantly different manner than protein systems do. These interactions are characterized not only by rather low affinities, but also often by limited specificity – understood both in terms of selectivity towards a certain ion type (metal-specificity) an in terms of the degree of binding site definition is space (site-specificity). These features can severely hamper both structural and functional characterization of these interactions by experimental methods. Despite these limitations, interactions with metal ions are a key factor in shaping the spatial structures of nucleic acid molecules and in case of catalytically functional systems – ribozymes and DNAzymes – also their activities. NMR structural studies of nucleic acid complexes with metal ions are additionally hampered by the impossibility of direct observation of most among biologically relevant metal ions. For this reason, to describe their binding sites one has to apply a series of methods based on the observation the effects these ions induce in the NMR parameters of atoms belonging to the DNA/RNA molecule. Similar studies are also often complemented by the use of computational methods. For nucleic acid structures whose very formation is induced by metal ion binding, the first and often most work intense step in the study of the interaction is the determination of the DNA/RNA structure itself.

# Chapter 2

# Nucleic acid complexes with metal ions studied within the presented habilitation achievement

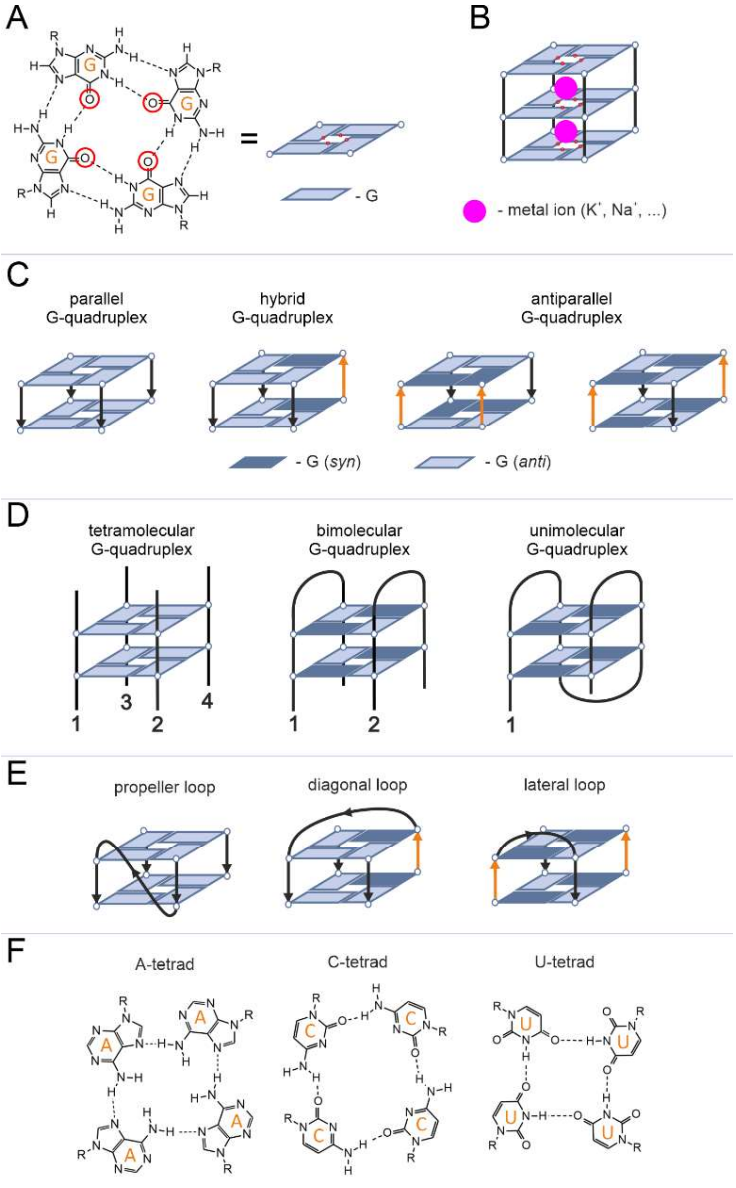
## 2.1 Papers H1 and H2 – determination of the spatial structure of r(UGGUGGU)$_4$ G-quadruplex and explaining the reason behind its exceptional thermal stability.

***Structural diversity of G-quadruplexes***

DNA and RNA G-quadruplexes are a biologically important[42,43] and in consequence intensely studied type on noncanonical nucleic acid structures. At the same time, they constitute the arguably most famous example of DNA/RNA structures induced through site-specific metal ion binding. The basic structural blocks of G-quadruplexes are G-tetrads, formed through Hogsteen-type hydrogen bonding interactions between four guanosine nucleobases (Figure 2.1.1a).[44] The simplest G-quadruplexes consist of only two stacked G-tetrads. However, even these simplest structures cannot form without specific binding of a metal ion (Figure 2.1.1b).[45] This is because stacking interaction between two G-tetrads puts eight guanosine carbonyl group oxygen atoms (O6) in a several angstroms range from each other. Each of these atoms carries a partial negative charge and such an accumulation of unnaturalized charge would result in strong electrostatic repulsion. Metal ion binding in the so-called central channel of the G-quadruplex, around which the mentioned O6 atoms are located, screens

their electrostatic repulsion and provides instead stabilizing interactions with a positive charge of the bound cation. In living cells G-quadruplex structures are most often stabilized by monovalent $K^+$ ions that locate themselves between G-tetrads and directly coordinate all eight O6 atoms of the neighboring tetrads.[45] Cations with a smaller ionic radius can assume somewhat different positions within the channel – for example $Na^+$ ions usually locate themselves closer to the plane on a single G-tetrad[45] – yet the interaction is always based in direct coordination of the O6 atoms and requires full dehydration of the bound ion. Despite the presence of direct metal ion coordination by DNA/RNA, binding sites forming in G-quadruplex channels are characterized by a limited metal-ion selectivity and can accept a broad range of different cations.

**Figure 2.1.1.** Structural diversity of G-quadruplexes. A) chemical structure of a G-tetrad with carbonyl O6 oxygen atoms marked in red, B) a scheme of the G-quadruplex core with bound metal ions marked, C) strand orientations in G-quadruplexes, D) G-quadruplex molecularity, E) G-quadruplex loop types, F) chemical structures of non-guanine tetrads.



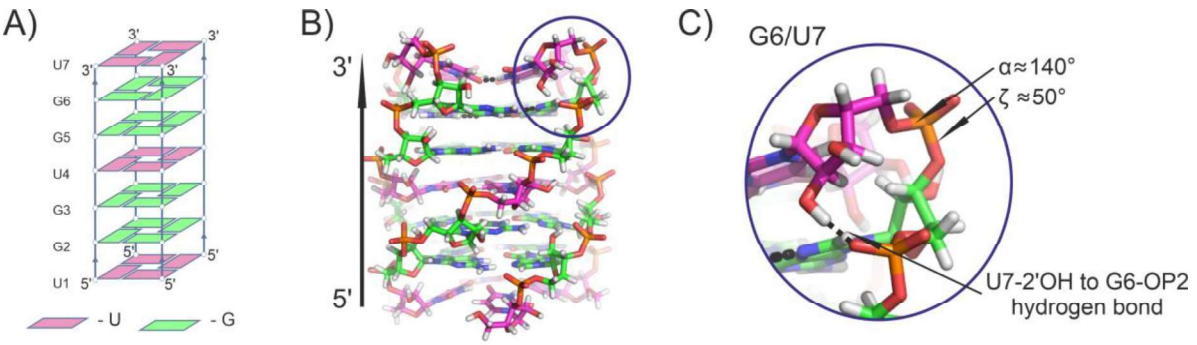
A very important feature of G-quadruplexes is their great structural heterogeneity. The G-quadruplex core – a stack of G-tetrads – is composed of four independent DNA/RNA strands, each of which contributes a single guanosine residue to each of the tetrads. G-quadruplex structures accept any combination of relative strand directions (Figure 2.1.1c). We can thus distinguish parallel (same direction of all four strands), hybrid (one strand with direction opposite to the other three) and antiparallel (two strands going in one direction and other two in the opposite) G-quadruplexes. Additionally, while in the construction of the G-quadruplex core the four guanosine tracts forming it can be treated as independent strands, in the covalent structure of the DNA/RNA molecule involved they can be connected by additional runs of unpaired nucleoside residues, forming the so-called G-quadruplex loops. Next to tetramolecular G-quadruplexes, in which each strand in the quadruplex core indeed constitutes a separate DNA/RNA molecule, we can also encounter bi- and unimolecular G-quadruplexes (Figure 2.1.1d). For these last two groups we can additionally distinguish three types of loop region conformations – the so-called propeller, lateral and diagonal loops (Figure 2.1.1e). The presence of these numerous and independent degrees of freedom lead to the situation in which even limiting ourselves to unimolecular G-quadruplexes we can encounter 26 different topologies. This amazing multitude of G-quadruplex folds is the reason why, even after over three decades of intense studies and determination of over 500 high-resolution structures, the task of predicting the topology of a G-quadruplex molecule based on its primary sequence remains a largely unsolved problem.

***Special features of RNA G-quadruplexes and non-guanine tetrads***

A seemingly unimportant structural detail of all G-quadruplex topologies, apart from the parallel one, is the presence of guanosine residues assuming the *syn* conformation around their glycosidic bonds.[46] This fact, however, is of key importance for RNA G-quadruplexes, as in this biopolymer *syn* conformations of purine residues are highly energetically unfavorable. In consequence, for RNA G-quadruplexes, regardless of their molecularity, the parallel topology is almost exclusively observed.[47,48] It does not however signify that these systems do not display conformational diversity. Nucleobases other than guanosine do not form homo-tetrads stable enough to become basic building blocks of their own DNA or RNA quadruplex architectures (hypothetical A-, C- or U-quadruplexes). Nevertheless, these kinds of tetrads were experimentally observed for DNA and RNA molecules in which a stable G-quadruplex core has provided a scaffold to properly position four copies of the appropriate base. Such A-,[49] C-,[50] T-[51] and U-tetrads[52] constitute an additional building block characteristic for G-quadruplexes (Figure 2.1.1f), further increasing the degree of their conformational diversity and sometimes imbuing their structures with unexpected physical properties. An important example of an RNA G-quadruplex, whose properties are to a large degree molded by the presence of U-tetrads is the tetramolecular G-quadruplex r(UGGUGGU)$_4$ studied by me in papers **H1** and **H2**. The studies of this system commenced in the research group led by prof. dr hab. Zofia

Gdaniec at IBCh PAS already before I joined her group in January 2017 and trace their beginnings in the studies of the propensity of trinucleotide repeats AGG, CGG and UGG to form G-quadruplex structures.[53] Initial NMR studies, complemented by a mass spectrometry analysis, have shown the ability of the rUGGUGGU RNA molecule to fold into a tetramolecular G-quadruplex containing three U-tetrads in its structure (Figure 2.1.2a), in the presence of both $K^+$ and $Na^+$.[53] Additionally, a UV-melting analysis of this G-quadruplex have revealed that it is characterized by a very high thermal stability. Interestingly, a series of unexpected NOE contacts was measured in the proximity of the 3'-terminal U-tetrad, suggesting the presence of an atypical RNA conformation in this region. However, the nature of this conformation, as well as its possible links to the system's exceptional stability remained a mystery.

**Figure 2.1.2.** r(UGGUGGU)$_4$ G-quadruplex structure A) secondary structure determined in ref. 53, B) 3D spatial structure, C) hydrogen bonding between 3'-terminal uridine 2'OH group and a phosphate group. The figure contains elements of Figure 1 from paper **H1**.




***The determination of r(UGGUGGU)$_4$ G-quadruplex structure***

For this reason, the first scientific endeavor I have undertaken after joining the team of prof. dr hab. Zofia Gdaniec was the determination of the high-resolution three-dimensional structure of the r(UGGUGGU)$_4$ G-quadruplex using NMR data collected in the presence of $K^+$ ions. As I did not yet possess tested procedures for the extraction of structural restraints from NMR spectra, setup of structure calculations and analysis of the obtained structures I had to develop all of these procedures during the process of r(UGGUGGU)$_4$ structure determination, based on available literature counterparts. For the structure calculations I have employed the classical NOE and torsion angle restraints. In similar endeavors for each new system studied one has to ascertain that the applied calculation scheme provides broad enough sampling of the conformational space of the biomolecule as well as that it allows to easily surpass the energy barriers between different local minima in order to provide a chance of uncovering the global minimum – the final NMR structure. In the case of r(UGGUGGU)$_4$ G-quadruplex a serious complication arose due to four-fold degeneracy of observed NMR resonance frequencies, caused by the symmetry of the studied structure. This translated to the

fact that for each NOE restraint extracted from the NMR spectra it was impossible to determine whether the observed NOE is due to the spatial proximity of two atoms belonging to the same RNA molecule or is it a contact to one of the other three interacting strands or perhaps a sum of both contributions. For this reason, in the calculations, I have treated every NOE as an ambiguous restraint, that can be satisfied in more than one way. After a tedious optimization of the calculation procedure I have settled for a two-step approach, described in detail in paper **H1**. It was based on first finding an approximate structure starting from a random RNA conformation and using a simplified force field for the description of the systems covalent parameters and electrostatics, followed by a second round of calculations starting from this approximate structure and using a more sophisticated force field to better capture the local geometry of the system. This procedure allowed me to obtain a well-defined NMR structural bundle, which I deposited at the PDB, under the access code 6GE1 (Figure 2.1.2b).

***The r(UGGUGGU)$_4$ G-quadruplex contains a "reversed U-tetrad" structural motif***

The determined structure revealed some atypical backbone dihedral angle values in the vicinity of the central U-tetrad, however its most unexpected feature was the conformation assumed by the 3'-terminal region of the molecule. In the solved structure I observed a rapid, almost 180 degrees turn of the sugar-phosphate backbone just before the final uridine residue (Figure 2.1.2b). Such a conformation inverts the polarity (the identity of hydrogen bond acceptors and donors) of the last U-tetrad with respect to all the other tetrads present in the G-quadruplex and completely changes the stacking interactions geometry of the 3' uridine residues with the preceding guanosines. I called this atypical conformation a "reversed U-tetrad". As the most distinctive feature of the solved structure this element appeared as a plausible source of its high thermal stability. Following up on this assumption I have performed an in-depth literature query for information regarding G-quadruplexes containing a 3'-terminal uridine residue. It revealed that similar "reversed U-tetrads" were observed before in some crystal structures,[54] however their properties were never discussed, perhaps due to assumptions that such an atypical conformation of a peripheric residue may be simply an artifact of crystal packing. I have also encountered several articles not containing structural data, yet observing that the presence of a 3'-terminal uridine in tetramolecular G-quadruplexes can increase their thermal stability by up to 30 °C.[55,56]

These data motivated me to analyze the structure of the "reversed U-tetrad" motif more closely, in search of specific interactions that may so significantly stabilize this conformation. The first candidate was the likely formation of an additional hydrogen bond between 2'OH group of the terminal uridine and the phosphate group of the preceding guanosine (G-1), made possible thanks to the sharp turn of the sugar-phosphate backbone accompanying the formation of the "reversed U-tetrad" (Figure 2.1.2c). The likely involvement of this 2'OH group in hydrogen bonding interactions

was made apparent already in the early stages of structural studies by its high resistance to exchange with the solvent in NMR spectra, yet only the determination of the three-dimensional structure allowed to proposed the phosphate group as the hydrogen bond acceptor. In the **H2** paper, which continued the analysis of the "reversed U-tetrad" motif, I managed to obtain a direct experimental confirmation of this hydrogen bond, through the observation of a J-coupling between the proton of the 2'OH group and the phosphorus atom of the appropriate phosphate group. However, while preparing the **H1** paper I did not yet possess this information and thus I decided to study the kinetic stability of this hydrogen bond as well as other factors possibly stabilizing the "reversed U-tetrad", such as for example an atypical hydration pattern, by the means of explicit solvent molecular dynamics (MD) simulations. This aspect of the study I conducted in collaboration with dr Joanna Sarzyńska, who specializes in MD simulations of nucleic acid systems. In all the performed simulations, 6 μs in total, the expected hydrogen bond was present throughout the entire run. The analysis of the obtained MD trajectories also revealed the formation of a long-lived hydration structure, yet around the central U-tetrad, not the 3'-terminal one (Figure 2.1.3). Instead, it suggested the existence of an additional $K^+$ binding site in the structure of the "reversed U-tetrad", located outside of the G-quadruplex central channel (Figure 2.1.3). This ion would bind in the G-quadruplex groove and directly coordinate four atoms belonging to the RNA molecule. This binding site would also be, according to the simulation results, to at least some degree selective, as no interactions were observed for the runs performed in the presence of the $Na^+$ ion instead. Based on these results, in the paper **H1** I have proposed the presence of four additional hydrogen bonds and four potassium binding sites (one for each RNA molecule within the G-quadruplex) in the structure of the "reversed U-tetrad" motif as factors potentially responsible for its stabilizing influence on G-quadruplex structures.

**Figure 2.1.3.** Positions preferentially assumed by $K^+$ ions (magenta) and water oxygen atoms (green) during MD simulations of r(UGGUGGU)$_4$ G-quadruplex**.** The figure is a copy of Figure 4 from paper **H1**.

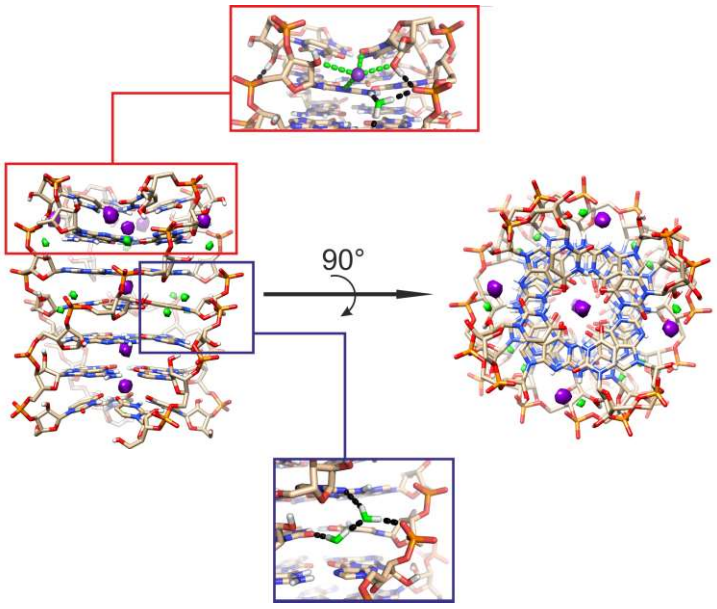
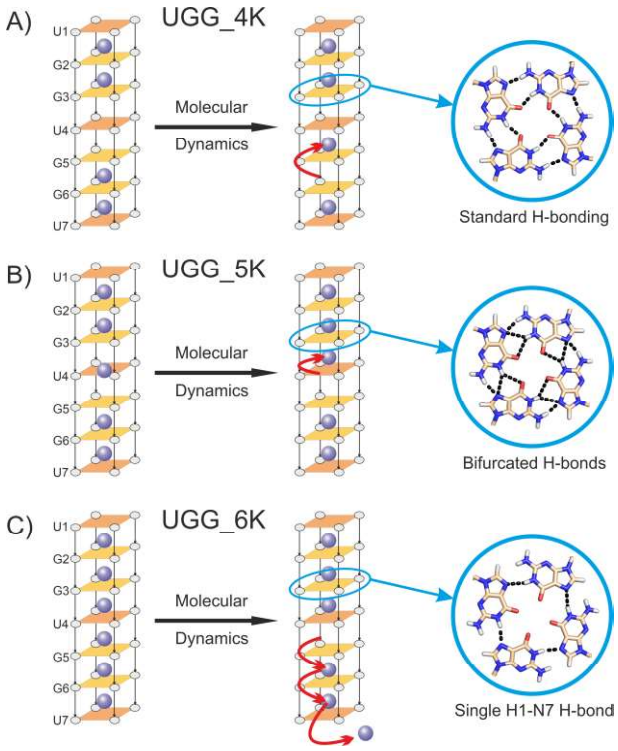
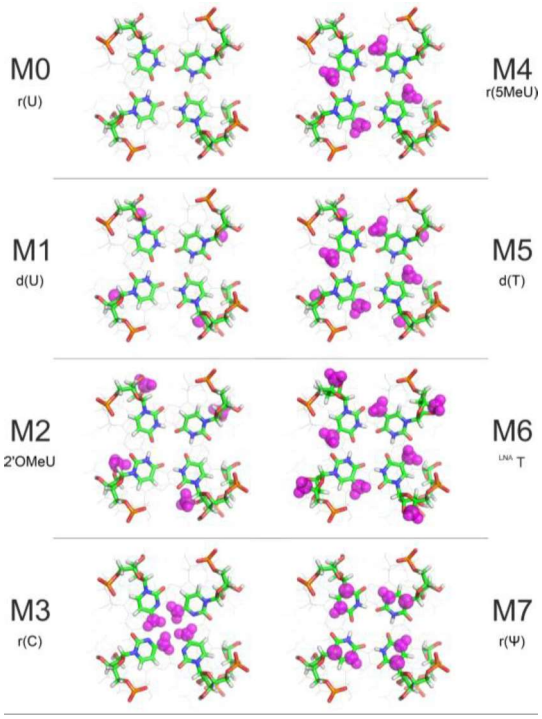


**Figure 2.1.4.** The behavior of $K^+$ ions in the central channel of r(UGGUGGU)$_4$ G-quadruplex observed in MD simulations. The simulations were conducted with four (A), five (B) or six (C) $K^+$ ions initially present in the channel. The figure is a copy of Figure 3 from paper **H1**.




Next to an in-depth analysis of the structural features of the "reversed U-tetrad", the second motivation for the MD study of the r(UGGUGGU)$_4$ G-quadruplex was to investigate the influence of the presence of three U-tetrads in its structure on metal ion binding in the central channel. A G-quadruplex consisting of seven consecutive G-tetrads should theoretically be able to bind six $K^+$ cations. However, the smaller size of a U-tetrad leads to the tightening of the central channel in its vicinity which may lead to the inability to accept this many ions by the r(UGGUGGU)$_4$ molecule. Due to the involvement of $K^+$ ions bound in the channel in the G-quadruplex formation process itself, an NMR characterization of this interaction is severely hindered (as I will discuss further in Chapter 3.1). Computational studies are in turn limited by the inertness of the channel-bound ions, which precludes simulation convergence in terms of their equilibrium properties. Nevertheless, performing MD simulations starting with either four, five or six $K^+$ ions in the channel allowed me to conclude that the presence of four channel-bound $K^+$ ions is the most likely. This is because in the other simulations either one of the ions was expulsed from the channel during the MD run or significant hydrogen bonding geometry distortions were observed in G-tetrads (Figure 2.1.4).

**Figure 2.1.5.** The set of chemical modifications introduced to the 3'-terminal residue of r(UGGUGGU)$_4$ G-quadruplex in order to understand the sources of the thermal stability of the "reversed U-tetrad" motif. The figure is a copy of Figure 2 from paper **H2**.




***Experimentally verifying the sources of the "reversed U-tetrad's" stabilizing effect***

The connection between structural features of the "reversed U-tetrad" motif and its stabilizing effect on G-quadruplex folds was also the subject of paper **H2**. In this work I have experimentally tested the contributions to the stabilizing effect arising from the previously proposed sources – additional hydrogen bonds and $K^+$ binding in G-quadruplex grooves – as well as other potentially relevant factors, such as stacking geometry and steric factors. To achieve this goal, I have introduced a series of carefully selected chemical modifications into the studied G-quadruplex, whose task was to selectively influence some among the potential stabilizing factors listed above. For example, through the substitution of the 3'-terminal (ribo)uridine residue by its deoxy- counterpart I expected to selectively remove the 2'OH-to-phosphate hydrogen bonding, without affecting the stacking interactions or hydrogen bonding between nucleobases. The series of introduced 3'-terminal modifications is presented in Figure 2.1.5. The effects of each such modification were analyzed on two planes: structural (whether the modification changes the conformation of the 3'-terminal part of the G-quadruplex) and thermodynamic (how the thermal stability is affected in the modified system). Within the structural part my experimental contribution concentrated around the NMR analysis of each modified oligomer. I have also acquired and analyzed circular dichroism (CD) spectra for all studied systems, in order to check whether the "reversed U-tetrad" motif possesses a characteristic CD-

spectroscopic fingerprint. The described NMR and CD analysis was performed twice – in the presence of $K^+$ and $Na^+$ – with qualitatively similar results. This confirmed the lack of influence of the bound ion type on the conformational preferences of the studied systems. In parallel a co-author of the **H2** paper, dr Karolina Zielińska, measured the melting temperature of each G-quadruplex using UV spectroscopy methods. A combined analysis of the collected data allowed us to single out the factors truly responsible for the stability of the "reversed U-tetrad" motif. For example, for the deoxyU modification (M1 in Figure 2.1.5) the structural analysis has shown the formation of an unaltered "reversed U-tetrad" conformation, while the UV data revealed a decrease of the melting temperature of the system by around 10 °C. The same effect was observed in the presence of both $K^+$ and $Na^+$. The introduction of the dU modification selectively removed the four hydrogen bonds formed by the 2'OH groups and thus the observed decrease of the melting temperature was interpreted as proof that these interactions have a significant contribution to the stability of the reversed tetrad. This conclusion was not so clear cut at first glance, as the uridine O2' oxygen atoms also constitute one of the direct ligands of the $K^+$ ions, observed to bind in the G-quadruplex grooves in MD simulations. Their removal may thus disrupt the metal interaction with this site. A significant influence of this factor was however excluded thanks to the observation that the destabilizing influence of the dU modification was independent on whether $K^+$ or $Na^+$ was present. Based on this observation we have concluded that $K^+$ binding in the G-quadruplex grooves has no significant effect on its thermal stability, contrary to our previous assumptions. The structural analysis of other modified oligomers has demonstrated that formation of a "reversed U-tetrad" is very sensitive to the presence of additional steric crowding in its vicinity and that a potential second element important for its stability are favorable stacking interactions. This thesis was corroborated by an additional *in silico* analysis using the MM-GBSA method. These calculations compared the energetics of the native $r(UGGUGGU)_4$ G-quadruplex form, containing "reversed U-tetrad", to its hypothetical conformation featuring a "standard" U-tetrad and the most significant differences were indeed observed in the energetic term containing the stacking interactions.

***The study of $r(UGGUGGU)_4$ G-quadruplex and the "reversed U-tetrad" motif – a summary***

In papers **H1** and **H2** I proposed an explanation for an unexpectedly strong stabilization of tetramolecular RNA G-quadruplex folds, related to the presence of a 3'-terminal uridine residue in their structures. I demonstrated that such systems form a structural motif that I dubbed a "reversed U-tetrad" and that additional hydrogen bonds formed by 2'-OH groups and favorable stacking interactions are important factors contributing to its stabilizing effect. In the course of these studies, I also suggested the formation of additional $K^+$ binding sites in G-quadruplex grooves, however their presence was only observed *in silico* and have not yet been experimentally confirmed (see Chapter 3.1).

## 2.2 Paper H3 – determination of the spatial structure of the lanthanide binding aptamer using paramagnetic PCS restraints

***Lanthanide binding nucleic acids and paramagnetic effects in biomolecular NMR***

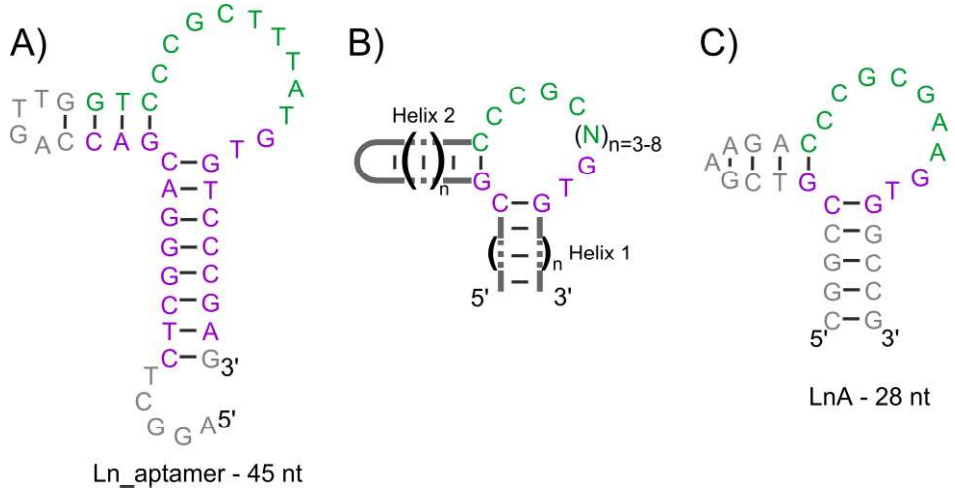
An important motivation behind my interest in nucleic acid molecules capable of tight and specific metal ion binding is the prospect of using such systems to expand the capabilities of biomolecular NMR methods. As described in Chapter 1.2, the local character of classic NMR restraints negatively impacts the accuracy and precision of the determination of global features of complex nucleic acid structures.[29] A solution to this problem lies in the application of long-range NMR restraints – RDC, PCS, PRE – however access to them for nucleic acid studies remains limited. Methods for inducing the PCS effect, whose observation requires a rigid attachment of a paramagnetic metal ion with $\Delta\chi \neq 0$ to the studied molecule, are especially lacking.[32] The most valuable source of PCS restraint are lanthanide ions that, thanks to their similar chemical properties within the series, usually all bind to the same sites allowing for the measurement of several sets of complementary PCS restraints. Due to limited affinity of nucleic acids to metal ions, discussed in Chapter 1.1, as well as often less defined interaction geometry, for these polymers it was not yet possible to identify a lanthanide binding structural motif capable of inducing strong PCS, usable as structural restraints. For this reason I turned my attention to a DNA aptamer discovered in 2016,[26] capable of binding lanthanide ions with a sub-micromolar affinity, as well as to a series of DNAzyme molecules activated by lanthanide binding[24] and decided to study the binding sites present in these systems. This research topic – the application of lanthanide binding by specific DNA structural motifs to induce paramagnetic NMR effects – was not pursued before in my Institution. It constitutes my own research idea for which I secured funding in the framework of NCN OPUS19 grant program (research grant entitled: "Lanthanide binding oligonucleotides (LBOs) as paramagnetic tags for NMR spectroscopy of nucleic acids"). Paper **H3** presents the results of a study of the lanthanide binding aptamer structure (denoted LnA from now on) and the geometry of its metal binding site. Within this paper I also demonstrate the high value of PCS induced in this system as NMR structural restraints.

***Structure determination of LnA using paramagnetic restraints***

Before discussing results I obtained it is important to note that aptamer and DNAzyme molecules are most often identified by the so called SELEX method (abbreviation for *"Systematic Evolution of Ligands by Exponential Enrichment"*),[57] also known as *in vitro* selection.[58] This methodology allows one to select sequential motifs interacting with a chosen molecular target from a broad pool of DNA or RNA molecules. It does not however directly provide any information regarding the structure assumed by the selected aptamers nor on the geometry and location of the interaction site. Some preliminary data can be obtained through the analysis of nucleotide conservation

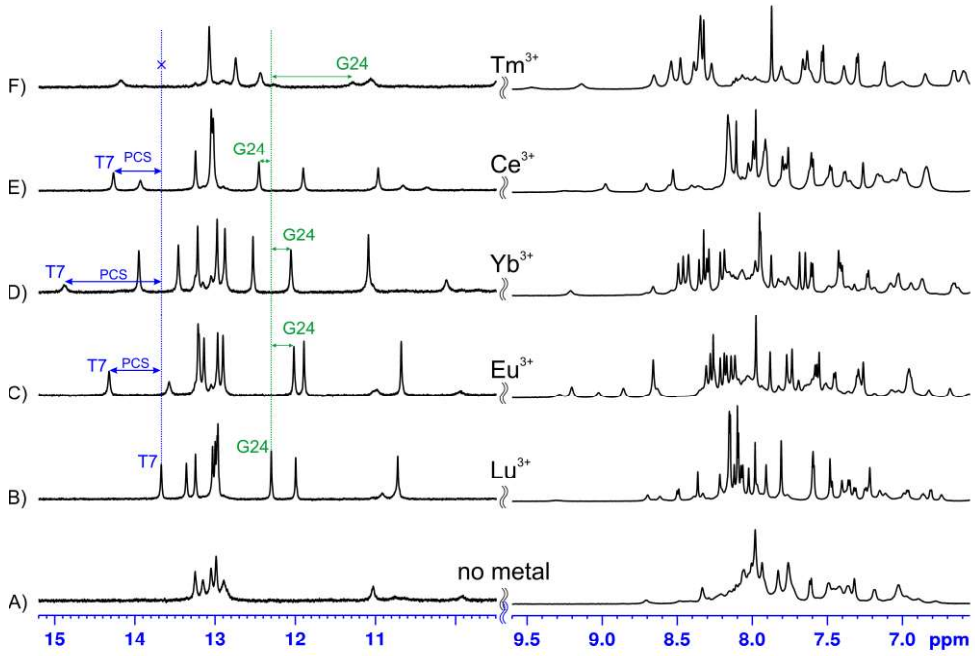
among the selected DNA/RNA molecules, however the determination of aptamer structures and binding site architectures require separate studies using one of the high-resolution structural biology methods.

**Figure 2.2.1.** Secondary structure of the lanthanide binding aptamer. A) construct from original publication – „Ln_aptamer",[26] B) nucleobase conservation between variants obtained in the SELEX experiment, C) construct optimized for NMR studies – „LnA". Residue color code: magenta – the constant region during SELEX, green – the variable region during SELEX, gray – residues introduced post-SELEX. The figure contains elements of Figure 1 from paper **H3**.




In the case of LnA, the initial analysis of nucleotide sequences uncovered by the SELEX method, performed by the authors of that experiment allowed to create a 45 nucleotide long aptamer construct and propose its secondary structure, presented in Figure 2.2.1a.[26] High level of conservation of some of the nucleobases in the unpaired region connecting two helical elements (Figure 2.2.1b) allowed to propose this region as the likely site of interactions,[26] yet without suggesting any details of the occurring interactions. Before commencing deeper structural studies of this system using NMR methods I decided to attempt a further optimization of LnA sequence to obtain a shorted variant, greatly facilitating the NMR analysis. I designed a series of shorter aptamer constructs and evaluated the lanthanide binding capabilities of each among them using UV and CD spectroscopic methods. This was made possible by the fact that lanthanide binding induces significant changes in both the thermal stability and structure of the aptamer, observable by the two mentioned techniques. In this way I was able to obtain a LnA variant containing only 28 nt (Figure 2.2.1c), while preserving unperturbed lanthanide binding affinity ($K_d \approx 0.3$ μM).

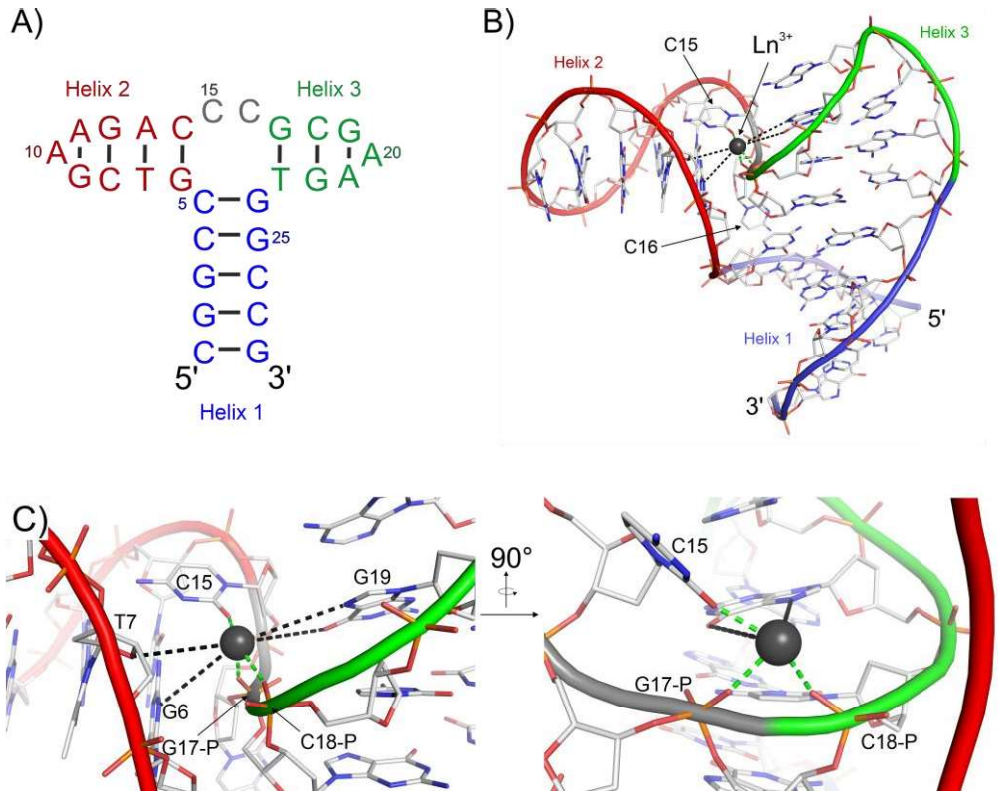
**Figure 2.2.2.** Imino and aromatic regions of $^{1}$H NMR spectra of LnA_28 either free (A) or in presence of: $Lu^{3+}$ (B), $Eu^{3+}$ (C), $Yb^{3+}$ (D), $Ce^{3+}$ (E), or $Tm^{3+}$ (F). The figure is a copy of Figure 2 from paper **H3**.




Already a preliminary analysis of $^{1}$H NMR spectra recorded for this system demonstrated the formation of several new base pairs as a consequence of the presence of a diamagnetic lanthanide ion-lutetium $Lu^{3+}$ (Figure 2.2.2a vs b). Titration of LnA with a series of other, paramagnetic, lanthanide ions lead to the observation of strong NMR signal shifts with respect to the diamagnetic sample which after resonance assignment could be transformed into PCS structural restraints (Figure 2.2.2b-f). Already the assignment of imino proton resonances allowed for the identification of one new structural element forming in response to metal ion binding. Namely the observation of wobble base pairing between residues G23 and T23 as well as of an additional GC base pair neighboring it in space demonstrated the structuring of a large portion of the supposed loop region into a third, short helical element (Helix 3, Figure 2.2.3a). To determine the relative orientation of the three helical elements in space, and thus the tertiary structure of the aptamer, I performed the full resonance frequency assignment, derived NOE and PCS structural restraints and performed molecular dynamics simulations restraints with this NMR data. In the first stage of structure calculations I determined a preliminary spatial structure of the system using only classic NMR restraints (NOE and torsion angle). I then used this structure to approximately locate the metal ion using PCS data. In the second calculations stage I refined the structure of the entire complex using both classic restraints and PCS which are sensitive to the position of each atom in the DNA molecule with respect to the metal ion. To perform PCS-restrained simulations I had to edit the source code of the AMBER18 molecular dynamics software to properly describe these effects in the simulations. The performed calculations

allowed me to determine the structure of LnA in complex with lanthanide ions (Figure 2.2.3b), which I deposited in the PDB database under access code 7QB3.

**Figure 2.2.3.** The structure of LnA aptamer induced by lanthanide ion binding. A) The real secondary structure, B) the spatial strucutre, C) the lanthanide binding site. In panel C the green lines illustrate direct DNA-metal contact, while the black ones DNA-metal distances corresponding to second sphere interactions. The figure is a compilation of elements of Figures 3, 4, and 5 from paper **H3**.




***Tertiary structure features of LnA and the geometry of lanthanide binding site***

LnA assumes a rigid tertiary fold formed through co-axial stacking of Helices 1 and 3 (Figure 2.2.3b). In turn Helix 2 is anchored at around a right angle with respect to the other two. Such an arrangement of helical elements leads to a fragment of the sugar-phosphate backbone becoming buried between Helices 2 and 3. Importantly in the structure deposited to the PDB the lanthanide ion is also present, as its position was directly determined experimentally using PCS restraints. The lanthanide ion interacts directly with three ligands belonging to the DNA molecule – a O2 atom of the unpaired residue C16 locate in the narrow groove of Helix 2 and two phosphate groups belonging to the "buried" fragments of the sugar-phosphate backbone (Figure 2.2.3c). It is thus the tertiary arrangement of the system that is responsible for the creation of this binding site. The presence of only three direct coordination bonds between the and the metal ion suggests that the latter remains partially hydrated, as lanthanide ions usually assume coordination numbers in the range of 7-9. To study whether the water molecules coordinating the bound metal ion form additional interactions with DNA I performed and

analyzed a series of additional explicit solvent MD simulations. Their results suggest the existence of a complex network of hydrogen bonds between the hydration waters and electronegative atoms belonging to the DNA molecule. Including these MD-observed interactions up to nine atoms belonging to the DNA molecule appear to be involved in lanthanide ion binding – three through direct coordination and six through hydrogen bonding – meaning that we are dealing with a highly complex binding site.

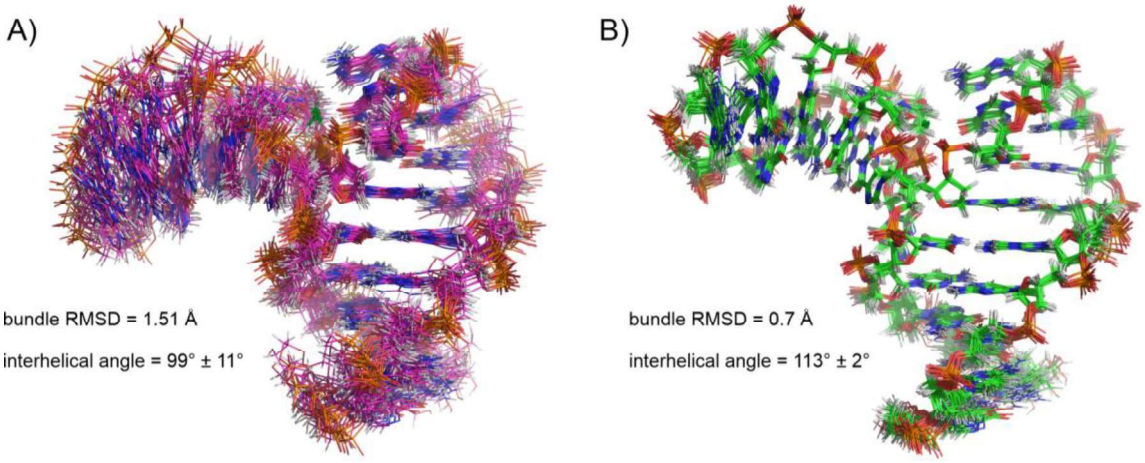
Structural characterization of a metal binding site specific towards lanthanide ions is important in the general context of understanding the sources of metal-specificity in DNA-metal interactions. However, uncovering the structural details of an interaction site constitutes only a departure point for attempts to explain how these details determine its physical properties – in this case metal-specificity. In the case of lanthanide binding site form LnA direct interactions are observed between the metal ion and two phosphate groups as well as a carbonyl group belonging to a cytidine residue. Phosphate groups, due to the negative charge that they carry, are among the most ubiquitous ligand groups in nucleic acid-metal ion interactions.[4] On the other hand, cytosine is the only free nucleobase to which lanthanide ions bind.[59] For this reason in paper **H3** I suggest that the reason for the observed lanthanide-specificity of LnA might be the presence of a cytidine nucleobase as one of three direct ligands of the bound metal ion.

***The influence of PCS restraints of accuracy and precision of the determined LnA structure***

The presence of a relatively rigid and asymmetric first coordination sphere of the metal ion is responsible for the non-zero values of magnetic susceptibility anisotropy $\Delta\chi$ of its unpaired electrons and in consequence for the possibility to observe the PCS effect. The measured PCS values turned out to the well interpretable in terms of DNA structural parameters, as demonstrated by a very good agreement between the experimentally measured effects with those predicted using the structure (through equation 1 in paper **H3**).

This means that the lanthanide ions bound to LnA aptamer constitute a promising source of long-range PCS restraints. I demonstrated their usefulness for structural studies already in the **H3** paper through a comparison of LnA structures determined with and without using these restraints (Figure 2.2.4). The NMR structural bundle obtained using PCS was significantly better defined. Global features of the LnA fold, such as the angle between Helix 2 and other helical elements, were determined with a much higher precision. Identification of a lanthanide binding site in DNA allowing for the observation of high-quality PCS restraints opens the prospect of a much broader use of paramagnetic effects in NMR spectroscopy of nucleic acid systems. This would be achieved through the isolation of the minimal lanthanide-binding motif from LnA and its reintroduction into other DNA/RNA systems which I proposed in paper **H3** and will discuss further in Chapter 3.2.

**Figure 2.2.4.** The influence of PCS restraints of the precision of LnA structure determination. NMR structural bundle obtained using A) only classical NOE and dihedral angle restraints, B) classical and PCS restraints. The figure contains elements of Figure 8 from paper **H3**.




***LnA aptamer structural studies – a summary***

Paper **H3** presents the spatial structure of a lanthanide binding aptamer solved using NMR methods. Two features distinguish the results obtained in this work from typical NMR-based structural studies, these are the precise experimental localization of the interacting metal ion and the rarely achieved convergence of the determined structural bundle (average pairwise RMSD of 0.7 Å). Achieving these results was made possible by the application of long-range paramagnetic NMR effects – PCS structural restraints. The structure of LnA is the first DNA structure in the PDB, for which a direct interaction between a lanthanide ion and a nucleic acid system allowed for the observation of strong PCS effects. The discovery of a system with this property opens the perspective towards a broader application of paramagnetic NMR restraints in DNA/RNA studies (Chapter 3.2).

## 2.3 Paper H4 – determination of the spatial structure of 8-17 DNAzyme in complex with a $Zn^{2+}$ ion and the formulation of a new model of the structure-function relationship for this system

***Structural studies of DNAzyme molecules***

Catalytically proficient RNA and DNA molecules, ribozymes and DNAzymes respectively, constitute an important type of nucleic acid systems in which metal ion binding can not only play a structural role, but also directly support their function. By far the most common reaction catalyzed by either type of nucleic acid catalyst is RNA hydrolysis, however RNA and DNA enzymes differ significantly in both their origin and the degree to which the structural basis of their function has be elucidated to date. While the most prominent types of RNA-hydrolyzing ribozymes were initially discovered in the genomes of a wide spectrum of living organisms, the existence of DNAzymes in

Nature was not yet observed.[60] All known systems of this type were obtained synthetically in the process of in vitro selection, already mentioned in Chapter 2.2. Despite not being present in Nature, DNAzymes possess a series of advantages over ribozymes for biotechnological applications, such as much higher chemical stability, including in cellular conditions or lower production costs.[61] Already since their discovery in the 1990s the prospect of using DNAzymes as therapeutic agents, capable of modulating mRNA levels in living organisms, attracted much interest.[62] However, such an application of DNAzymes still remains unrealized. This is primarily due to their limited activity in cellular conditions, in which their cofactors – divalent metal ions – are scarce.[63] Studies aiming at the design of DNAzyme variants more active at low metal ion concentrations are hindered by the lack of mechanistic understanding of these systems. While DNAzyme crystallization was attempted ever since their discovery, the first paper describing a catalytically proficient DNAzyme structure – for the 9DB1 DNAzyme capable of RNA ligation – was published only in 2016.[18] In the following year, crystallographic structures of 8-17 DNAzyme in presence of $Pb^{2+}$ and in the absence of a cofactor were also published.[19] These structures provided first high-resolution information for the elucidation of the molecular mode of action of DNAzyme systems. They also uncovered a previously unforeseen level of structural complexity in DNAzyme folds. Namely, they demonstrated that many sequence regions in DNAzymes, previously thought to be disordered, form stable secondary and tertiary structure elements, based on non-canonical base-pairing and pseudoknot formation.

***Motivation for NMR studies of the 8-17 DNAzyme***

These discoveries inspired me to attempt to study DNAzyme molecules using NMR methods. When I entered this field of research in 2018, no DNAzyme structures solved using NMR spectroscopy were available in the literature. Even the feasibility of such studies was not trivial to estimate, due to rather large sizes of typical DNAzyme systems – in the range of at least 60-70 nt. For this reason, using experience obtained during LnA aptamer studies (paper **H3**), I proposed a sequence optimization method of DNAzyme molecules for NMR studies. It allowed me to create an 8-17 DNAzyme construct consisting of only 35 nt while preserving unaltered catalytic activity. Thanks to the innovative character of the proposed studies, I managed to obtain funding for their continuation from NCN SONATA 14 grant program (project entitled: "NMR structural studies of 8-17 and I-R2 DNAzymes"). In the framework of this research grant I planned to determine the spatial structure of the 8-17 DNAzyme induced by $Zn^{2+}$ binding. Solving this structure was important, even though the crystallographic structures of this system induced by $Pb^{2+}$ binding and without any metal ion cofactor present were already known at that time.[19] This is because over twenty years of intense studies of this system suggested that it possesses two different catalytically proficient folds.[64] One of them, structurally identical to the apo form, was supposed to be employed by the 8-17 DNAzyme in presence of $Pb^{2+}$ cofactor (Figure 2.3.1a). On the other hand, the presence of other cofactors, such as

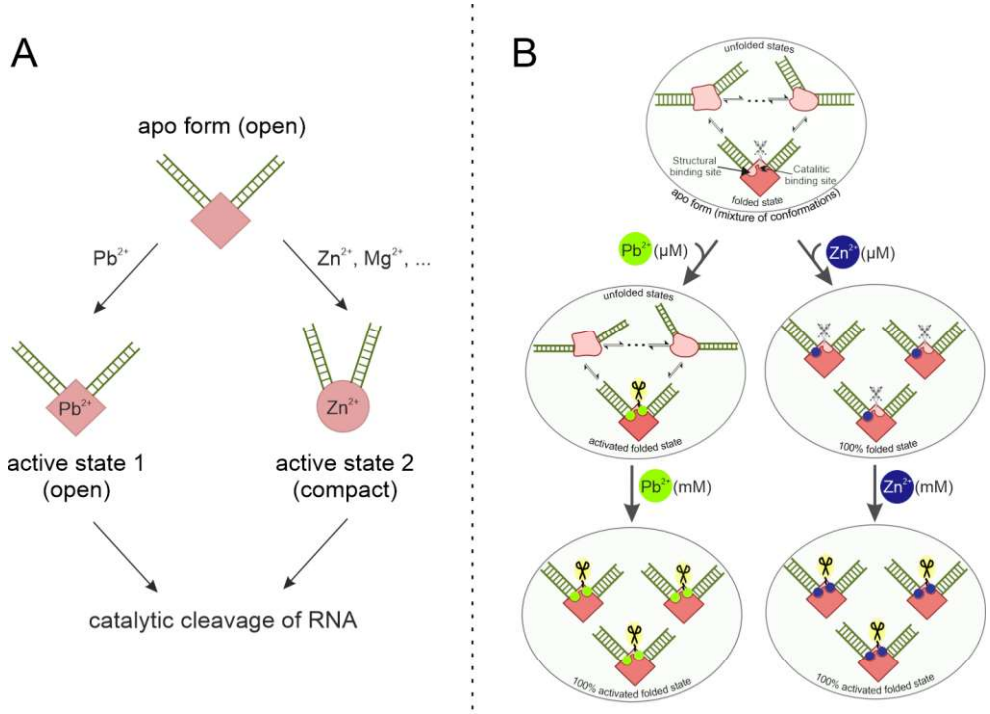
$Zn^{2+}$ or biologically relevant $Mg^{2+}$, was observed to induce a global structural rearrangement towards a second, more compact active structure (Figure 2.3.1a). This model was formulated based on results obtained using lower resolution techniques, such as CD spectroscopy[65] and FRET measurements.[66,67] This unique property of the 8-17 DNAzyme signified that the available crystallographic data do not provide information regarding the active fold important for the activity of this system in cellular conditions, as the "compact" structure induced by $Mg^{2+}$ ions remained unsolved. The **H4** paper presents the bulk of results the I obtained throughout the mentioned SONATA 14 project. Information gathered during the project led to new research avenues and in consequence to the expansion of the 8-17 DNAzyme study significantly beyond the scope originally planned in the project. Apart from the high-resolution structure determination in the presence of $Zn^{2+}$, I also conducted NMR studies in the presence of a series of other metal ions: $Pb^{2+}$, $Mg^{2+}$ and $Na^{+}$. Additionally, the system was also once more studied using other techniques applied before in previous studies. A combined analysis of all these results allowed me to propose a new model for 8-17 DNAzyme interactions with metal ions. For this reason, I decided to include them all in a single publication in a high-profile journal (paper **H4**), even though their volume and importance might have allowed them to become the basis for an entire series of publications.

***8-17 DNAzyme structure determination in presence of $Zn^{2+}$***

The research presented in paper **H4** is mostly based upon an 8-17 DNAzyme sequential variant optimized for structural studies by NMR methods. The basis for this optimization was an observation that the spatial structure of the DNAzyme's catalytic domain (red/pink in Figure 2.3.2a) should be independent from the sequence of the DNAzyme arms that bind the RNA strand cleaved by the system (green in Figure 2.3.2a). For this reason, I embarked on a search for 8-17 DNAzyme sequential variants containing shortened arms, yet still being capable of quantitatively binding the RNA substrate in NMR conditions. I decided to compensate the decreased affinity of the DNAzyme and substrate strands, related to the reduced number of base-pairing interactions between them, by significantly increasing the GC-content in the DNAzme's arms region. When it turned out that this simple adjustment was not enough to preserve the propensity of the two strands to hybridize into the active complex, I decided to covalently link them into a single construct using a stable trinucleotide loop (Figure 2.3.2b). The single molecule character of the new construct significantly reduced the entropic penalty related to the formation of the desired structure, however in conjunction with a very high GC-content, led to a high probability of alternative, non-native base pairing arrangements. For this reason, I designed a broad series of 8-17 DNAzyme constructs following the outlined blueprint. In this process I have employed the UnaFold[68] software to quicky estimate *in silico* the propensity of each construct to assume he classic 8-17 DNAzyme secondary structure as well as alternative base pairing schemes. A series of the most promising constructs was synthesized and their conformational

properties in the presence of $Zn^{2+}$ were experimentally characterized using NMR spectroscopy. For further structural studies I selected a construct showing only a single spectral form in the presence of $Zn^{2+}$ and having the highest thermal stability, denoted from now on as "8-17_short" (Figure 2.3.2b).

**Figure 2.3.1.** Models describing the influence of metal ions on the structure and activity of 8-17 DNAzyme A) the model proposed in the previous studies, B) model proposed in paper **H4**. The figure is a compilation of elements of Figures 1 and 8 from paper **H4**.




Even for the obtained 35 nucleotide-long construct the feasibility of NMR resonance assignment without isotope enrichment was not obvious, due to a high level of spectral crowding. However, thanks to the mentioned relatively high thermal stability of "8-17_short" I was able to record 2D NMR spectra in a broad range of temperatures, which allowed for the assignment of even those among the spectral signals whole frequencies were severely overlapped in some of the studied temperature conditions. The collected set of NMR restraints turned out to be sufficient to determine the spatial structure of "8-17_short" in the presence of $Zn^{2+}$ (Figure 2.3.3a), which I deposited to the PDB database under the access code 8OR8.

**Figure 2.3.2** The 8-17 DNAzyme secondary structure. A) A standard construct B) construct optimized for NMR structural studies – "8-17_short". The catalytic domain is marked in red/pink, while the substrate-binding arms are shown in green. The figure contains elements of Figure 1 from paper **H4**.

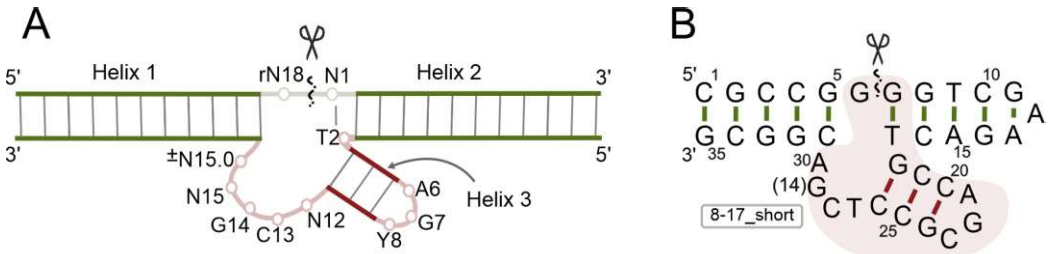
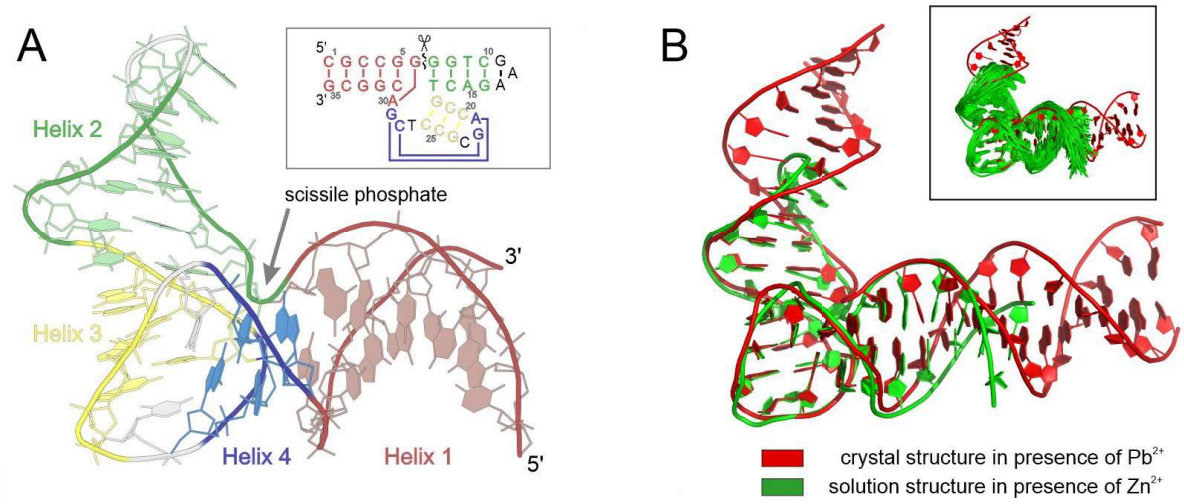



**Figure 2.3.3** The spatial structure of 8-17 DNAzyme. A) the structure of "8-17_short" determined in the presence of $Zn^{2+}$ B) a comparison of structures determined in the presence of $Pb^{2+}$ and $Zn^{2+}$. The figure contains elements of Figures 3 and 4 from paper **H4**.




### *8-17 DNAzyme NMR structural studies in the presence of other metal ions*

The determined 8-17 DNAzyme structure in the presence of $Zn^{2+}$ became a source of multiple new research questions, as despite the literature consensus[64] it turned out to be practically identical to the structure assumed by this system in the presence of $Pb^{2+}$ that was solved previously using crystallographic methods (Figure 2.3.3b). To explain this unexpected result, I decided to study the 8-17 DNAzyme in solution also in the presence of $Pb^{2+}$ as well as $Mg^{2+}$ and $Na^{+}$, both of which were supposed to induce the "compact" active structure postulated before. Titrating "8-17_short" with $Pb^{2+}$, $Mg^{2+}$ and $Na^{+}$ ions allowed me to observe a structural transition from a series of coexisting spectral forms in the absence of the metal to a single well-defined structural form. NMR data analysis allowed me to demonstrate that in each case the addition of the metal ion leads to the formation of the same fold that I observed in the presence of $Zn^{2+}$. These results show that $Pb^{2+}$ ions also induce the structural transition leading to the formation of the compact active structure. Several questions remained open however, such as why this effect was never observed in the previous studies or why the

crystallographic structure of the apo form of the enzyme also revealed this fold that in the NMR experiments was only induced by metal ion titration. Analyzing once more that available literature data regarding the interactions between the 8-17 DNAzyme and metal ions I noticed a glaring inconsistency between DNA-metal complex dissociation constants ($K_d$) obtained in experiments observing structural changes induced by metal ion binding (such as CD spectroscopy or FRET experiments) [65–67] and those determined through the measurement of the DNAzyme's catalytic activity in the function of metal ion concentration.[66,69] This observation suggested that the folding and activation of the 8-17 DNAzyme is related with metal binding to two different sites with significantly different activities. To verify this hypothesis, it was necessary to determine the affinity of "8-17_short" for each studied cofactor using two different methods: one based on the observation of the structural transition and other based on activity measurements. I also decided to repeat the same experiments for a classic 8-17 DNAzyme construct, to eliminate the possibility that the obtained results might be artifacts of the DNAzyme arm shortening. Characterization of RNA cleavage rates by two DNAzyme constructs in a broad range of concentrations of $Zn^{2+}$, $Pb^{2+}$ and $Mg^{2+}$, as well as performing CD and NMR spectroscopy-monitored titrations with these ions was a large experimental burden. It was the task of a master student Julia Wieruszewska and a PhD student Aleksandra Pawłowicz employed in my group. The obtained results confirmed the hypothesis formulated based on the literature review: for both constructs the $Zn^{2+}$ and $Mg^{2+}$ binding constants estimated using structural changes were several orders of magnitude larger than the values derived from activity measurements. This suggested that for both these ions the 8-17 DNAzyme folding occurs quantitatively already at cofactor concentrations much lower than those necessary to induce catalytic activity. In consequence, in the conditions in which the system is active, and thus those commonly used in its studies, the 8-17 DNAzyme exists exclusively in its structured, compact form. On the other hand, for $Pb^{2+}$ the binding constants derived from the two methods turned out to be much closer – consistent within the order of magnitude. This means that for this cofactor appreciable catalytic activity is present already at metal concentrations at which the active, compact form still coexists with other, inactive conformations.

***Proposing a new model for 8-17 DNAzyme interactions with metal ion cofactors***

The observed observations constitute the basis for a new model for 8-17 DNAzyme interactions with metal ions, proposed by me in paper **H4**. Instead of the previously postulated two active DNAzyme folds (Figure 2.3.1a), it assumes the existence of only one such structure (solved previously by X-ray crystallography and again in paper **H4** by NMR spectroscopy). The presence of most cofactors – like $Zn^{2+}$ or $Mg^{2+}$ – in concentrations necessary to induce catalytic activity of the system lead to the assumption of this active conformation by the entire population of DNAzyme molecules (Figure 2.3.1b – right panels), while the $Pb^{2+}$ ion is able to activate some DNAzyme molecules already in concentrations not influencing the apo conformational ensemble – by binding to

the sub-population of the properly folded molecules (Figure 2.3.1b – left panels). It is this difference that is responsible for the macroscopically distinct picture of the DNAzyme structure in the presence of $Pb^{2+}$ with respect to the other cofactors. The molecular basis for this behavior is the existence of two metal binding sites – that I denoted "structural" and "catalytic", respectively – within the active structure of the 8-17 DNAzyme and the differences in affinities of different cofactors to these two sites. In this model the apo form of the DNAzyme is a mixture of several conformational forms (Figure 2.3.1b – upper panel), including among them some fraction of the compact active structure (due to, for example, the presence of monovalent cations neutralizing the DNA negative charge). Such a picture also explains the observation of a well-folded form of the 8-17 DNAzyme in the crystals obtained without any divalent metal ions, as the result of the selection of this structural form from the solution ensemble by the forming crystal lattice.

***Identification of the $Zn^{2+}$ binding site responsible for 8-17 DNAzyme folding***

Independently from the experiments performed in the presence of $Pb^{2+}$, $Mg^{2+}$ and $Na^{+}$, I also continued the study the NMR studies of $Zn^{2+}$ interactions with "8-17_short". The goal of these additional experiments was to elucidate the geometries of the newly postulated "structural" and "catalytic" $Zn^{2+}$ binding sites. Since $Zn^{2+}$ is a diamagnetic ion, the only available method of locating its binding sites was the measurement of chemical shift perturbations (CSP) it induces for DNA atoms. However, as I already noted (Chapter 1.2), the application of this method to systems in which metal binding is accompanied by a structural transition in the DNA molecule is significantly hindered, as it is impossible to distinguish CSPs induced by the direct proximity of the bound metal ion from those caused by the structural changes alone. For the 8-17 DNAzyme I was however able to find a method to circumvent this limitation, thanks to demonstrating that $Na^{+}$ ions induce the same DNAzyme fold as $Zn^{2+}$, yet bind with an orders of magnitude lower affinity and thus can be easily replaced by added $Zn^{2+}$ without accompanying changes to the DNA structure. To locate $Zn^{2+}$ binding sites in 8-17 DNAzyme I thus applied the CSP method by first saturating the structural transition using $Na^{+}$ ions and then performing a $Zn^{2+}$ titration to the already structured DNAzyme sample. This experiment allowed me to locate the "structural" binding site in the vicinities of a G-T base pair present at the peripheries of the catalytic domain. This site was identified as the "structural" one by its characteristic high affinity. Unfortunately, the localization of the "catalytic" site using the same method turned out to be impossible, as its metal affinity is so low ($K_d > 10^{-3}$ M; based on the performed catalytic activity measurements) that a series of other weak metal binding sites started to become populated alongside it. For this reason, at the later stages of $Zn^{2+}$ titration I observed non-zero CSP values in multiple regions of the DNAzyme structure, including its substrate-binding arms, which made their interpretation in terms of binding site location uncertain.

*A summary of the 8-17 DNAzyme NMR study*

The paper **H4** apart from presenting the spatial structure of the 8-17 DNAzyme induced by $Zn^{2+}$ ions, also proposes a new interpretation of the mode in which interactions with divalent metal ions shape the conformational preferences of this system. The proposed model allows for a reinterpretation of the results obtained throughout over 20 years of 8-17 DNAzyme studies, by showing that they can be explained by the existence of only a single catalytically proficient fold. In this manner the **H4** paper also provides the first structural basis necessary to understand the mechanism of 8-17 DNAzyme's action in the presence of $Mg^{2+}$ ions that are the most important cofactor in the context of the intracellular application of this system.

# Chapter 3

# A comparative analysis of the results obtained for the three studied systems and an attempt at a synthetic summary of the impact of the presented work on the research field

## 3.1 A critical methodological analysis of the presented research

*A summary of the methods used to characterize the metal binding sites*

The goal of NMR studies in all the papers presented in Chapter 2 was the same – to elucidate the noncanonical DNA/RNA spatial structure induced by metal ion binding and to identify and characterize the interaction site. However, in practice for each of the studied systems – $r(UGGUGGU)_4$ G-quadruplex, LnA aptamer and 8-17 DNAzyme – I used a different subset of the NMR methods described in Chapter 1.2, which translated into a very varied degree of precision with which I was able to characterize the position assumed by the metal ions in each structure. In the first part of the current Chapter I would like to take another look at each of the three systems, compare the sets of methods I used for their study and attempt a critical assessment to what degree additional experiments, not included in the respective papers might have helped to improve the level of understanding of the DNA/RNA-metal interactions occurring in each molecule.

A short summary of the information obtained about the positions of the metal binding sites in each of the studied systems presents itself as follows:

1) In the case of $r(UGGUGGU)_4$ G-quadruplex, while the spatial structure of this RNA system was determined using NMR methods with good accuracy, the entirety of information regarding the positions of the interacting $K^+$ ions was obtained using molecular dynamics simulations. This means that the determined positions assumed by the $K^+$ ions in the G-quadruplex channel and the presence of

additional binding sites in the groove might both constitute only artifacts of the applied computational methods. To reduce the chances of such an outcome I repeated the MD simulations with several different force field parameters describing water molecules and metal ions (paper **H1**), yet an experimental verification of the observed interaction details would constitute a more convincing argument.

2) On the other hand, the structure of LnA aptamer complex with a lanthanide ion constitutes an example of the determination of bound metal ion position with an accuracy level rarely achieved in NMR studies (metal ion position RMSD between the twenty conformers deposited to the PDB of 0.11 Å). This result was made possible by the application of several hundred long-range PCS restraints, each of which carried information about the position of one of the atoms belonging to the DNA molecule with respect to the metal ion. In this case further refinement of the metal ion position does not appear to be necessary.

3) The results obtained in paper **H4** regarding the interaction geometry between 8-17 DNAzyme and a $Zn^{2+}$ ion constitute in turn a rather typical example of information NMR spectroscopy can deliver in case of interactions involving diamagnetic metal ions. For this system, through the application of CSP data, I was able to identify the set of nucleotides most likely involved in the formation of the $Zn^{2+}$ binding site. However, the determination of the exact position assumed by the metal ion or the identification of the DNA functional groups forming specific interactions with it lies outside the capabilities of this method. For this reason, in paper **H4** I had to speculate about the details of the interaction site using trends typically observed in metal coordination by DNA ligands.

Was it possible to obtain further experimental data to improve the metal binding site characterization for the systems in which the exact position of the metal ion remains uncertain? If so, why have I not done so in the published papers?

***Supplementary methods – molecular dynamics simulations***

I used MD simulations as one of the methods in the papers describing the first two of the discussed systems, yet I did not apply them in the **H4** paper describing 8-17 DNAzyme interactions with $Zn^{2+}$. This may seem particularly puzzling given the fact that for this system I was able to collect a set of CSP data, that may optimally be complemented by simulation results to describe the geometric details of the occurring interactions. While gathering materials for the **H4** paper I attempted to use this method, yet it did not yield clear cut results. I performed a series of MD simulations in which the $Zn^{2+}$ ion was either initially manually put in the vicinities of the expected interaction site or randomly positioned close to the DNA surface. In each independent run however, the ion positioned itself in the first deeper local minimum it encountered (usually in the first geometry featuring a direct contact with

a DNA ligand) and remained there until the end of calculations. This means that the simulation timescales were orders of magnitude too short to achieve convergence with respect to the metal ion preferred positions. Overcoming this problem would require the use of more advanced simulation methods allowing for an improved sampling the conformational space by the metal ion. Similar analysis may however itself constitute the material for a separate publication, similarly to what was the case with the study of $Pb^{2+}$ binding to the crystallographic structure published previously.[70] For this reason, I decided to leave this subject to research groups specializing in computational methods and abandoned the use of MD simulations in paper **H4**.

***Supplementary methods – chemical shift perturbation (CSP) measurement***

On the other hand, I applied the CSP method only in the **H4** paper, while similar experiments theoretically constitute a perfect approach to experimentally verify the presence of $K^+$ binding sites in the grooves of the r(UGGUGGU)$_4$ G-quadruplex. I did not however decide to perform this kind of measurements when preparing papers **H1** and **H2** because all studied samples already contained manyfold molar excess of the metal ion with respect to RNA – the 50 mM KCl, customarily used in G-quadruplex studies, to ensure the quantitative formation of the G-quadruplex fold. This means that, even if the $K^+$ binding sites in the grooved are weak (with millimolar dissociation constants), they would still be filled with the metal ion to a significant extent already at the starting point of hypothetical further $K^+$ titration. The addition of more KCl may thus not induce significant CSPs regardless of whether these binding sites exist or not. Lowering KCl concentration during sample preparation may in turn interfere with G-quadruplex formation.

***Supplementary methods – intermolecular NOE***

To confirm the geometries of metal ion binding the G-quadruplex channel one could potentially perform additional experiments based on measurement of intermolecular NOEs, that I did not apply in any of the presented papers. They would require the use of $NH_4^+$ ions as substitutes for $K^+$ during G-quadruplex formation. The positions of these ions in the central channel can be directly assessed experimentally by NOE restraints,[71,72] similarly to what I described in Chapter 1.2 for $Co(NH_3)_6^{3+}$ ions. Similar measurements would however need to be preceded by repeated structural studies to ascertain that $NH_4^+$ induce the same G-quadruplex fold as $K^+$ (which appears plausible based on preliminary NMR measurements).[53] Another limitation of this method is the fact that interaction geometries of $NH_4^+$ and $K^+$ in the channel of even identical G-quadruplexes may still differ and thus even a precise localization of $NH_4^+$ ions does not guarantee that the obtained results would be transferable to the biologically relevant $K^+$ ions.

In theory a similar experiment may also be performed for the 8-17 DNAzyme using the $Co(NH_3)_6^{3+}$ ion, as a substitute for hydrated $Mg^{2+}$. In this case an important unknown is to what extent the ions interacting with 8-17 DNAzyme remain hydrated. If cofactor binding involves their partial dehydration, then the geometry of interactions observed for $Co(NH_3)_6^{3+}$ may not constitute a good approximation for the complexes of metal ions actually relevant for the activity of this system.

***Supplementary methods – paramagnetic effects***

The potential of applying paramagnetic NMR effects, PRE and PCS, to the study of $r(UGGUGGU)_4$ G-quadruplex and 8-17 DNAzyme interactions with metal ions is an interesting question. While there are literature examples of lanthanide ion binding both in the grooves[73] and the central channel of G-quadruplexes[74] the potential for similar interactions was not verified for $r(UGGUGGU)_4$. Similarly to the use of $NH_4^+$ ions discussed in the previous point a significant limitation of this method would be the impossibility to verify to what extent the observed interaction geometries are representative of those assumed by the $K^+$ ions.

The perspective of applying paramagnetic effects to the study of the 8-17 DNAzyme appeared more promising. Paramagnetic $Co^{2+}$ ions are among the known cofactors of this system,[64] and these ions can give rise to strong paramagnetic shifts, when binding to the biomolecule with a well-defined first coordination sphere geometry. Lanthanide binding to 8-17 DNAzyme is also confirmed.[75] However, these ions act as inhibitors of its catalytic activity and thus it remains unclear whether they interact with this system in a similar manner to other ions. Nevertheless, during the studies preceding the preparation of the **H4** paper I performed the NMR analysis of 8-17 DNAzyme interactions with both $Co^{2+}$ and several lanthanide ions. In each case however the only effect I observed was line broadening (PRE). The lack of observable paramagnetic shifts (PCS) meant that these ions while interacting with 8-17 DNAzyme are characterized by close to zero $\Delta\chi$ values and thus do not assume a single interaction geometry or remain largely hydrated. The observed line broadening was indeed the most pronounced for the residues belonging to the catalytic domain which in itself confirmed preferential paramagnetic ion interaction with this part of the system. Quantitative PRE measurement is however difficult for samples with natural isotope levels[39] which prevented me from using these effects as structural restraints. A repeated DNA synthesis with full $^{13}C$ and $^{15}N$ isotope enrichment would probably allow for PRE restraint generation from HSQC-type spectra, yet I did not attempt it due to its very high costs.

Paramagnetic broadening, in theory translatable into PRE restraints, was also present next to the PCS effect in the NMR spectra of the LnA aptamer, studied in paper **H3**. Also, in this case I did not attempt their quantitative interpretation as structural restraints, as once again samples without isotope enrichment were used. For this system however the use of additional restraints was hardly necessary, as the position of the bound metal was already very well determined using PCS restraints.

***A summary of the comparative methodological analysis***

The availability of a broad range of NMR methods sensitive to the interactions of a given biomolecule with metal ions leaves for each studied system an opportunity to plan additional experiments potentially capable to improve the description of these interactions. However, as demonstrated by the presented examples, the practical feasibility of such additional measurements, as well as the actual usefulness of the results they may bring are often impossible to accurately predict *a priori*. In the studies I conducted within the papers **H1**-**H4** the choice of the methods used was informed by the physical properties of each studied system, that determine the likelihood of successfully applying any given NMR method. This does not mean however that the methods used in the published works exhaust the full range of information obtainable by NMR spectroscopy for the studied systems, as discussed in this Chapter.

## 3.2 Impact of the published results on the research field

The systems that I studied belong to two types of biologically important noncanonical nucleic acids architectures: G-quadruplexes and so called "multiple way junctions", that is structures composed of a central loop to which a number of helical elements is attached.

In the last decades G-quadruplexes have become arguably the most studied type of noncanonical nucleic acid structures, due to both the important roles they were shown to play in living organisms[43] and due to their overwhelming structural diversity which allows for new G-quadruplex folds to still be uncovered, despite hundreds of their high-resolution structures being already known. Such a newly discovered G-quadruplex structural element is for example the "reversed U-tetrad" motif, whose identification and characterization are the subject of papers **H1** and **H2**. This element is a building block available only to RNA G-quadruplexes, which makes it rather unique in the general context of G-quadruplex conformational variability, as there exist many structural motifs observed only in DNA G-quadruplexes,[47,48] but the opposite situation is rarely encountered. On the other hand, this motif can only form in the structures of tetramolecular G-quadruplexes, whose biological relevance is still debatable. For this reason, apart from a structural curiosity, I see it rather as an element potentially useful in the context of nucleic acid nanotechnology, a research field concerned with the design of DNA/RNA systems that thanks to correctly tuned structural properties are capable of carrying out specific technological tasks. The "reversed U-tetrad motif" due to its capability to strongly stabilize the architecture of a tetramolecular RNA G-quadruplex, could be used for example to promote the self-association of four RNA strands in response to a stimulus consisting of an increased $K^+$ or $Na^+$ concentration in solution. Enabling the system to use this mechanism would only involve the addition of a GGU sequence at the 3'-end of the target RNA, as this is the minimal sequential motif shared by the systems, for which the "reversed U-tetrad" was observed to date.

The other two studied systems, while sharing a multiple way junction type architecture, should be considered separately in terms of the effects the determination of their structures had on the research discipline I practice. My results related to the effects that divalent ions have on the structure of the 8-17 DNAzyme allowed to reinterpret the results obtained throughout over 20 years of structural studies of this system[64] and potentially opened a path towards a deeper understanding of its mechanism of action in the presence of its biologically relevant cofactor $Mg^{2+}$ and perhaps even towards rational design of its variants more active in cellular conditions. The demonstration that the 8-17 DNAzyme assumes only one active structure regardless of the available cofactor shed new light on the structural data available to date and showed that both the new NMR structure solved in the presence of $Zn^{2+}$ and the previously available crystallographic structure determined with a bound $Pb^{2+}$ ion can be directly used as starting points to elucidate the molecular mode of action of this system also in presence of other cofactors – such as the biologically relevant $Mg^{2+}$. The importance of results presented in paper **H4** in well-illustrated by the fact, that this work became the subject of a Comment also published in Nature Communications[76] and that in less than a year since its came out it was cited over ten times. What however, remains a significant unknown is the geometry of the catalytically relevant binding site for ions such as $Zn^{2+}$ or $Mg^{2+}$. Assuming however that it is not fundamentally different from the geometries observed for $Pb^{2+}$ in the crystal and MD simulations,[70] then the available structural data are sufficient to enable further studies aimed at the improvement of 8-17 DNAzyme's $Mg^{2+}$ binding capabilities through rational, structure-based, introduction of chemically modified nucleotide building blocks into its sequence. Similar studies are currently underway in my research group, as a direct continuation of the structural studies described in paper **H4**.

The published results are also of significant importance in the context of DNAzyme structural studies in general, as thus far high-resolution structures of similar systems remain very scarce. Apart from the crystal structure of DNAzymes 9DB1[18] and 8-17[19] and the NMR structure of the latter system, described in paper **H4** only one other high-resolution structure of a DNAzyme is available in the literature –the structure of 10-23 DNAzyme solved using NMR methods.[20] Even this limited assortment of structural data have already dramatically changed our perspective on this kind of systems. Before 2016 (when the first high-resolution DNAzyme structure was published) multiple failed attempts at DNAzyme structural characterization led to the rise of supposition that these systems probably do not form stable tertiary structures and even that the presence of substantial single-stranded regions may be pivotal for their mechanisms of action.[60] The currently available set of spatial structures have fundamentally inverted this paradigm, as all of them revealed complex, rigid DNAzyme tertiary architectures. May however this new view be equally misleading? Those among DNAzyme systems that may potentially not assume rigid folds would naturally be less susceptible to studies using typical structural biology methods, as this property would make them on one hand more difficult to crystallize and on the other to prove large scale conformational disorder in solution the use

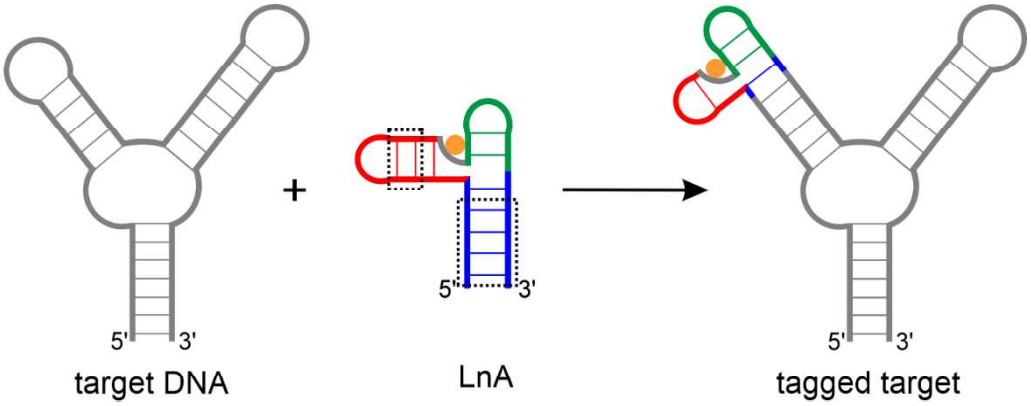
of long-range NMR restraints may be necessary, which are not easily obtained for DNA/RNA molecules. This problem is actually relevant to all DNA and RNA structures based on "multiple way junction" type architectures. Similar elements constitute a common building block of secondary structures assumed by the most biologically relevant, complex RNA molecules, such as mRNA, lncRNA ("long non-coding RNA") or genomic RNA of virial pathogens.[77–81] The scarcity of methods to assess which among the present "multiple way junctions" remain conformationally labile and which form stable tertiary structure, as well as how these preferences depend on external factors (the presence of protein partners, metal ions etc.) remains in my opinion one among the major obstacles in the prediction of tertiary structures of complex RNA molecules.

In this context the results I reported in paper **H3** have a very significant, not yet realized, potential to provide the tools necessary. The identification of a DNA structural element, capable of inducing strong paramagnetic shifts in the NMR spectra, translatable into high-quality long-range PCS structural restraints, opens the perspective of using these valuable effects in structural studies of a broad range of other DNA and RNA systems. The requirement for such studies would be, as I proposed in paper **H3**, the isolation of the lanthanide binding site from the determined LnA aptamer structure, followed by its introduction into other DNA or RNA molecule whose structural properties we would like to study. Such a procedure would allow to measure long-range PCS restraints in many DNA/RNA systems that in normal conditions do not interact with lanthanide ions, which would greatly facilitate their structural studies using biomolecular NMR methods.

The structural features of the binding site formed by the LnA aptamer, determined in paper **H3**, make it the most suitable for introduction by substitution of apical loops originally present in the "host" molecule, as illustrated in Figure 3.2.1. This means that nucleic acid systems such as DNAzymes and biologically relevant multiple way junctions are among the spectrum of molecules into which paramagnetic sites could be introduced using this approach. It has to be noted that PCS values measured for these systems would provide crucial information about their conformational preferences regardless of whether a rigid structure is present or not. This is because in case of large scale conformational variability being present in the studied system the observed PCS – unlike other spectral properties that are related to nuclear relaxation, such as NOE or PRE – undergo a simple averaging over the ensemble of conformations assumed by the biomolecule.[82] This allows them to be used to unequivocally determine whether the structural elements present in the studied system form one rigid tertiary architecture or are they moving with respect to one another and to what extent, as was already demonstrated multiple times for protein systems.[83] The application of PCS restraints would allow to greatly broaden the capabilities of NMR to study the structures of complex RNA molecules. Pilot studies of this type are currently underway in my research group. In parallel I also carry out structural studies of a series of other DNA molecules, for which literature data shows the ability to strongly interact with lanthanide ions, yet the locations and geometries of their metal binding

sites remain unknown. The goal of these studies is to identify additional systems capable of inducing strong PCS effects, potentially with structures much different from LnA. This would allow me to compose a library of paramagnetic tags, that would be possible to introduce to an even greater variety of target DNA and RNA molecules.

**Figure 3.2.1** A general scheme of paramagnetic tagging using a LnA-derived lanthanide binding site. Black dotted frames – residues from LnA_28 that can likely be omitted in this process. The figure is a copy of Figure 9 from paper **H3**.




In summary, the studies I performed allowed for the identification of a new structural element in RNA G-quadruplexes, deciphering the structure of a lanthanide binding site to a DNA molecule, characterized by a rarely encounter interaction affinity and metal specificity, as well as for the elucidation of key structural information necessary to advance mechanistic studies of the 8-17 DNAzyme. They thus constitute, in my opinion, a notable contribution to the understanding of the properties of nucleic acid structures induced by specific metal ion binding. Moreover, thanks to demonstrating that lanthanide ions bound to LnA give rise to strong PCS in the NMR spectra my research opens the potential of a broader use of paramagnetic NMR effects in nucleic acid structural biology.

References

1. Schnabl, J. & Sigel, R. K. Controlling ribozyme activity by metal ions. *Curr. Opin. Chem. Biol.* **14**, 269–275 (2010).

2. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **83**, 813–841 (2014).

3. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335–343 (2004).

4. Zheng, H., Shabalin, I. G., Handing, K. B., Bujnicki, J. M. & Minor, W. Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection. *Nucleic Acids Res.* **43**, 3789–3801 (2015).

5. Sigel, R. K. O. & Pyle, A. M. Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem. Rev.* **107**, 97–113 (2007).

6. François, J. M., Gerday, C., Prendergast, F. G. & Potter, J. D. Determination of the $Ca^{2+}$ and $Mg^{2+}$ affinity constants of troponin C from eel skeletal muscle and positioning of the single tryptophan in the primary structure. *J. Muscle Res. Cell Motil.* **14**, 585–593 (1993).

7. Crow, J. P., Sampson, J. B., Zhuang, Y., Thompson, J. A. & Beckman, J. S. Decreased Zinc Affinity of Amyotrophic Lateral Sclerosis-Associated Superoxide Dismutase Mutants Leads to Enhanced Catalysis of Tyrosine Nitration by Peroxynitrite. *J. Neurochem.* **69**, 1936–1944 (1997).

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

9. Koizumi, M. & Ohtsuka, E. Effects of phosphorothioate and 2-amino groups in hammerhead ribozymes on cleavage rates and $Mg^{2+}$ binding. *Biochemistry* **30**, 5145–5150 (1991).

10. Chowrira, B. M., Berzal-Herranz, A., Keller, C. F. & Burke, J. M. Four ribose 2'-hydroxyl groups essential for catalytic function of the hairpin ribozyme. *J. Biol. Chem.* **268**, 19458–19462 (1993).

11. Erat, M. C. & Sigel, R. K. O. Determination of the Intrinsic Affinities of Multiple Site-Specific $Mg^{2+}$ Ions Coordinated to Domain 6 of a Group II Intron Ribozyme. *Inorg. Chem.* **46**, 11224–11234 (2007).

12. Pan, T., Long, D. M. & Uhlenbeck, O. C. Divalent Metal Ions in RNA Folding and Catalysis. *Cold Spring Harb. Monogr. Arch.* **24**, 271–302 (1993).

13. Qu, H. *et al.* Rapid and label-free strategy to isolate aptamers for metal ions. *ACS Nano* **10**, 7558–7565 (2016).

14. Hofmann, H. P., Limmer, S., Hornung, V. & Sprinzl, M. Ni2+-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA* **3**, 1289–1300 (1997).

15. Rajendran, M. & Ellington, A. D. Selection of fluorescent aptamer beacons that light up in the presence of zinc. *Anal. Bioanal. Chem.* **390**, 1067–1075 (2007).

16. Johnson-Buck, A. E., McDowell, S. E. & Walter, N. G. Metal Ions: Supporting Actors in the Playbook of Small Ribozymes. *Struct. Catal. Roles Met. Ions RNA* 175–196 (2011).

17. Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M. & Scott, W. G. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.* **5**, 587–595 (1998).

18. Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Höbartner, C. & Pena, V. Crystal structure of a DNA catalyst. *Nature* **529**, 231–234 (2016).

19. Liu, H. *et al.* Crystal structure of an RNA-cleaving DNAzyme. *Nat. Commun.* **8**, 2006 (2017).

20. Borggräfe, J. *et al.* Time-resolved structural analysis of an RNA-cleaving DNA catalyst. *Nature* **601**, 144–149 (2021).

21. Bhattacharyya, D., Arachchilage, G. M. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, 38 (2016).

22. Miyake, Y. *et al.* MercuryII-mediated formation of thymine-HgII-thymine base pairs in DNA duplexes. *J. Am. Chem. Soc.* **128**, 2172–2173 (2006).

23. Ono, A. *et al.* Specific interactions between silver(I) ions and cytosine–cytosine pairs in DNA duplexes. *Chem. Commun.* 4825–4827 (2008).

24. Huang, P.-J. J., Vazin, M., Lin, J. J., Pautler, R. & Liu, J. Distinction of individual lanthanide ions with a DNAzyme beacon array. *ACS Sensors* **1**, 732–738 (2016).

25. Feig, A. L., Scott, W. G. & Uhlenbeck, O. C. Inhibition of the hammerhead ribozyme cleavage reaction by site-specific binding of Tb(III). *Science* **279**, 81–84 (1998).

26. Edogun, O., Nguyen, N. H. & Halim, M. Fluorescent single-stranded DNA-based assay for detecting unchelated gadolinium(III) ions in aqueous solution. *Anal. Bioanal. Chem.* **408**, 4121–4131 (2016).

27. Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 287–387 (1998).

28. Otting, G. Protein NMR using paramagnetic ions. *Annu. Rev. Biophys.* **39**, 387–405 (2010).

29. Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat. Struct. Biol.* **4**, 443–449 (1997).

30. Chen, K. & Tjandra, N. The use of residual dipolar coupling in studying proteins by NMR. *Top. Curr. Chem.* **326**, 47–67 (2012).

31. Miao, Q. *et al.* Paramagnetic Chemical Probes for Studying Biological Macromolecules. *Chem. Rev.* **122**, 9571–9642 (2021).

32. Täubert, S. *et al.* Lanthanide Tagging of Oligonucleotides to Nucleobase for Paramagnetic NMR. *ChemBioChem* **21**, 3333–3337 (2020).

33. Wu, Z. *et al.* New lanthanide tag for the generation of pseudocontact shifts in DNA by site-specific ligation to a phosphorothioate group. *Bioconjug. Chem.* **28**, 1741–1748 (2017).

34. Shelke, S. A. & Sigurdsson, S. T. Site-Directed Spin Labelling of Nucleic Acids. *European J. Org. Chem.* **2012**, 2291–2301 (2012).

35. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).

36. Krepl, M. *et al.* An intricate balance of hydrogen bonding, ion atmosphere and dynamics facilitates a seamless uracil to cytosine substitution in the U-turn of the neomycin-sensing riboswitch. *Nucleic Acids Res.* **46**, 6528–6543 (2018).

37. Colmenarejo, G. & Tinoco, I. Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme. *J. Mol. Biol.* **290**, 119–135 (1999).

38. Marathias, V. M. *et al.* Determination of the Number and Location of the Manganese Binding Sites of DNA Quadruplexes in Solution by EPR and NMR in the Presence and Absence of Thrombin. *J. Mol. Biol.* **260**, 378–394 (1996).

39. Cetiner, E. C. *et al.* Paramagnetic-iterative relaxation matrix approach: extracting PRE-restraints from NOESY spectra for 3D structure elucidation of biomolecules. *J. Biomol. NMR* **73**, 699–712 (2019).

40. Bertini, I., Luchinat, C., Nagulapalli, M., Parigi, G. & Ravera, E. Paramagnetic relaxation enhancement for the characterization of the conformational heterogeneity in two-domain proteins. *Phys. Chem. Chem. Phys.* **14**, 9149–9156 (2012).

41. Gochin, M. A high-resolution structure of a DNA-chromomycin-co(II) complex determined from pseudocontact shifts in nuclear magnetic resonance. *Structure* **8**, 441–452 (2000).

42. Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **12**, 832–837 (2020).

43. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).

44. Webba da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. - A Eur. J.* **13**, 9738–9745 (2007).

45. Largy, E., Mergny, J. L. & Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met. Ions Life Sci.* **16**, 203–258 (2016).

46. Karsisiotis, A. I., O'Kane, C. & Webba da Silva, M. DNA quadruplex folding formalism – A tutorial on quadruplex topologies. *Methods* **64**, 28–35 (2013).

47. Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A. & Gdaniec, Z. Overview of the RNA G-quadruplex structures. *Acta Biochim. Pol.* **63**, 609–621 (2016).

48. Mendoza, O., Porrini, M., Salgado, G. F., Gabelica, V. & Mergny, J.-L. Orienting Tetramolecular G-Quadruplex Formation: The Quest for the Elusive RNA Antiparallel Quadruplex. *Chem. - A Eur. J.* **21**, 6732–6739 (2015).

49. Pan, B., Xiong, Y., Shi, K., Deng, J. & Sundaralingam, M. Crystal Structure of an RNA Purine-Rich Tetraplex Containing Adenine Tetrads: Implications for Specific Binding in RNA Tetraplexes. *Structure* **11**, 815–823 (2003).

50. Patel, P. K., Bhavesh, N. S. & Hosur, R. V. NMR Observation of a Novel C-Tetrad in the Structure of the SV40 Repeat Sequence GGGCGG. *Biochem. Biophys. Res. Commun.* **270**, 967–971 (2000).

51. Patel, P. & Hosur, R. V. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats. *Nucleic Acids Res.* **27**, 2457–2464 (1999).

52. Cheong, C. & Moore, P. B. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **31**, 8406–8414 (1992).

53. Malgowska, M. *et al.* Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. *Nucleic Acids Res.* **42**, 10196–10207 (2014).

54. Fyfe, A. C., Dunten, P. W., Martick, M. M. & Scott, W. G. Structural Variations and Solvent Structure of r(UGGGGU) Quadruplexes Stabilized by $Sr^{2+}$ Ions. *J. Mol. Biol.* **427**, 2205–2219 (2015).

55. Xu, Y., Ishizuka, T., Kimura, T. & Komiyama, M. A U-Tetrad Stabilizes Human Telomeric RNA G-Quadruplex Structure. *J. Am. Chem. Soc.* **132**, 7231–7233 (2010).

56. Zhou, J. *et al.* Unexpected Position-Dependent Effects of Ribose G-Quartets in G-Quadruplexes. *J. Am. Chem. Soc.* **139**, 7768–7779 (2017).

57. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

58. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

59. Fu, P. K.-L. & Turro, C. Energy transfer from nucleic acids to Tb(III): selective emission enhancement by single DNA mismatches. *J. Am. Chem. Soc.* **121**, 1–7 (1998).

60. Silverman, S. K. Catalytic DNA (deoxyribozymes) for synthetic applications—current abilities and future prospects. *Chem. Commun.* **30**, 3467–3485 (2008).

61. Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem. Sci.* **41**, 595–609 (2016).

62. Zhou, W., Ding, J. & Liu, J. Theranostic DNAzymes. *Theranostics* **7**, 1010–1025 (2017).

63. Wang, F., Saran, R. & Liu, J. Tandem DNAzymes for mRNA cleavage: Choice of enzyme, metal ions and the antisense effect. *Bioorg. Med. Chem. Lett.* **25**, 1460–1463 (2015).

64. Schlosser , K. & Li, Y. A Versatile Endoribonuclease Mimic Made of DNA: Characteristics and Applications of the 8-17 RNA-Cleaving DNAzyme. *ChemBioChem* **11**, 866–879 (2010).

65. Mazumdar, D. *et al.* Activity, Folding and Z-DNA Formation of the 8-17 DNAzyme in the Presence of Monovalent Ions. *J. Am. Chem. Soc.* **131**, 5506–5515 (2009).

66. Kim, H.-K. *et al.* Metal-Dependent Global Folding and Activity of the 8-17 DNAzyme Studied by Fluorescence Resonance Energy Transfer. *J. Am. Chem. Soc.* **129**, 6896–6902 (2007).

67. Kim, H. K., Rasnik, I., Liu, J., Ha, T. & Lu, Y. Dissecting metal ion–dependent folding and catalysis of a single DNAzyme. *Nat. Chem. Biol.* **3**, 763–768 (2007).

68. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

69. Brown, A. K., Li, J., Pavot, C. M.-B. & Lu, Y. A Lead-Dependent DNAzyme with a Two-Step Mechanism. *Biochemistry* **42**, 7152–7161 (2003).

70. Ekesan, Ş. & York, D. M. Dynamical ensemble of the active state and transition state mimic for the RNA-cleaving 8–17 DNAzyme in solution. *Nucleic Acids Res.* **47**, 10282–10295 (2019).

71. Winnerdy, F. R. *et al.* Unprecedented hour-long residence time of a cation in a left-handed G-

quadruplex. *Chem. Sci.* **12**, 7151–7157 (2021).

72. Trajkovski, M., Ket, P. & Plavec, J. Cation localization and movement within DNA thrombin binding aptamer in solution. *Org. Biomol. Chem.* **7**, 4677–4684 (2009).

73. Beger, R. D., Marathias, V. M., Volkman, B. F. & Bolton, P. H. Determination of internuclear angles of DNA using paramagnetic-assisted magnetic alignment. *J. Magn. Reson.* **135**, 256–259 (1998).

74. Galezowska, E., Gluszynska, A. & Juskowiak, B. Luminescence study of G-quadruplex formation in the presence of $Tb^{3+}$ ion. *J. Inorg. Biochem.* **101**, 678–685 (2007).

75. Kim, H.-K., Li, J., Nagraj, N. & Lu, Y. Probing metal binding in the 8-17 DNAzyme by Tb III luminescence spectroscopy. *Chem. - A Eur. J.* **14**, 8696–8703 (2008).

76. Schmuck, J. F., Borggräfe, J. & Etzkorn, M. The dynamic world of the 8–17 DNAzyme. *Nat. Commun. 2024 151* **15**, 1–3 (2024).

77. Cao, C. *et al.* The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.* **12**, 1–14 (2021).

78. Imai, S., Suzuki, H., Fujiyoshi, Y. & Shimada, I. Dynamically regulated two-site interaction of viral RNA to capture host translation initiation factor. *Nat. Commun.* **14**, 1–14 (2023).

79. Walter, F., Murchie, A. I. H. & Lilley, D. M. J. Folding of the four-way RNA junction of the hairpin ribozyme. *Biochemistry* **37**, 17629–17636 (1998).

80. Novikova, I. V., Hennelly, S. P., Tung, C. S. & Sanbonmatsu, K. Y. Rise of the RNA Machines: Exploring the Structure of Long Non-Coding RNAs. *J. Mol. Biol.* **425**, 3731–3746 (2013).

81. AbuQattam, A., Serrano-Quílez, J., Rodríguez-Navarro, S. & Gallego, J. An exon three-way junction structure modulates splicing and degradation of the SUS1 yeast pre-mRNA. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861**, 673–686 (2018).

82. Andrałojć, W. *et al.* Information content of long-range NMR data for the characterization of conformational heterogeneity. *J. Biomol. NMR* **62**, (2015).

83. Bertini, I. *et al.* Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* **132**, 13553–13558 (2010).

## 5. *Presentation of significant scientific or artistic activity carried out at more than one university, scientific or cultural institution, especially at foreign institutions.*

Over the span of my scientific career thus far, I have conducted research activities in one Polish and three foreign scientific institutions - Institute of Bioorganic Chemistry Polish Academy of Sciences, University of Florence (Italy), Synchrotron SOLEIL (France) and Leiden University (The Netherlands). Below I shortly summarize the subject matter and results of studies conducted in each institution:

*Institute of Bioorganic Chemistry PAS (employment 2017- )*

My scientific activity at IBCh PAS concentrates around the study of structure, function and interactions of nucleic acid molecules using NMR spectroscopy, complemented by a series of other biophysical methods. The projects I conduct touch upon a broad spectrum of DNA and RNA molecule types. Over the years I was involved in the studies of, among others, the factors determining the stability of RNA G-quadruplexes, the influence of modified nucleotide units on

the structures of RNA duplexes and hairpins or the interactions of small molecules ligands with DNA duplexes. Recently the most prominent subjects of my scientific activity include the search for and structural characterization of lanthanide-binding DNA structural motifs and high-resolution structural studies of DNAzyme systems aimed at understanding their mechanisms of action. My research activity at IBCh PAS resulted in 15 scientific papers (positions 8-22 on the publications list provided as point I.4 of the "List of scientific or artistic achievements …" form).

*University of Florence (employment 2013-2016)*

During my work at the University of Florence the main subject of my scientific activity was the application of paramagnetic NMR effects to the study of biomolecular structure and dynamics. My most significant achievement from that period, and the cornerstone of my PhD thesis, was the development of a computational method allowing for a partial reconstruction of the ensemble of conformations assumed by multidomain biological polymers from the motionally averaged NMR structural data. This method was applied, among others, to determine the preferred conformations in solution of a two-domain protein – calmodulin – and a "two-way junction" RNA element – HIV1 TAR and to identify very lowly populated transition states of a protein-protein complex. My research activity at the University of Florence resulted in 6 scientific papers (positions 2-7 on the publications list provided as point I.4 of the "List of scientific or artistic achievements …" form).

*Synchrotron SOLEIL (Master thesis internship 2013)*

During my Master thesis internship at Synchrotron SOLEIL my research work was related to an attempt to experimentally observe the, co called "rotational Doppler effect" in photoelectron spectroscopy spectra. The conducted research resulted in one publication (position 1 on the publications list provided as point I.4 of the "List of scientific or artistic achievements …" form). My main contribution to this paper was the development of a method to deconvolute photoelectron spectra of an HCl molecule into bands origination from different rotational energy states.

*Leiden University (two research stays 2014 and 2016, over 6 months in total)*

My stay at Leiden University took place within the scope of research internships planned in my PhD project (co called "secondments" in the Marie-Curie Actions ITN program). It was related to the application of the paramagnetic NMR data analysis method that I developed, to the identification of lowly populated minor states in a protein-protein complex. This project resulted in one scientific publication (positions 6 on the publications list provided as point I.4 of the "List

of scientific or artistic achievements …" form; my affiliation for this publication is the University of Florence, as this was the institution of employment during my PhD studies).

6. *Presentation of teaching and organizational achievements as well as achievements in popularization of science or art.*

*Lectures and workshops/laboratory training for university students:*

The institutions at which I was employed thus far did not require me to conduct teaching activities as a part of my employment contract. Nevertheless, over the years from my own volition I involved myself in the organization and teaching of courses related to the subject of NMR spectroscopy. These courses were conducted both as parts of teaching plans of Poznań-based higher education institutions (points 1-3 below) and specialized workshops (points 4-5 below):

1) Poznań University of Technology, lectures and laboratory training on NMR spectroscopy for the students of the "Inżynieria Farmaceutyczna" (Pharmaceutical Engineering) curriculum. Academic years 2019/2020, 2020/2021, 2021/2022, 2022/2023.

2) Adam Mickiewicz University - NanoBioMedical Center, lecture and laboratory training on NMR spectroscopy for the students of „*Środowiskowe interdyscyplinarne studia doktoranckie w zakresie nanotechnologii*" PhD program (POWR.03.02.00-00-I032/16). Academic years 2018/2019 and 2019/2020.

3) Institute of Bioorganic Chemistry PAS, lectures: **„***Modern Methods of NMR Spectroscopy"* for PhD students of Poznań Doctoral School of the Institutes of the Polish Academy of Sciences; Academic years 2020/2021 and 2024/2025

4) „AMPERE Nuclear Magnetic Resonance Summer School" (organized by Adam Mickiewicz University - NanoBioMedical Center), workshop: "*Two-dimensional NMR spectroscopy – an introduction*", Editions: 2018, 2019, 2022

5) Institute of Bioorganic Chemistry PAS (Department of Biomolecular NMR), workshop: „*Jednodniowa szkoła 2D NMR*", co-organizer of the event and lecturer, 24-26.04.2019 (three editions)

*Scientific advisor for master and PhD students:*

- Julia Wieruszewska, master thesis supervisor (currently in review): „*Wyznaczenie struktury przestrzennej DNAzymu 8-17 w celu zrozumienia jego mechanizmu działania*", Adam Mickiewicz University

- mgr Aleksandra Pawłowicz, PhD co-supervisor, 2021- , IBCh PAS

- mgr Ewa Połomska, PhD co-supervisor, 2022- , IBCh PAS

- mgr Mikołaj Podlewski, PhD co-supervisor, 2022- , IBCh PAS

The three PhD students listed above were employed at IBCh PAS entirely from the funds that I secured from NCN OPUS19 and OPUS22 grant programs.

PODPIS ZAUFANY
WITOLD JACEK
**ANDRAŁOJĆ**
06.04.2025 13:57:10 [GMT+2]
Dokument podpisany elektronicznie
podpisem zaufanym

……..……..……
(Applicant's signature)