From Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

RNA SYSTEMS FOR NMR STUDIES *IN VITRO* AND *IN VIVO*

Hannes Feyrer

Karolinska Institutet

Stockholm 2022

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN 978-91-8016-574-7

Cover illustration: **'Broccoli makes Broccoli'**. Fluorescence image of different concentrations of the Broccoli aptamer and its ligand DFHBI-1T in 0.5 µL-droplets arranged as a Broccoli floret by the Mantis pipetting robot. Image created by **Natali Papanicolaou** with assistance of Hannes Feyrer.

RNA systems for NMR studies *in vitro* and *in vivo*

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Hannes Feyrer

The thesis will be defended in public at CMB lecture hall, Berzelius väg 21, Solna, Sweden 29th of April 2022, 9:00 am.

Principal Supervisor:
Assoc. Prof. Katja Petzold
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Structural Biology

Co-supervisor(s):
Assoc. Prof. Emma R. Andersson
Karolinska Institutet
Department of Cell and Molecular Biology

Assoc. Prof. Alessandra Villa
Royal Institute of Technology
Center for High Performance Computing

Opponent:
Prof. Frédéric Allain
ETH Zürich
Department of Biochemistry

Examination Board:
Prof. Lena Mäler
Stockholm University
Department of Biochemistry and Biophysics

Prof. Björn Högberg
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Biomaterials

Prof. Neus Visa
Stockholm University
Department of Molecular Biology

“These books aren’t for reading. They’re more for writing.”

– Terry Pratchett, Small Gods.

POPULAR SCIENCE SUMMARY OF THE THESIS

Virtually every important process in our body, and hence in our cells, is performed by biological molecules such as proteins, RNA, DNA, carbohydrates or lipids. Scientists have known for over 70 years that proteins are chains of amino acids that fold into a specific structure which is necessary for the protein to perform its function and also the famous double-helix structure of DNA has been solved around that time. It was only in the last 30 years that we started to recognize the importance of such folded structures for RNA molecules. The field of RNA biology has really experienced a revolution since we have recognized that 90% of the human genome, despite not coding for proteins, is transcribed into RNA and performs a multitude of functions which turned out understanding of biology on its head.

Knowing that a functional RNA molecule has a structure inevitably creates the need to solve that structure in order to understand fundamental biological processes, but also to develop new medicines, since we have learned how many diseases are associated with RNA malfunction. To study RNA structure, we usually synthesize some of that RNA and obtain structural information with a method central to the later work of this thesis: NMR spectroscopy. Using NMR, we can obtain full structural models including all atoms, but we can also use it to obtain somewhat crude information like an overall folding pattern.

For the actual experiment, the molecules are in a tightly controlled environment, which basically only consists of water and some salts. This is arguably very different from the interior of a cell, which is full of other molecules, small and large, which can interact with the RNA. Over time, scientists have recognized that this can actually make a difference to the observed structure, even though it does not always. The fundamental physical parameters that determine whether a different environment leads to a different RNA structure are poorly understood. It is possible to observe structural features of RNA inside living cells with NMR spectroscopy (called ‘in-cell NMR’), but it is a challenging experiment. Up until now, only ~5 publications have measured RNA with in-cell NMR, and even including very similar DNA molecules it is less than 20 publications in total. The field is rather young and is still figuring out basic experimental approaches, which is where my thesis ties in.

The goal of my thesis was to further develop the methods for in-cell NMR experiments and to extend the range of possible molecules that we can study with in-cell NMR. This includes approaches to introduce RNA samples into the cells, to keep them stable in there, to keep the cells alive during the measurement and to carefully monitor all those processes. To do that, I adapted a method to have the cells circularize the RNA chain, which effectively increases their lifetime in the cells, and thus allows us to perform longer experiments. Furthermore, I developed methods that facilitate the production of RNA in large amounts and high purity, which is needed for NMR studies *in vitro* and in cells.

ABSTRACT

NMR spectroscopy is an excellent tool to study the structure-function relationship of RNA. Such measurements are usually performed *in vitro*, which requires large amounts of isotope-labeled sample in high purity and can give access to individual atoms, structure of the molecule and conformational dynamics. This is contrasted by measurements in living cells, where researchers struggle with low signal intensity, line-broadening and rapid sample degradation. In this work, we developed sample preparation methods for NMR studies to expand the range of RNA constructs that are accessible for NMR studies *in vitro* and in cells.

Firstly, we improved yield and purity of *in vitro* transcription of short RNA constructs by transcribing several repeating target sequences from a tandem template, and cleaving them to the target length with RNase H. This abolishes issues with suboptimal initiation sequences and creates higher purity due to the high sequence-specificity of RNase H guided by a chimeric oligo. We demonstrated the high yield and purity of several such RNA molecules and incorporated the protocol into a workflow for studies of conformational dynamics with relaxation dispersion NMR.

Secondly, we demonstrated the site-specific incorporation of a $^{13}C/^{15}N$-labeled adenosine into a 46 nt RNA molecule with the use of purely enzymatic methods. Such site-specific labeling is an effective approach to overcome resonance overlap in larger RNAs, which can preclude further structural and dynamics studies. We showed the facile production of such a sample and reported on a second conformation which would in a uniformly labeled sample be hidden by overlapping resonances.

Lastly, we furthered method development for in-cell NMR methods by exploring transfection strategies, cell culture methods and RNA systems. We adapted a protocol for the production of circular RNA at high concentration in HEK293T cells to generate the first in-cell NMR spectra of intracellular expressed RNAs. Furthermore, we produced the same circular RNAs by *in vitro* transcription and ligation to assess their improved stability against cellular exonucleases. As circular RNA model systems, we used the fluorescent aptamer Broccoli and a small hairpin RNA, called GUG, which proved useful for relaxation dispersion NMR measurement previously. The expression of both circular constructs at was possible at micromolar concentration in HEK283T cells and both constructs could be transcribed and circularized *in vitro*. In-cell NMR of the expressed circular RNA did however not yield detectable signals, indicating that either the intracellular concentration is too low, or the location of the expressed RNA precludes free tumbling.

LIST OF SCIENTIFIC PAPERS

I. **Feyrer H**, Munteanu R, Baronti L, Petzold K. One-pot production of RNA in high yield and purity through cleaving tandem transcripts. *Molecules*. 2020 Jan;25(5):1142.

II. **Feyrer H**, Schlagnitweit J, Petzold K. Practical Aspects of Sample Preparation and Setup of ^{1}H $R_{1\rho}$ Relaxation Dispersion Experiments of RNA. *Journal of Visualized Experiments: Jove.* 2021 Jul 9(173).

III. **Feyrer H**, Gurdap CO, Marušič M, Schlagnitweit J, Petzold K. Purely enzymatic incorporation of an isotope-labeled adenine into RNA for the study of conformational dynamics by NMR. 2022. *Submitted*

Scientific papers not included in the thesis:

IV. Karlsson H, **Feyrer H**, Baronti L, Petzold K. Production of Structured RNA Fragments by In Vitro Transcription and HPLC Purification. *Current Protocols*. 2021 Jun;1(6):e159.

CONTENTS

LIST OF ABBREVIATIONS

bp	Base pair
DFHBI	3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one
DNP	Dynamic nuclear polarization
FID	Free induction decay
FT	Fourier transformation
G4	G-quadruplex
HSQC	Heteronuclear single quantum correlation
iPPase	Inorganic pyrophosphatase
IVT	*In vitro* transcription
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Spectroscopy
nt	Nucleotides
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
racRNA	Ribozyme-assisted circular RNA
RCT	Rolling circle transcription
RNA	Ribonucleic acid
Rnl2	T4 RNA ligase 2
SNR	Signal-to-noise ratio
SOFAST	band-Selective Optimized Flip-Angle Short-Transient
ssDNA	Single-stranded DNA
dsDNA	Double-stranded DNA
T7RNAP	T7 RNA polymerase
TORNADO	Twister-optimized RNA for durable overexpression

1 INTRODUCTION

1.1 The RNA paradigm

The expert in structural biology, especially an NMR spectroscopist studying nucleic acids, may not need an introduction and motivation to the topic of this thesis. For the wider audience, of scientific background or not, I will use this chapter to provide some general context for both the field of NMR studies of RNA, and this thesis itself. In the next chapter (Background and literature review), I will then provide technical detail and the state of the art of the field.

Scientific progress is often seen as different paradigms chasing one another, with each new paradigm being a more accurate description of the subject matter at hand. This view on science was published by Thomas Kuhn in 1962[1], and since then, paradigm shifts have frequently been claimed to have occurred. The application of Kuhn's concept to molecular biology has been disputed[2], on the ground that a true Kuhnian paradigm shift requires the rejection of old models when accepting a new one, but molecular biology seems to merely expand its models as new biological entities are discovered.

Whether this is valid or not, it is undisputable that the focus of molecular biology, and thus the contemporary world view, has been on changing biological entities. With the discovery of the DNA structure[3], biology in health and disease was explained through genes and DNA. Recombinant protein production allowed entirely new ways to study the workings of life and its molecular machines, but also created new therapeutic agents from recombinant insulin to monoclonal antibodies. This is not to say that a previous focus is entirely lost, and that several such focus fields cannot exist in parallel. Genetics is still relevant, and new recombinant antibodies are still being developed.

The (or maybe better: a) current paradigm focusses on RNA. While its central role in protein synthesis (mRNA, tRNA, rRNA) has been known for a long time, it has been discoveries like microRNAs, long non-coding RNAs, circular RNAs and other regulatory RNAs, and their involvement in e.g., gene regulation and development, which has been driving the revolution of RNA.

Contrary to Kuhn's view on science, Galison claimed that technological advancement, not a new concept, is the main driving force for new scientific progress[4]. He claims that the tools preceded the ideas, as they are required for the new data to be obtained before it can be molded into a scientific model. Adopting this view, we can find many groundbreaking tools of molecular biology that allowed us to see RNA in a new light. Such methods include fluorescence microscopy, x-ray crystallography, and especially RNA sequencing and chemical probing of RNA structure. We can now see the cell as more than just the genes it contains, and the proteins it expressed, but also through the RNA that is transcribed, spliced, regulated and translated. The picture of the inner workings of a cell appears to be more complete than ever.

1.2 Societal impact of RNA

No scientific revolution ever comes without leaving a mark on the public sphere, and so does the current RNA paradigm. With the knowledge of a new biological entity and its abundant involvement in disease comes the desire to target the entity to cure the disease. Just like proteins, RNA can be both the target, the therapeutic agent or the delivery vehicle for a therapeutic agent.

For example, mRNAs and microRNAs have been targeted with antisense therapies, where short, modified RNA molecules direct the cell's regulatory machinery to degrade the target RNA. In such antisense therapies, RNA is both the target, as well as the therapeutic agent (even though chemically modified). During the Covid-19 pandemic, the newly approved mRNA vaccines could show their potential with unprecedented success, making sure that almost every person on the planet now has heard of mRNA. In an approach that reminds more of classical enzyme inhibition, RNAs are also considered as targets for small molecule drugs, so far without clinical approval except for ribosomal inhibitors. Therapeutic RNA technology is still in its infancy, but the medical approval of several examples makes it a promising strategy for diseases which seemed undruggable in the past.

1.3 Biophysics and NMR spectroscopy

Any new technology is expected to improve with increased understanding of the underlying mechanisms. For RNA, this includes especially sequencing approaches and structural analysis. A cornerstone of the RNA paradigm is the realization that part of RNA function lies in the property to adopt complex secondary and tertiary structures, which goes well beyond mere sequence information. A similar concept lies at the heart of epigenetics and the importance non-canonical DNA structures.

For example, microRNA downregulation cannot fully be explained by sequence matching alone and seems to involve well-placed mismatches between the mRNA and microRNA strands. mRNA seems to be defined mostly by its sequence, through which it codes for proteins. However, when developing mRNA vaccines (pre-Covid-19), Moderna found that more structured mRNAs have a longer half-life in cells, leading to higher expression[5]. And ultimately, sequences are molecular entities, which interact via chemical bonds that follow the laws of thermodynamics, kinetics, and quantum mechanics – the realm of physical chemistry.

Biophysics (which should more accurately be called 'biophysical chemistry') offers a whole toolbox of techniques to study chemical and physical parameters of RNA, of which NMR spectroscopy is one of the most central methods. As the go-to technique to solve molecular connectivity for small molecules (next to x-ray crystallography), solution NMR can also solve 3D structures of biological macromolecules at atomistic detail and investigate kinetic and thermodynamic properties in the dynamic ensemble. Despite drawbacks like high sample purity, low sensitivity, considerable cost and a steep learning curve, NMR has given insight into the molecular architecture of RNA like few other techniques. In other words, an NMR project is risky to begin. Once successful however, we have access to almost every atom in the

entire chain and can study their responses to changing physical and chemical changes all at once.

1.4 In-cell NMR as a new tool for RNA research

The science philosopher Hans Jonas claimed that science and technology (ger: Technik) operate in a circular dialectic[6], which means that new technologies create new tools for science to work with and problems for science to solve, and new science allows new technologies to emerge. This approach seems to be a bridge between Kuhn's and Galison's view on scientific progress, the polar opposites of progress from either ideas or tools[4]. NMR spectroscopy is certainly a tool that has shaped a lot of progress, as molecules have become accessible in a simpler way as before. This is evidenced by the fact that virtually every institute of chemistry in the developed world houses at least one spectrometer for day-to-day analysis of synthesized compounds. Similarly, the impact of high-resolution NMR spectroscopy to our understanding of protein and RNA structures, and especially methods allowing the real-time study of detailed conformational change in solution, cannot be underestimated.

A new technology has recently left a footprint in the scientific literature: In-cell NMR spectroscopy, which is the approach to perform high-resolution NMR measurements in living cells. The underlying assumption is, that the structural ensemble of biomolecules (essentially proteins, DNA and RNA) is affected to some degree by the conditions they are measured in, and the aqueous buffers used for high-resolution NMR is a poor mimic for the intracellular environment. Among pH, weak and strong interactions, it is especially the high viscosity and intracellular crowing effect that characterized the interior of the cell. These discrepancies may cause scientists to solve structures that would not actually be formed in the cell. This assumption does have some evidence in its favor, as shown by studies of intracellular structure since the 1980s. A few structures have been shown to be altered in the cell, but in some cases, researchers have found the conclusions from their *in vitro* studies being disproven. Most of these were performed on proteins, and in the question how the intracellular environment drives protein folding, the excluded volume effect has largely been rejected as the main driving force[7,8].

This thesis ties in with method development of in-cell NMR studies for RNA targets. Only few papers have been published that have successfully measured RNA with in-cell NMR. Less than 20 publications have measured DNA constructs, which are in many ways similar to RNA and have some experimental advantages. Several studies agree that solutions to the problem of telomeric G-quadruplex topology, which concluded the parallel fold based on *in vitro* studies mimicking cellular crowding are actually not formed in cells[9,10], and found that another topology is most prevalent in the cell[11–14]. Apart from this major finding, the field is struggling with basic experimental steps so far, since RNA sample production is stunningly laborious and expensive, the handling of the cells in such large amounts if troublesome and biological relevance is challenged by open questions about cell cycles, intracellular localization.

Successful in-cell NMR experiments promise an increased understanding of what forces drive molecular structure and interaction in living cells and would hopefully allow us to model them better *in vitro*. This PhD thesis attempts to push these methods forward, and even where unsuccessful, attempts to describe obstacles and challenges for others to learn from.

2 BACKGROUND AND LITERATURE REVIEW

The following chapter will introduce the technical background of the methods used, and provide an overview of the current state of the literature. NMR spectroscopy as a tool to study biomolecular structure is rather complex in theoretical background (i.e., working principles derived from quantum mechanics) and rich in application to various aspect of RNA structure and dynamics. Hence, the background will be kept relevant to the applications that have been performed in this work and are necessary to understand the articles and preliminary results in chapter 5 (Results and discussion) and thus a lot of the mechanistic description of NMR spectroscopy will be omitted. For the methods which have been developed further in this thesis, namely *in vitro* transcription and in-cell NMR, a more thorough account of the current literature is provided, to put presented developments in perspective.

2.1 NMR studies of RNA

2.1.1 Principles of NMR spectroscopy

Spectroscopy refers to methods which probe the interaction between matter and electromagnetic radiation, often in a frequency-resolved manner to determine precise energy transitions. In NMR spectroscopy, the matter-component is the atomic nucleus, the transition occurs in the magnetic property – spin – and the electromagnetic component is radiofrequency, which is where nuclear spin transitions are excitable. The exact transition frequency depends on the external magnetic field B_0, which creates a preferential alignment of the otherwise thermally disordered spins. Furthermore, NMR spectroscopy is a resonance spectroscopy, in contrast to absorption or emission spectroscopies like UV or fluorescence spectroscopy. The following NMR background is based on the textbook by Keeler[15], if not otherwise stated.

Spins in a magnetic field

For the sake of simplicity, only nuclei with spin quantum number I = ½ will be considered here, even though it applies to all nuclei of I ≠ 0. A spin with I = ½ has two energy levels, m = ½ and m = -½, also known as α or 'up' and β or 'down' respectively, which if matching the resonance condition, will lead to the spin transition. The transition energy between those levels is proportional to the external magnetic field and the gyromagnetic ratio γ of the nucleus: $\Delta E = -\gamma B_0$, known as Zeeman splitting (Figure 1). Also, electrons around the detected nucleus shield the external magnetic field, adding another term to the equation: $\Delta E_{eff} = -\gamma(1-\sigma)B_0$, where σ is the shielding factor specific for sample properties.

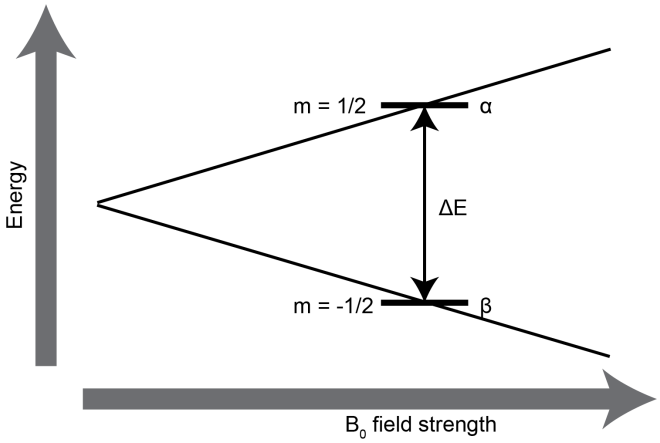

Figure 1: The Zeeman effect describes the increased energy gap between two spin states proportional to the magnetic field strength B_0.

While nuclear spins orientations at room temperature are randomly distributed, they exhibit partial alignment when placed in an external magnetic field B_0. In the magnetic field, the spins are still not perfectly oriented, but also not entirely random. A preferential alignment can be detected, called the bulk magnetization.

To describe the measurement of actual samples, the vector model is usually used, where a vector M_z is defined to represent the bulk magnetization in the sample. At equilibrium, this vector points along the z-axis, which is also the vector of the magnetic field B_0. In NMR, the excitation of nuclear spin transitions occurs via a second magnetic field B_1, which is perpendicular to B_0 and in the vector model, flips M_z to the xy-plane. Magnetization in the xy-plane corresponds to equal population of α and β spins and rotates in the xy-plane with the transition frequency corresponding to the transition energy described above ($\omega_0 = -\gamma B_0$), called the Larmor frequency.

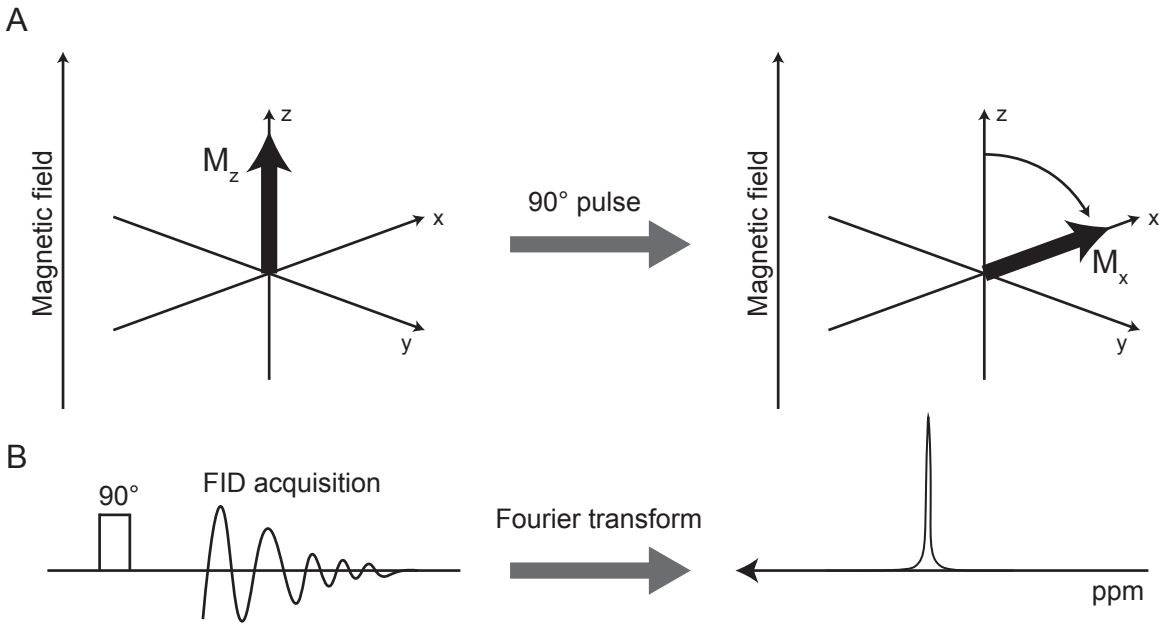

Figure 2: The basic NMR experiment. A: The bulk magnetization M_z is oriented along the magnetic field and flipped along the x-axis upon application of a 90° pulse. B: Pulse sequence of a 1D experiment. Once the magnetization is flipped along the x-axis, the FID is recorded which gives the NMR spectrum after Fourier transformation.

Technically, continuous excitation via the B_1-field leads to rotation of the M_z in the zx-plane (assuming a B_1 vector along the y-axis and following the right-hand rule). In order to flip the magnetization along the x-axis, the B_1 excitation is 'pulsed', with a duration t_p determined by $\theta = \omega_1 t_p$, where θ is the flip angle (here: 90°) and ω_1 is the B_1-field strength, written as the Rabi frequency (nutation frequency of the magnetization vector in the zx-plane).

Chemical shift and 1D NMR

For further description of the magnetization vector and its manipulation, the B_1 frequency is adopted as a rotating reference frame. A spin with Larmor frequency equal to the rotating frame frequency will not appear to rotate in the xy-plane, when flipped to the x-axis by a 90°-pulse. Different nuclei of the same isotope within the same or different molecules experience different shielding factors, leading to different effective magnetic fields. Deviation from ω_0, the 'on-

resonant' Larmor frequency, result in rotation in the xy-plane. Thus, different atoms in a molecule are discernible by the difference in rotation frequency of their nuclear magnetization vector with respect to a reference frequency. This difference in frequency is called the **chemical shift** and is comparable between machines as it is written in parts per million (ppm) relative to the spectrometer frequency. These frequencies are detected within the NMR spectrometer by a sensitive electric coil, in which the rotating magnetic moments induce a current. Due to relaxation phenomena (discussed below), these currents decay, naming the detected signal a 'free induction decay' (FID, Figure 2B).

The simplest NMR experiment (i.e., pulse sequence) thus consists of excitation of the bulk magnetization along the x-axis and detection of the FID during relaxation (Figure 2A and B). The FID contains the time information of all atoms of the measured isotope in the sample, which can be turned into frequency information by Fourier transformation (FT) thus yielding a spectrum (Figure 2B). Such a spectrum contains the referenced and normalized chemical shift in parts per million (ppm) on the x-axis, and the resonance intensity on the y-axis.

Relaxation

After the magnetization vector has been moved away from equilibrium, relaxation will occur in two pathways that are distinct and must be treated as such. The longitudinal relaxation (or spin-lattice transversal relaxation) brings the bulk magnetization back to thermodynamic equilibrium, which is alignment along the z-axis and is captured by the rate constant R_1 (or relaxation time $T_1 = 1/R_1$). In contrast, transverse relaxation (or spin-spin relaxation) describes the dephasing of the bulk magnetization, which results in effective shortening of the magnetization vector. This process is called R_2 (or $T_2 = 1/R_2$). The relaxation rates are described as exponential decays of a first order kinetic process, known as Bloch equations[16].

In practical terms, relaxation is relevant for two reasons. Firstly, the T_1 relaxation determines how long a user has to wait between individual recordings of the NMR experiment. For biological application, NMR experiments are recorded with dozens to thousands of identical scans to also detect resonances of low signal-to-noise ratio (SNR). Long T_1 times, as found in large molecules, demand longer waiting time between experiments, effectively increasing measurement times (Figure 3). Pulse sequences to reduce the time delay between scans have been developed, such as the ASAP, SOFAST or paramagnetic approaches[18–20]. Secondly,

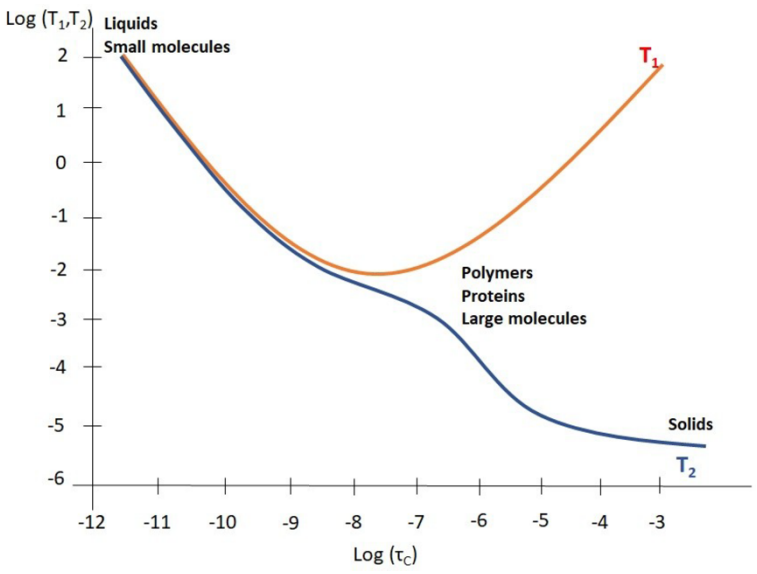

Figure 3: Dependence of T_1 and T_2 relaxation times on the molecular correlation time τ_c. Reprinted from Besghini et al.[17]

transverse relaxation times T_2 are inversely correlated with correlation time τ_c (and thus molecular weight) as shown in Figure 3[17]. A rapid exponential decay is Fourier-transformed to a Lorentzian line with an increased line-width, compared to a slower-decaying exponential function. In first approximation, the full width at half maximum (FWHM) is the inverse of the transverse relaxation rate: FHWM = $1/(\pi*T_2)$. Thus, the larger the molecule is, the broader the lines and consequently the lower SNR and higher the risk of resonance overlap. For biomolecules, T_1 is expected to be in the order of seconds, while T_2 is expected to be in the order of tens of milliseconds for the ^{1}H nucleus.

2D NMR and isotopes

Sequences of pulses increase the information density of NMR experiments by transferring magnetization in different pathways between the nuclei. Most, but not all these experiments are measured as 2D experiments. It would go beyond the scope of this introductory chapter to describe the mechanistic principles of 2D NMR and magnetization transfer in detail, so the overview is limited to relevant concepts.

The 'second dimension' in 2D NMR refers to a second chemical shift information, which is encoded in the resulting signal. This can either come from a second isotope, or from the same isotope after magnetization has been manipulated, e.g., to link nuclei with their nearest neighbors in space. The first nucleus is almost always ^{1}H and lays the foundation for the vast majority of biological NMR applications (except ^{19}F and ^{31}P), due to its ubiquity, sensitivity to chemical environment and excellent gyromagnetic ratio[21]. Thus, ^{1}H-connected heteronuclei (especially ^{13}C and ^{15}N, Figure 4) are detected via ^{1}H in coherence transfer experiments, resulting in 2D spectra. The most important such experiments are called heteronuclear single-quantum coherence (HSQC) or heteronuclear multi-quantum coherence (HMQC). Such experiments can also be designed for ^{19}F-^{13}C correlation for example[22].

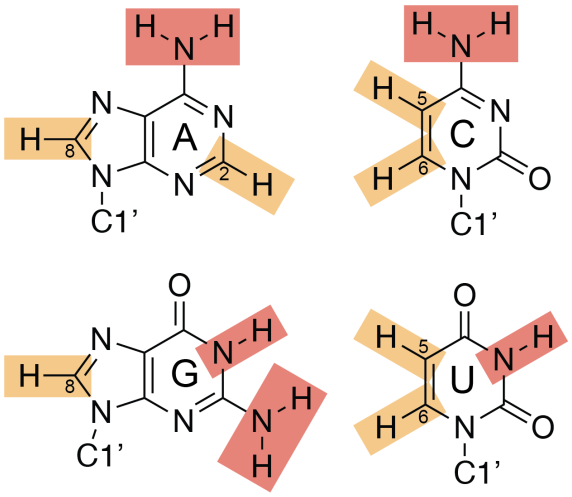

Figure 4: RNA bases and important resonances. Yellow: C-H pairs observable in ^{1}H,^{13}C-HSQC. Red: Amino and imino N-H pairs for detection with ^{1}H,^{15}N-HSQC experiments.

Magnetization transfer

To transfer magnetization in pulse sequences, two general pathways are utilized. Firstly, J-couplings (or scalar couplings) is magnetization transfer along covalent bonds. This effect also creates line-splitting, but usually removed with so-called decoupling schemes[23]. J-couplings are most commonly used to create coherences between different nuclei, e.g. ^{1}H-^{13}C, ^{1}H-^{15}N, ^{19}F-^{13}C, but also ^{1}H-^{1}H in order to identify direct covalent connections within a molecule. In contrast to scalar couplings, dipolar couplings function as electrostatic interactions through

space. They are of great relevance in the famous NOESY experiment, where the nuclear Overhauser effect (NOE) is exploited to correlate nuclei with their other nuclei within ~6 Å[24].

Technical considerations

High-resolution NMR is extremely sensitive to small changes in magnetic field in time and space. Therefore, two technical features are implemented to ensure field homogeneity. Firstly, a deuterium spectrum is constantly measured and used as internal reference to account for magnetic field drift over time. This is called 'lock' and requires a small amount of deuterated solvent (usually 5 – 10% D_2O) in the sample. Solvents can be used entirely deuterated, but this would render solvent-exchangeable hydrogens (e.g. imino protons) invisible[25]. Secondly, a set of small tunable coils, called shim-coils, around the NMR sample ensures high B_0 field homogeneity at the location of the sample, which is crucial to achieve narrow and symmetric lines in high-resolution NMR. Samples with uneven magnetic susceptibility are difficult or impossible to shim, which leads to poor line-shape.

2.1.2 Study of RNA structure

RNA bases and sequence

RNA is composed of a chain of ribonucleotides, which consists of a phosphate/ribose backbone, and a base linked to the C1' atom of the sugar (Figure 5A). For every nucleotide of a given chain, the phosphate and ribose moieties are identical and vary only in the base, of which four are available: Two purine bases, adenine (A) and guanine (G), and two smaller pyrimidine bases, uracil (U) and cytosine (C)[26]. In biosynthesis, these appear as nucleotide triphosphates, which entails the base, ribose and a triphosphate moiety at the C5' atom of the ribose. In RNA biosynthesis, the α-phosphate connects one ribose's C5' atom with the following C3' atom by forming a phosphodiester upon release of a pyrophosphate molecule[27].

The sequence of an RNA is thus the order of bases as they appear on the RNA chain. The 'first' nucleotide is obvious for a chain where a nucleotide has a free C5'-linked hydroxyle group (potentially carrying one to three phosphates), called the 5'-end or 5'-terminus. In case of an RNA segment in a longer chain or a circular RNA, no such end can be identified. Then, the chemical connectivity of the phosphates and riboses signifies the directionality, which is generally described in a 5'-3' manner (Figure 5A). Through this chemical property, the 4-letter code of RNA (and DNA) cannot just be read backwards but includes an information of direction[26]. Unless otherwise indicated, all sequences in this thesis are written in 5'-3' direction.

As classically described for DNA duplexes, the bases arise in two canonical pairs, the Watson-Crick base pairs: adenine pairs with uracil (A-U), and guanine pairs with cytosine (G-C) (Figure 5B), formed via amino and imino hydrogen bonds. In the large structural space of RNA however, many more base pair types have to be considered, which may use both the Watson-

Crick side (front), as well as the Hoogsteen side (back) of purine bases (and additionally hydrogen bonds from sugar moieties)[28].

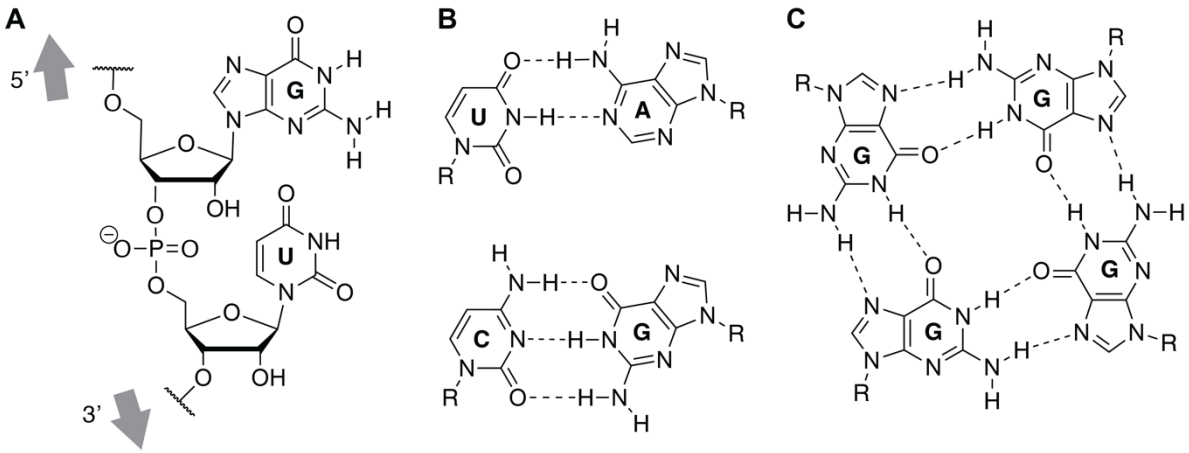

Figure 5: A: Dinucleotide showing the ribose-phosphate backbone and 5'-3' direction. B: Watson-Crick base pairs A-U and C-G including hydrogen bonds formed. C: G-quadruplex tetrad composed of four guanosines forming two hydrogen bonds to each neighbor.

Secondary and tertiary structure

RNA is usually single-stranded and forms many secondary structures through self-folding but is also found to make heteroduplexes with other RNAs and DNAs in biomolecular interactions. The helix-type formed by RNA strands is called A-form helix, which is shorter and more compact than the B-form helix found for most DNA structures. The A-form helix is characterized by a C3'-endo ribose conformation, a diameter of 2.3 nm and a stronger base pair tilt of ~20°. The double helix structure forms a shallow and a deep groove, allowing for interactions with other molecules, especially proteins[29]. Next to hydrogen bonds in base-pairing, stacking interactions between adjacent aromatic nucleobase rings form the energetic basis for the formation of a double-stranded helix. A heteroduplex between a DNA and RNA strand forms also an A-form helix. The lacking 2'-functionality of DNA allows the formation of a B-form helix, which is denied to RNA strands, due to a steric clash of the 2'-OH group and the previous phosphate moiety[26]. In contrast, certain 2'-modifications (e.g., 2'-F, 2'-OMe, locked nucleotides) can increase the interactions of the A-form helix, resulting in tighter binding of such modified RNA strands[30–32].

A helix formed by self-pairing of a single strand is called a stem-loop (or hairpin), as it requires the turn of the strand to loop back on itself (Figure 6C). The unpaired nucleotides at the apical end of that stem are called loop. When designing artificial RNA structures that should form a stable hairpin structure, the UUCG loop sequence is usually used, because of its thermodynamic stability[33,34].

A continuous, canonical A-form helix is of course contingent on the correct sequence of nucleotides (Figure 6A). Nucleotides not fulfilling the conditions for base pairing (mainly G-C, A-U, G-U) result in mismatches, but can still retain the properties of the A-form helix, as even for unpaired nucleotides the most favorable position may be stacked into the helix (Figure 6B). Unpaired nucleotides on only one of the helical stands are called single bulges or signified by the number of unpaired nucleotides (Figure 6B). Longer stretches of unpaired nucleotides

are called internal loops, classified in either symmetric or asymmetric, depending on their number of unpaired nucleotides on each strand. Large asymmetric bulges which form stem-loop structures can result in so-called junctions, which is a stem-loop forming out of a helix. Junctions can also appear when several strands of RNA come together, forming double helices in different directions (Figure 6D). A famous four-way junction can be found in the cloverleaf structure of the tRNA[36].

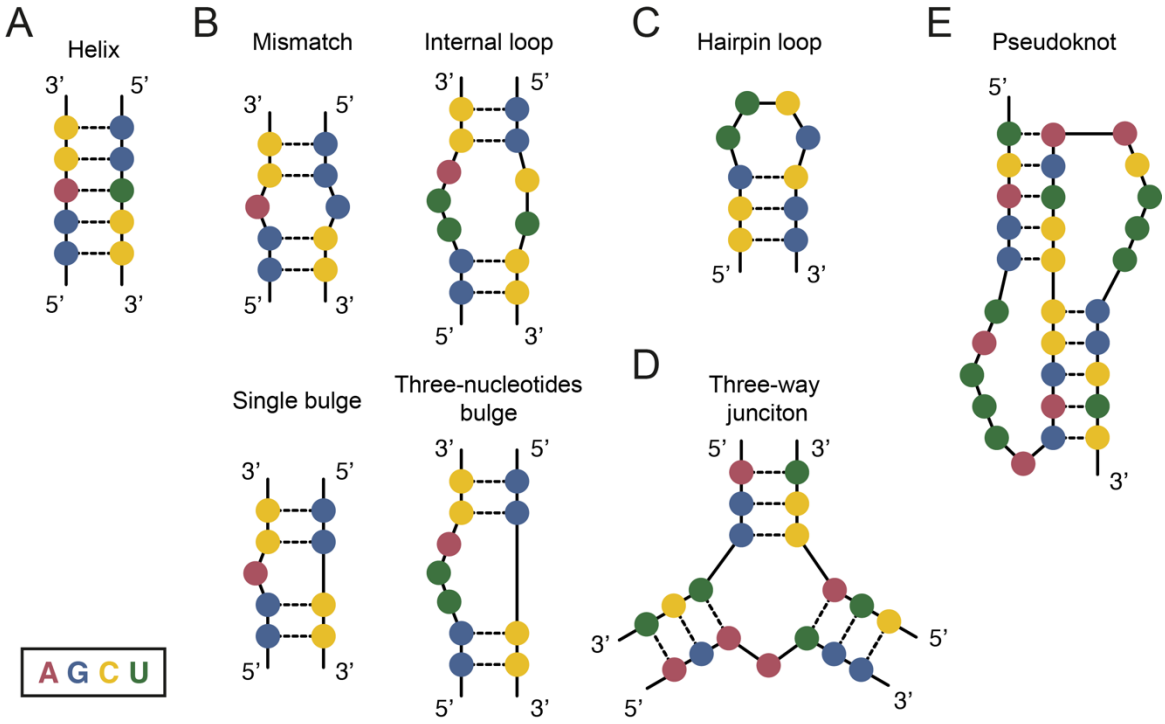

Figure 6: Secondary structures of RNA. A: Antiparallel double helix with Watson-Crick base-pairing. B: Mismatches and bulges in helical context. C: Hairpin loop, closing the end of a helix stem. D: Three-way junction, connecting 3 RNA strands. E: Pseudoknot structure with two stems in opposite directions. Figure adapted from Baronti[35].

Guanosine-rich sequences have a propensity to form so-called G-quadruplex (G4) structures, which is characterized by four planar guanosines interacting with their respective neighbors via two hydrogens on each the Watson-Crick and Hoogsteen edges[37,38] (Figure 5C). Several such guanosine tetrads stack on top of one another, which a monovalent cation, usually potassium, chelated by their O6 atoms[39]. The backbone topology of G4s can vary significantly between different sequences, but also solvent conditions. G4s are often found in DNA, especially in promoter regions and the telomers, but are also found by RNAs both in the telomeric RNA, but also a variety of cellular transcripts. Another noteworthy secondary structure is the pseudoknot fold, which is formed by a second helix with a segment further up- or downstream a first helix by forming a loop structure[40] (Figure 6E).

Next to secondary structures, RNA can form tertiary structures by the interaction of secondary structure elements to each other, or to single strands. A triplex RNA for example contains of an A-form double helix and a third pyrimidine strand pairing to the Hoogsteen side of the purine bases via minor-groove interactions. Aptamers and riboswitches form more intricate tertiary structures, which often contain stem-loops, junctions or G4s. The landscape of aptamer and riboswitch folds is extensive, so I will limit this to one aptamer which is relevant for the understanding of the Results section: The Broccoli aptamer.

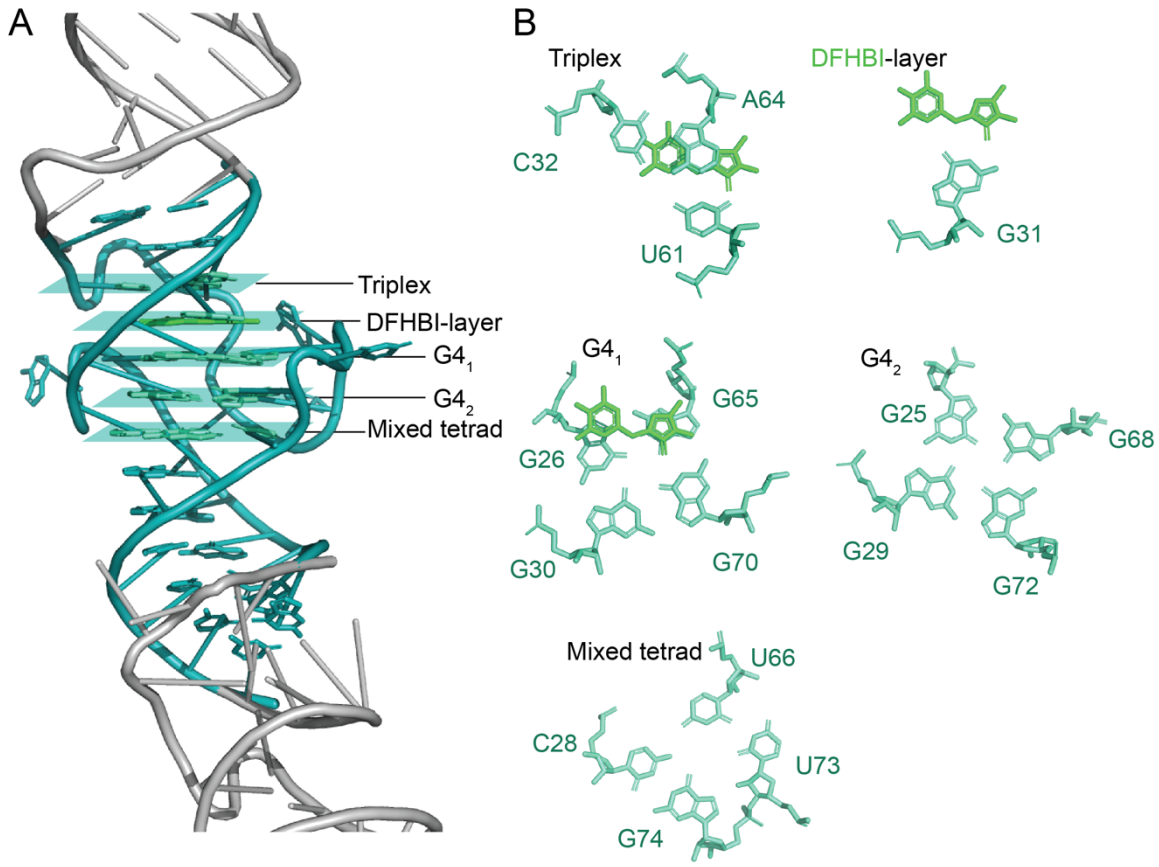

Figure 7: Tertiary structure representation and zooms of the planar core structure components of the Spinach-DFHBI structure (pdb: 4TS2). A: Spinach structure with the core structure identical to Broccoli in cyan (cartoon structure and bases as atomistic structure) and Spinach-stem loops in gray (cartoon structure). The 5 layers are highlighted and labeled, which can be seen zoomed in. B: Planar layers of the Spinach core as highlighted from panel A. DFHBI (green) is shown below the triplex layer, in hydrogen bonds with G31 and on top of the G4$_1$. Hydrogens and hydrogen bonds have been omitted in the representation. Nucleotide numbering taken from pdb 4TS2, published by Warner and colleagues[41]. The images were created with PyMOL.

The aptamer termed Broccoli is a mutated version of the Spinach aptamer, which was generated by SELEX (systematic evolution of ligands by exponential enrichment) in order to bind the synthetic analogue of the GFP chromophore DFHBI (or derivates thereof)[42,43]. Broccoli itself was selected via an advanced SELEX protocol and exhibits brighter fluorescence and stronger folding than Spinach[43]. Spinach was found to form a core structure consisting of several planes, with the ligand DFHBI stacked between a base triple and a G-quadruplex (G4$_1$), shown in Figure 7. Below, another G-quadruplex (G4$_2$) and a mixed tetrad, consisting of G, C, U and U is formed. The core structure is scaffolded by two stem loops. Compared to Spinach, or its many mutants [43,44], Broccoli only has 5 mutations, and none of them are in crucial positions of either DFHBI-binding or plane formation. Accordingly, Broccoli was simulated to retain the same overall structure as Spinach, but no structure has been solved until now. For further interpretation, Broccoli is assumed to form the same core structure[44], and thus structural interpretation are assumed to relate to both those structures. The numbering of nucleotides in the following review relates to the construct used by Warner and colleagues[41] (Figure 7) and is not directly comparable to other studies, since different stem lengths and sequences were used.

X-ray crystallography revealed that Spinach folds into a complex tertiary structure, shown in Figure 7, which contains four planes, of which two are G-quadruplexes[41,45]. These G-

quadruplexes planes were not expected from sequence alone. DFHBI binds on top of the G-quadruplex, and forms hydrogen bonds with a G28N3, G26C2'-OH and A64C2'-OH. Above DFHBI, an A-U base pair (U32-A64) forms a triplex interaction with U61, which has been called the 'lid', closing on top of DFHBI[46]. Below the two guanine tetrads, another plane is formed by C28, G74 and two uridines U66 and U73. The hydrogen bonds in the plane are the Watson-Crick base pair (C28 and G74), amino base pair between C28N3 and U66O6, imino hydrogen between U66O2 and U73N3 and a sugar base pair between U73C2'-OH and G74N7. While the backbone topology is very unlike a G-quadruplex, the bases in the mixed tetrad are all planar and they contribute to the coordination of the potassium ion via G73O4 and U73O2.

Spinach and its many derivates (Baby Spinach, iSpinach and Broccoli) bind its ligand DFHBI very tightly, at reported dissociation constants of ~290 – 1300 nM[43,47,48] and many derivates of DFHBI have been developed, which exhibit enhanced applicability for cellular imaging[49,50].

Structure prediction

Due to its rather low chemical variability (only four nucleobases), the space of possible secondary structures formed by a single RNA strand may seem rather limited. Contrary to this intuition, three different hydrogen-bonding edges have been defined for each individual nucleotide, opening a vast space for formation of complex structures[28]. In the prediction of secondary structure however, most of these interactions are not considered. Instead, these algorithms focus more on the most common forms of base-pairing and do not generally include e.g. the sugar edge. Sequence-based prediction of RNA secondary structure is rather reliable especially for duplex-based structures, but every algorithm has its own shortcomings, for example prediction of G-quadruplexes, pseudoknots, or high-energy structures, which could potentially imply conformational heterogeneity[51–53]. Nevertheless, secondary structure prediction is of utmost importance when approaching structural studies, especially following the divide-and-conquer approach with NMR spectroscopy, where a predicted structure allows a first expectation for resonance assignment and further validation.

For a complex tertiary fold like Spinach however (Figure 7), secondary structure prediction proved very unreliable and predicted a three-way junction and a large single-stranded region[42] instead of the four planes including two G-quadruplexes[41].

Important NMR experiments for RNA studies

The simplest and most standard NMR experiments is the **1D-^{1}H experiment**. In practice of biomolecular NMR, water-suppression schemes have to be applied[55,56], as the water protons (~100 M) would render the low-concentrated RNA (0.01 – 2 mM) almost undetectable. In 1D-^{1}H spectrum, resonances cluster to some extent by their chemical origin. From left to right, there are: imino resonances (from stable hydrogen bonds) at 10 – 15 ppm; amino, aromatic and H1' resonances at 5 – 10 ppm; other sugar resonances at 3 – 5 ppm (Figure 8). These might be divided further in types of base pair and specific atoms, for example the regions of pyrimidine H6 and H8 atoms are well-separated within the aromatic region[54,57,58]. 1D-^{1}H experiments can be recorded on unlabeled RNA samples within several minutes. For ^{13}C/^{15}N-labeled samples, a decoupling scheme should be employed to remove signal splitting from heteronuclear J-couplings.

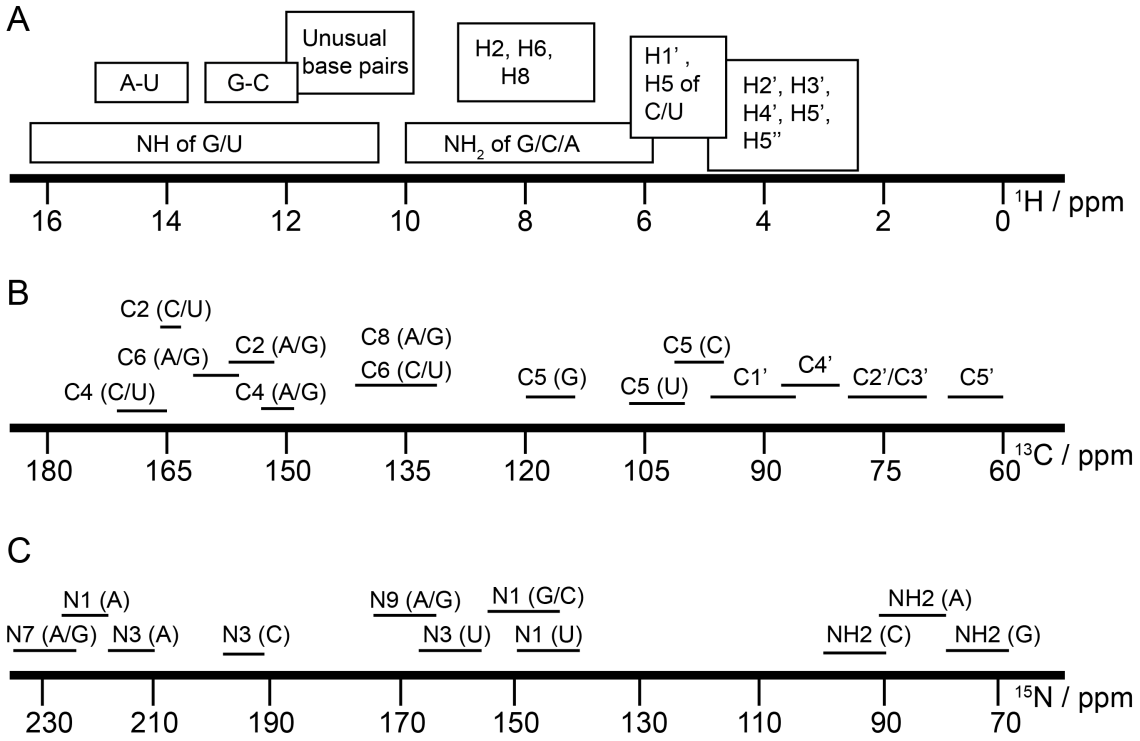

Figure 8: Range chemical shift values for ^{1}H, ^{13}C and ^{15}N for all atoms found in RNA. Redrawn from Barnwal and colleagues[54].

^{13}C/^{15}N-labeled RNA samples (from here on simply called 'labeled samples') allow the recording of **^{1}H,^{15}N- or ^{1}H,^{13}C-HSQC experiments**. Due to the large chemical shift dispersion of carbon resonances, these are either recorded as aromatic HSQC experiment, or as anomeric HSQC experiment (covering all sugar resonances), where the respective hydrogen atoms are correlated with their directly bound carbon atom. The ^{1}H,^{15}N-HSQC experiment reports on the correlation between base paired hydrogens and their directly bound nitrogen atom.

The results of 1D-^{1}H and the HSQC experiments can be very informative of the structure of a new RNA sample, given that a secondary structure prediction is available. Firstly, the chemical shift of imino resonances is to large extent determined by the nature of base pair and cluster as shown in Figure 8A[54]. This allows for a first comparison with the expected structure from a simulation to confirm homogenous folding. The ^{1}H,^{15}N-HSQC can help resolving overlapping

resonances by adding a second dimension. Secondly, chemical shift of ^{13}C resonances (especially aromatic and C1') reports in part on helicity of a resonance[59,60]. This can give further evidence if the right fold is present, give indications on initial resonance assignment and be invaluable when determining alternative confirmations (discussed in the next section).

For more accurate structural interpretation and resonance assignment, **NOESY** (Nuclear Overhauser Effect SpectroscopY) is invaluable for its ability to correlate protons not via covalent bonds, but through space[61]. This is important for resonance assignment but can also give information on the structural conformation of the RNA[58]. At this point, a similar work on protein NMR studies would explain how NOESY resonances (maybe paired with residual dipolar couplings, RDCs) can be fed into a molecular dynamics (MD) simulation in order to obtain a structural ensemble with atomistic resolution[62]. This is the main method to obtain 3D structures of proteins by NMR spectroscopy. While this is possible for RNA too, it is less common when performing structural investigations with NMR. This is in part because MD simulations for RNA are underdeveloped and cannot utilize NMR data as well as for proteins. Even though this bottleneck is actively worked on[63–65], the chemical structure of RNA and NMR properties allow researchers to tackle structural problems in a different way.

The arguably most important components in RNA structure are the nucleobases, which are well-accessible with NMR. The chemical shift dependence of the aromatic resonances and the imino protons in hydrogen bonds give invaluable information on orientation of a base in space with respect to neighboring bases. This allows to represent the bases as planes which guide the structure, and the backbone to follow (Figure 9B). Thus, it is common to represent RNA structures through its hydrogen bonds formed, and the relative orientation of the bases. While this is no atomistic description, it accurately captures the interaction bases at hydrogen-bond resolution, and allows the description of changing base pairs and their structural

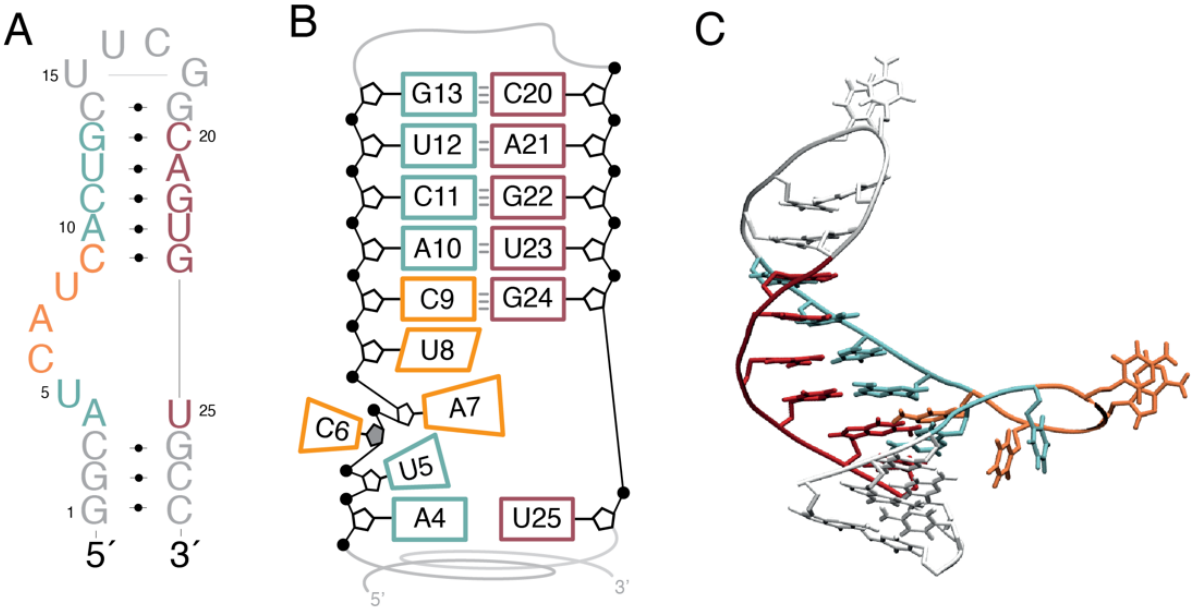

Figure 9: Different representations of the same RNA structure. A: Secondary structure prediction from Mc-Fold[51], showing the pairing of nucleotides. B: RNA structure with bases shown as planes, which includes relative orientation of the bases in the bulge. C: All-atom structure from molecular dynamics simulation shown with the backbone as a string and the orientation of bases at atomistic detail. Figures have been adapted from Baronti et al.[59]

rearrangements[59,66]. Only the full atomistic 3D model (Figure 9C) captures the full extent on all hydrogen bonds and relative orientation of bases, including sugar edge and backbone interactions. The Leontis-Westhof nomenclature is an approach to denote non-Watson-Crick interactions in a 2D sketch in order to capture all important information for a 3D interpretation[28].

Conformational dynamics by NMR

RNA molecules are highly dynamic molecules that exhibit conformational exchange across several time scales. NMR spectroscopy offers a plethora of methods to study such conformational exchange in different time regimes. Generally, such methods can be divided by the time scale they probe, which in first approximation relate to being faster, similar or slower than the relaxation parameters T_1 and T_2. Due to the energies associated with RNA dynamics, fast exchange is less frequently related to biological function and will thus be omitted in this summary. A thorough breakdown of the RNA dynamics over all time scales and discussion of relevant NMR experiments is available from Marušič and colleagues[67], and summarized in Figure 10.

Slow exchange between two alternative conformers occurs on a time scale slower than seconds and often involves large structural rearrangements. In the NMR spectrum, this is manifested as two distinct signals, one for each conformer. This can be probed by exchange spectroscopy (EXSY)[68] or hydrogen exchange experiments[69,70]. In EXSY, signals from conformational exchange appear next to NOESY cross peaks, which they can be delineated from using a

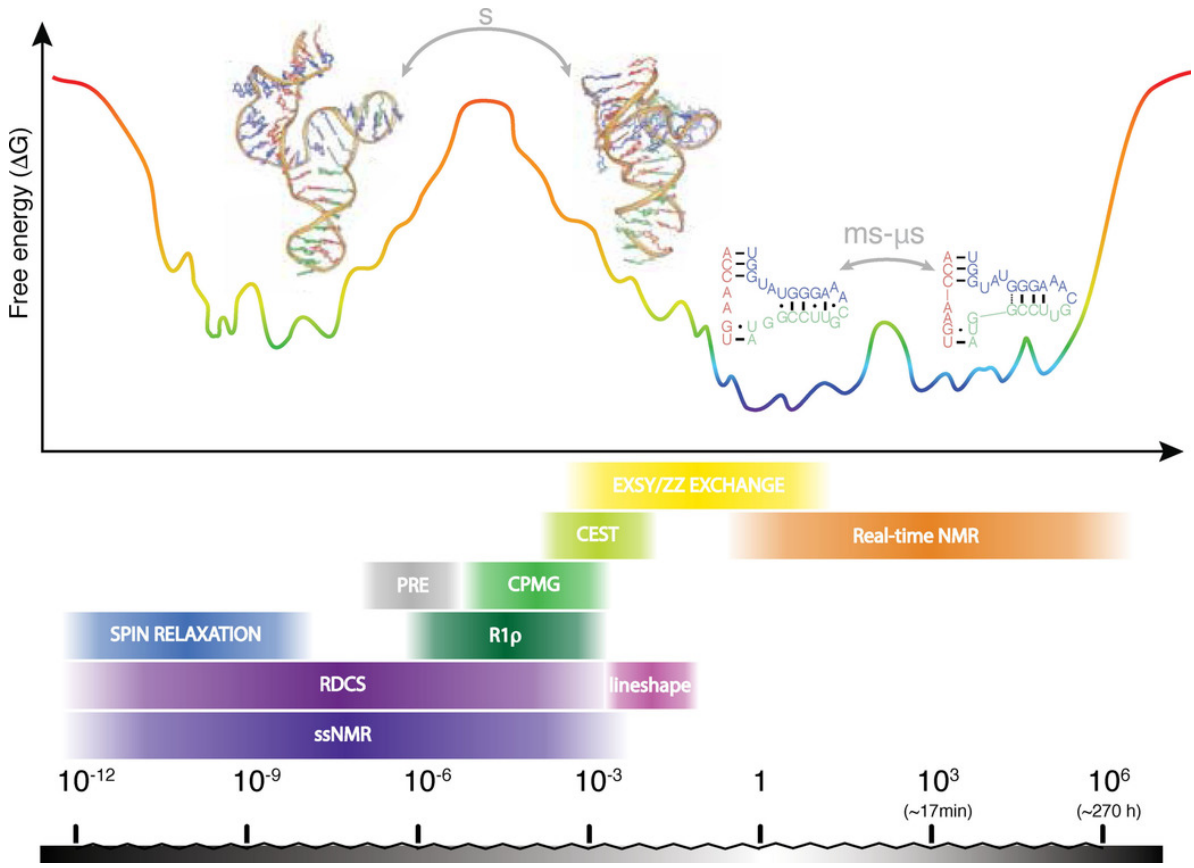

Figure 10: RNA dynamics methods over NMR-observable timescales. Reprinted from Marušič et al.[67]

ROESY (Rotating frame nuclear Overhauser Effect SpectroscopY) for comparison because the ^{1}H-^{1}H cross signals, but not exchanges signals are of opposite sign[67]. These exchange signals thus give the chemical shift for the alternative conformer, which can provide information on its structure. Hydrogen exchange methods report on the stability of hydrogen bonds by directly probing their exchange rates with bulk water. This is somewhat analogous to deuterium exchange experiments, where the signal of hydrogen-bonded protium (^{1}H) is slowly replaced by NMR-silent deuterium (^{2}H). In the CLEANEX experiment, water is selectively excited, which upon exchange to a hydrogen bonds leads to the appearance of signals in the imino region[69]. Alternatively, the water resonance can be saturated by continuous irradiation, which will lead to a decrease of imino signals[70].

Intermediate exchange is manifested in the NMR spectrum as one single line, which is shifted between the true chemical shift values of two conformers. Additionally, this line is subject to often severe broadening, up to the point where it is no longer discernible from experimental noise. The exchange rate for the intermediate regime is typically in the order of µs to ms, which corresponds roughly to the T_2 time of larger biomolecules. The difference of chemical shift of the observed signal between ground state and excited state depends to large extent on the relative populations of both conformers. Hence, if the population difference is very large, the observed line will almost match the ground state chemical shift, but the chemical shift of the excited conformer remains unknown.

Since chemical shift provides valuable information on RNA structure, NMR methods that reveal the chemical shift of the excited state are desired to probe intermediate exchange. This is the case for chemical exchange saturation transfer (CEST)[71] and relaxation dispersion (RD) methods[60,72]. For the latter, one uses the fact that different conformations have different relaxation properties and monitors the dispersion of relaxation rates in an exchanging system, while the relaxation of a non-exchanging system would remain constant. RD measurements can be executed using a longer R_2 measurement (CPMG sequence)[73] or in a spin-locked rotating frame ($R_{1\rho}$ RD)[74], which slightly deviate in experimental setup and time scales they can access. Both give access to exchange rate, relative populations and chemical shift difference. CEST functions similar to hydrogen exchange by saturating water resonances, in that the ground state resonance is selectively saturated, and the increase of a second signal appear, which readily provides the excited state chemical shift and thus allows for modelling of an excited state structure.

Sample considerations

NMR spectroscopy, especially in solution-state, is a rather insensitive method. The minimum requirement for typical RNA samples is ~250 µL at a minimal concentration of 1 – 5 µM, which equals 0.25 – 1.25 nmol. Unlike most proteins however, RNA can be concentrated to the low millimolar range without leading to precipitation or phase separation. Therefore, NMR samples for structural characterizations are usually prepared to concentrations of 0.5 – 2 mM. As most organic buffers lead to signals in ^{1}H, ^{13}C and ^{15}N spectra, phosphate buffer systems are generally used to study RNA.

Basically, all isotopes with $I \neq 0$ can be measured with NMR spectroscopy, but in the practice of biomolecular NMR, only isotopes with $I = ½$ are used. Thus, while ^{14}N does have a spin of 1 and can be measured by NMR, it is beneficial to use the ^{15}N isotope ($I = ½$) for the study of biomolecules[21]. Both ^{15}N and ^{13}C ($I = ½$) have a low natural abundance (0.3 and 1.1% respectively). This can be sufficient if (a) the sample can be concentrated enough and (b) no magnetization is transferred between heteronuclei. As both are rarely the case in biomolecular NMR, isotopic labeling is a crucial practice for the studies of RNA (as well as proteins and DNA). Since isotopic labeling is a considerable cost factors, unlabeled samples are often used to acquire measurements on only ^{1}H-based experiments, which can be enough to get preliminary data on a new system. In the case of phosphorus, ^{31}P of $I = ½$ is already at 100% natural abundance and thus no labeling is used for ^{31}P NMR studies.

Many cellular RNAs tend to be 'large' from a perspective of NMR spectroscopy (> 100 nt), since the line-width is proportional to molecular tumbling time. Several small RNAs (< 50 nt) do exist and are of high abundance, for example microRNAs, small nucleolar RNAs or small hairpin RNAs. The highly abundant tRNAs are just at the size-limit of high-resolution NMR and have been studied[75,76]. rRNA and mRNA are for most applications out of the scope of solution NMR. Next to the increased line-width, NMR spectra of large RNAs also struggle with severe resonance overlap, since the chemical shift dispersion of RNA is rather low. Many methods have been developed to overcome those issues of course. The point here is to highlight that the study of large RNA molecules come with additional considerations for NMR methods and sample production.

2.2 Sample production

2.2.1 RNA synthesis

For RNA to be produced *in vitro*, two major methods exist: *In vitro* transcription (IVT) and solid-phase oligonucleotide synthesis (SPOS). Expression and purification in a host organism, e.g. *E.coli*, is possible, but rarely used, especially for preparation of NMR samples. For all samples made for this work, I used IVT to produce RNA, and naturally, this will be focus of background and literature review here. Nevertheless, I will briefly introduce and mention certain SPOS methods, because they harbor certain unique strengths and cannot be disregarded when discussing RNA production methods. Many different applications for synthetic RNA and modified nucleic acids have been developed and are of great interest also for biophysical studies, but a general literature review would be beyond the scope of this chapter. Thus, I will focus on RNA production methods relevant for NMR studies, specifically to high yield, purity and isotopic labelling.

Solid-phase oligonucleotide synthesis (SPOS)

During SPOS, an RNA chain is built nucleotide by nucleotide on a solid support resin, which the 3'-terminal nucleotide is attached to. Each incorporation cycle of a new nucleotide consists of several steps, as the nucleotide is not provided as a regular nucleotide triphosphate (NTP), but as a chemically modified building block (usually phosphoramidite), which contains an

activatable protection group on the 5'-end (usually DMT, dimethoxytrityl), and protecting groups for 2'-OH and amino groups on the nucleobases. A thorough review of solid-phase chemistry for nucleotide synthesis can be found in the cited literature[77–80].

The unique strength of SPOS is the versatility in chemistry, which allows the production of virtually any nucleotide within the realm of feasible chemical synthesis[81,82]. This opens the possibility to incorporate modifications for fluorescent imaging[83], for structural studies[84], to increase stability inside cells[85,86] or to alter functions of therapeutic RNAs[87,88]. Of high value for NMR studies is the possibility to incorporate isotopically labeled nucleotides (usually ^{13}C and/or ^{15}N) site-specifically. This can be either done uniformly, i.e. all atoms of a specific element are isotopically labeled, or atom-specific[89–91].

SPOS has two weaknesses: Firstly, as the RNA chain is built in iterative chemical reactions, the yield drops exponentially with increasing chain length, which puts the limit of current synthesis methods to ~50 – 100 nt. Secondly, isotopically labeled and certain modified building blocks are not commercially available, and thus require manual synthesis to produce the precursor phosphoramidites. Nevertheless, SPOS is an invaluable tool to produce RNAs especially for structural studies and therapeutic application.

T7 *in vitro* transcription (IVT)

The go-to enzyme for *in vitro* transcription is the RNA polymerase of the bacteriophage T7, called T7 RNA polymerase (T7RNAP)[92]. It is also the basis for the protein expression system in *E.coli*, combined with the LacZ induction system[93,94]. Viral polymerases, such as T7, but also T3, Sp6 and Syn5[95–98], have the advantages of usually being single-subunit, having low transcriptional regulation and high processivity. This makes them valuable tools for *in vitro* reactions, as they are simple to use and generate large amounts of RNAs. Also, T7 RNA polymerase is simple to overexpress in *E.coli*[99].

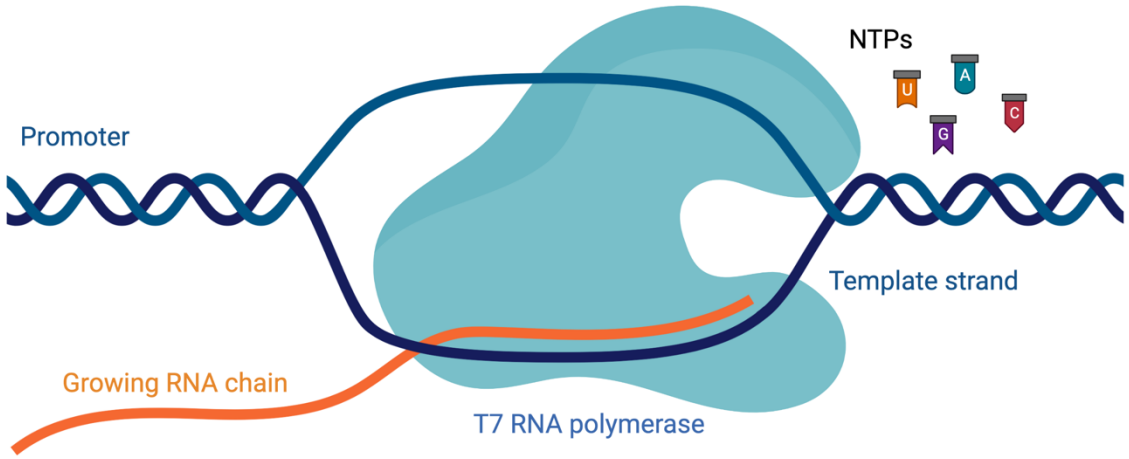

Figure 11: In vitro transcription by T7 RNA polymerase. After recognition of a specific promoter, the polymerase creates the RNA chain in the transcription bubble along the template strand. The RNA chain is built by condensation of the nucleotide triphosphates (NTPs) when they form Watson-Crick base pairs with the bases of the template strand. After the 17 nt-promoter region, the template can be either single-stranded or double-stranded. Figure created with Biorender.com

Details of a typical IVT reaction will briefly be laid out here, as they are crucial for the understanding of several choices in the work presented later. Next to the actual reacting components, nucleotide triphosphates (NTPs), the polymerase and the DNA template (Figure 11), an IVT reaction requires at minimum a buffering agent between pH 7.5 and 8 (typically TRIS or HEPES), a magnesium salt (required cofactor for T7RNAP), spermidine (a polyamine) and DTT (dithiothreitol, to control the redox state of T7RNAP). Additionally, crowders like DMSO or polyethylenglycol (PEG), or detergents like Triton-X are used[100].

Upon nucleotide condensation, a pyrophosphate moiety is released, which can co-precipitate with Mg^{2+} ions[101,102]. This precipitate can trap RNA and DNA, making them inaccessible for further processing like cleavage or ligation. The addition of inorganic phosphatase (iPPase) from *E.coli* is used to cleave pyrophosphate into phosphate ions, precluding the formation of such precipitate.

Application of T7 IVT to NMR samples

Other than SPOS, T7 IVT has no hard size limit and can generate RNA long enough to fold into nanoparticles[103,104]. In contrast to SPOS however, the starting sequence of the RNA transcription favors certain sequences and deviations results in potentially omitted 5' nucleotides[105–107] and reduced yield. These are in part determined by the viral promoter sequences found in nature[108–110], but have also been investigated further with structural and high-throughput means. It was found that the first guanosine nucleotide contacts the enzyme via its N7 atom[111]. Most recently, Conrad and colleagues screened a library of random sequences and identified the sequence GGGAAATA as the most efficient start of transcription[112]. In the practice of making short RNAs for NMR studies however, it is common practice to use 2 – 3 guanosines to ensure high-yield and homogeneous transcription, instead of a 6 – 8 nt sequence. Next to 5' heterogeneity, T7RNAP can also create heterogenous 3' termini, which have in part been attributed to self-templated extension[113].

To counteract end heterogeneity and the impact of the initiation sequence when producing NMR samples, researchers utilized different approaches to site-specifically remove 5' and 3' flanks to create a homogenous target RNA sample. These did either entail site-specific cleavage with *E.coli* RNase H[114,115] or cis- or trans-active ribozymes[115–117].

T7RNAP can use either single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) as transcription templates, as long as the 17 nt promoter consensus sequence is double-stranded[100,108]. A selection of typical templates is shown in Figure 12. PCR products or plasmids amplified in *E.coli* often serve as useful templates, because they are straight-forward to be produced in large amounts and long templates can easily be produced by primer assembly or similar gene synthesis methods. For short target sequences, templates can be synthesized by SPOS and annealed to the complementary strand of the T7 promoter sequence. This is particularly useful as SPOS is cheaper and easier for DNA than RNA, and it allows the use of 2'-OMe modifications of the last templating nucleotides, which was found to reduce 3'-heterogeneity of transcripts[118,119].

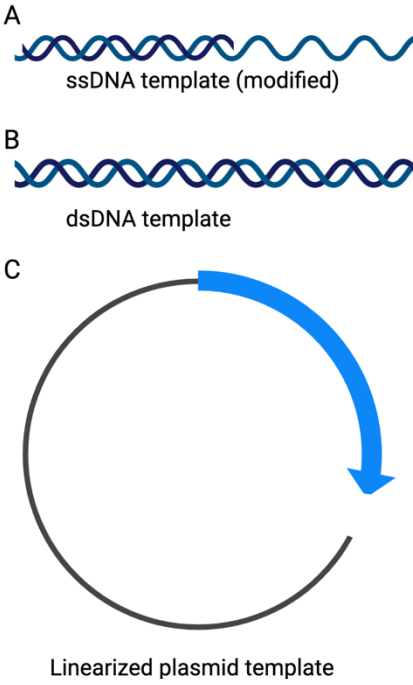

Figure 12: Different templates suitable for in vitro transcription: A: Single-stranded DNA (double-stranded in the promoter region) carrying 2'-OMe modifications at the last two nucleotides to enhance 3'-homogeneity. B: Double-stranded DNA, usually generated by PCR. C: Linearized plasmid template to ensure 3'-homogeneity through consistent polymerase run-off. Figure created with Biorender.com.

During polymerase-binding, the double helix of the template is melted, forming the so-called transcription bubble, which allows the binding of the priming nucleotide[111]. In this stage, T7RNAP is in the initiation conformation and synthesizes short transcripts (2 – 8 nt) without moving along the template[120]. These short transcripts are often aborted upon bubble collapse and ejected from the complex[121]. Only upon a change of conformation into the elongation conformation can the polymerase clear the promoter and extend the transcripts while moving along the template[122–125].

Yield of T7 IVT

Yield is expressed differently in the literature of IVT. While some articles express yield as the amount of how much the template was amplified, others express it as an amount of RNA generated, either in mass or molar amount. The latter then allows to calculate the yield as generally defined in synthetic chemistry, i.e. the stochiometric ratio of obtained product to limiting reagent. As the isotope-labeled nucleotides are the most expensive, and also the limiting reagents in an IVT, it makes sense to take this approach to quantify the yield of an IVT reaction. By the nature of the RNA being composed of 4 different reagents in different ratios, the limiting reagent is the most abundant one, given that all are present in the same concentration. For an IVT, the yield should only entail the target-length RNA molecules but ignore 3'-extended and 5'-shortened molecules.

The sequence-dependence of the initiation sequence, and continuous optimization of the IVT reaction suggest that an optimal reaction condition for all IVT reactions exist. However, strong discrepancies have been found between conditions for different types of templates, and for different RNA sequences. This knowledge is unfortunately rarely published, but instead described as an 'optimized RNA transcription', which is typically performed by varying concentrations of different reagents (Tris, Mg^{2+}, template, DMSO) in order to maximize yield. While this may be absent in certain fields, the NMR community producing isotope-labeled RNA sampled is rather concerned with not wasting labeled nucleotides in their large scale reactions. Thus, transcription conditions of new RNA sequences are typically optimized before producing isotope-labeled samples in large scale. So far, there is no model explaining these observations beyond the initiation sequence[112].

2.2.2 Isotopic labeling

NMR spectra of increasingly larger RNA molecules suffer from severe overlap, because the chemical shift dispersion is rather low. An approach to remove signal overlap are labeling schemes, which means that only selected nucleotides or atoms are labeled and thus de-crowd the spectrum. In the following section, I will introduce the general idea of different labeling schemes that can be produced with IVT.

Uniform labeling

In a regular IVT, all NTPs can be replaced with their $^{13}C/^{15}N$-labeled counterparts, which generates a fully labeled RNA sample. Often, only a specific type of nucleotide is labeled to reduce spectral crowding and allow the study of this specific type of nucleotide. As such, G/C or A/U-labeled samples are useful, because the important imino resonances arise from G and U only. Unfortunately, resonances of the same type of nucleotide tend to cluster in the same spectral region and thus local overlap is still likely.

Segmental labeling

Instead of labeling a certain species of nucleotides, an RNA segment of interested can be isotope-labeled, which allows a more detailed investigation of that section than with nucleotide-specific labeling. These methods are generally based on producing labeled and unlabeled fragments by IVT, and their successive ligation. The labeled and unlabeled fragments are often created by transcribing the full target RNA and separation into fragments by cleavage via RNase H or a ribozyme. This allowed the placement of labeled segments at termini[126,127], or in the center of a larger RNA fragment[115,128] (Figure 13A). The details of this approach have been discussed by Duss and colleagues[129].

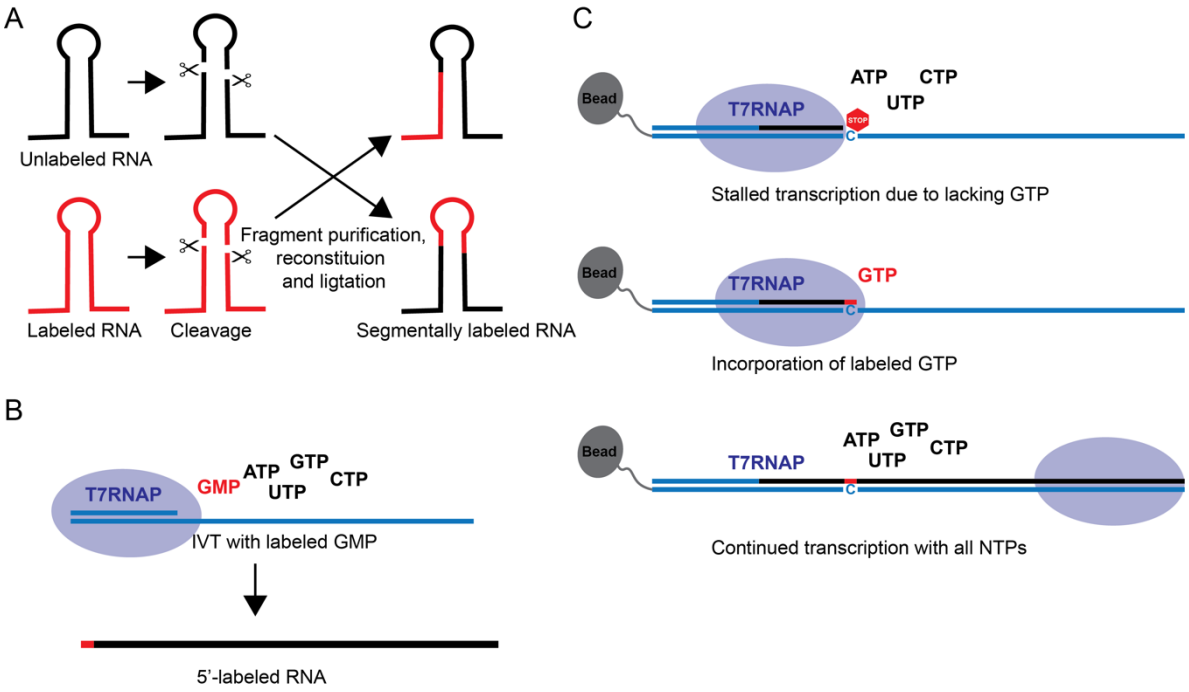

Figure 13: Isotopic labeling schemes. A: Segmental labeling by site-specific cleavage and successive ligation of isolated fragments. B: 5'-labeling of RNA by supplying labeled GMP to the IVT reaction. C: Position-selective labeling of RNA (PLOR). The IVT with 3 unlabeled NTPs is stalled when it reaches the missing nucleotide. After supplying the labeled NTP, IVT can be continued to complete the RNA construct.

Site-specific incorporation

To study only a single nucleotide, useful tools for site-specific labeling have been developed that rely on ligation, similar to segmental labeling. In those approaches, the single labeled nucleotide is attached to either the 3' or 5' end of an unlabeled strand, which is then ligated to another strand in order to place the single label in the center of a longer RNA. If the first nucleotide (usually G) is supplied as nucleotide monophosphate (GMP), it will be preferred over its triphosphate by T7RNAP. This approach was used to study only the 5' GMP of a 232 kDa fragment of the HIV genome, as they researchers supplied deuterated GTP and supplied only protonated GMP[130] (Figure 13B). Other methods for 5' labeling are the guanosine transfer reaction of group I introns[131], or at the 3' end via splinted ligation[132].

An alternative method for site-specific labeling is called PLOR (Position-selective labeling of RNA), a stop-and-go method[133]. For PLOR, the template is tethered to beads which allows rapid exchange of the IVT solution. First, IVT is performed in absence of a certain nucleotide which stalls transcription at the position of this nucleotide. Then, the solution is exchanged to supply the desired nucleotide, potentially isotope-labeled, and is then continued by supplying a new IVT buffer to continue the transcription (Figure 13C).

2.3 In-cell NMR

2.3.1 NMR studies in living cells

High-resolution structures of molecules, including proteins and RNA, are generally solved *in vitro*, i.e. outside of living cells with either X-ray crystallography, cryogenic electron microscopy or NMR spectroscopy. Among these methods, solution NMR has the advantage to easily access the conformational ensemble and taking it into account when modelling 3D structures. Nevertheless, the conditions that NMR structures are solved are undoubtedly very different from the inside of living cells. Most importantly, the interior of a cell is extremely dense and viscous and can have protein concentrations of 400 g/L, but may also vary in pH, salt content and concentration and is generally a spatially heterogenous environment which can change a macromolecular structure to be different from how it is found in a test tube.

Already in the 1980s, the physicochemical parameters that guide protein folding in living cells were investigated. Macromolecular crowding and excluded volume effect were identified as driving forces for protein folding[134–138]. The model of excluded volume describes cellular proteins as inert spheres, which drive a protein into its folded state while the unfolded state is destabilized. Chemical and electrostatic interactions are not considered here. To mimic molecular crowding *in vitro*, artificial crowding agents (PEG, DMSO, Ficoll, glycerol) or protein crowders (often bovine serum albumin) are used. However, the changes in protein stability when using crowders could not explain the destabilization of proteins inside living cells, which invalidates the model of macromolecular crowding as driving force for intracellular protein folding to some extent, and hints at other factors that are at play[7,137,139,140]. This highlights the importance of measurements directly in cells, and their interpretation in that context.

NMR spectroscopists have attempted to study proteins inside living cells since the early 2000s and since then, a whole field with different directions of study has emerged[141–144]. In-cell NMR faces unique challenges, as the cells are taken out of their optimal growth environment and are measured densely packed in a nutrient-scarce tube where they often die and lyse rapidly. Paired with the rather low sensitivity and long measurement times of NMR spectroscopy, a single in-cell NMR experiment is rather costly, laborsome and yield usually only poor spectral quality. Nevertheless, in-cell NMR experiments of proteins have successfully been measured in several cell lines of different organisms (e.g. *E.coli*[145–147], *X.laevis* oocytes[148–150], HEK293[151,152] and even primary mouse neurons[153]). In a few cases, structures could be solved based on the measurements in living cells[154,155], and even structural dynamics[156,157], protein-protein[158–160] and protein-small molecule interactions[161] could be observed.

Developments for protein in-cell NMR were direct inspirations for the studies of nucleic acids (NAs), even though they come with a few unique advantages and obstacles. The rest of this chapter will be dedicated to discussing the literature of in-cell NMR of nucleic acids with respect to the obstacles encountered, and the solutions that were applied or proposed.

2.3.2 Application to nucleic acids

Studies in *X.laevis* oocytes

Overexpression of proteins in *E.coli* or HEK293T is a common way to produce labeled and unlabeled in-cell NMR samples[142,162]. Contrary to than proteins, NAs are not easily overexpressed in cells. Instead, they have been introduced into cells in all published studies thus far. This allows the labeled or unlabeled sample preparation with SPOS or IVT as discussed earlier. A model system that allows simple and robust transfection of a defined sample amount is microinjection into stage IV *X.laevis* oocytes, which are ~1 mm in size and has been used for the injection for protein samples previously[148,163]. The injected volume can be up to 50 nL and only 200 oocytes are required for an entire NMR sample. Furthermore, the defined volume allows the injection of enough NA that a final NMR sample concentration of ~150 µM can be reached, given that the amounts injected are not apoptotic[164].

In a pilot study from 2009, Hänsel and coworkers injected several RNA and DNA molecules, bearing different modifications and labels to assess the feasibility of NA studies with in-cell NMR in oocytes[164]. They found that phosphorothioate modifications (PS) on the first two nucleotides, as well as 2'-OMe modifications improved lifetime of RNA and DNA in oocytes. As expected from protein studies, NMR spectra showed significant broadening and NAs degraded within less than 24 hours. Importantly, imino resonances could be observed in absence of cellular background signals between 9 and 15 ppm.

The study proved the general feasibility of in-cell NMR for nucleic acids and led the authors to pursue the investigation of telomeric G4 topology in living cells. First, the authors could use imino resonances in oocyte lysate and complementary PEG-crowded *in vitro* experiments to show that the assumed parallel topology was not the preferred one, as previously assumed[14]. In a follow-up study performed in living *X.laevis* oocytes, the hybrid-2 and 2-tetrad antiparallel basket structures could be identified as primary G4 topologies, alongside bringing insight into the importance of overhang sequence next to crowding and dehydration effects of the cellular environment[11].

These in-cell NMR results helped reject conventional PEG solutions as a meaningful crowding reagent to mimic the interior of a cell, and were supported by ^{19}F-NMR studies later[12,13]. The groups of Xu and Srivatsan independently tagged the telomeric repeat sequence TAGGG(TTAGGG)$_3$TT with two different fluorine-containing NMR reporters. Both groups found the hybrid-2 confirmation to be the dominant in HeLa cells and *X.Laevis* oocytes and K^+-containing buffer, but the parallel topology in PEG-crowded K^+-buffer, despite different ^{19}F-tags and cell lines.

Transfection of human cells

X.Laevis oocytes are useful systems to explore in-cell NMR experiments, but hardly a representative model system for human cells. In order to move to cultured human cells as more meaningful systems, transfection methods that can deliver the necessary amount of nucleic acid

material were required. Two such protocols were developed by the groups of Trantírek and Katahira.

The former is an adaption of the protocol used by Theillet and colleagues to electroporate α-synuclein into cells for in-cell NMR experiments[153,165] while the latter is a method previously for the delivery of proteins, where a pore-forming bacterial toxin Streptolysin O (SLO) is used to create holes in the cell wall of HeLa cells, which allows the diffusion of DNA or RNA from the medium. These pores can later be closed by addition of calcium[166].

Both protocols allowed the detection of different RNA and DNA constructs at comparable concentrations of 5 – 20 µM in the final HeLa-NMR sample. Individual experiments differed with regards to how much leakage was observed, and how many cells died in the delivery procedure. A comparable study is missing, but the general trend in the literature indicates that SLO transfection leads to more leakage and more dead cells[167], which have in part been removed by Percoll density gradient centrifugation[13,166,168].

Both methods use an input amount of 1 – 1.3 µmol of RNA and obtained a final amount of approx. 4 nmol (10 µM in 400 µL volume). To fill this volume, ~10^8 cells are needed in the NMR, which implies that each cell has to take up ~$24*10^6$ molecules which contribute to the same NMR signal. This is a considerable amount, especially when comparing these numbers to studies from other fields, that quantified intracellular RNA concentration. tRNA and rRNA concentrations were estimated to be in the low µM range[169], while mRNA sequencing revealed abundances of only a few to few hundred individual copies per mRNA[170]. MicroRNAs instead are of much higher abundance, with the highest-abundant microRNAs existing in up to 100,000 copies per cell[171]. The exogenous NA for an in-cell NMR experiment thus is in great excess over any individual RNA species in the cell.

Given the lower transfection yield of both SLO and electroporation compared to oocyte microinjection, the sample production is considerably more expensive, especially when using labeled samples. A trend in the field of in-cell NMR shows that researchers largely stopped using $^{13}C/^{15}N$-labeled samples after a few initial studies in oocytes[11,164,172]. This is probably owed to the fact that 2D experiments take considerably longer than 1D, and even isotope-filters to identify exogenous from intracellular signals did not yield expected results. Furthermore, imino signals and ^{19}F-NMR proved invaluable for structural studies, as will be shown in the following section.

Nucleic acids studied with in-cell NMR

An overview of all NA constructs studies with in-cell NMR are shown in Figure 14. A third of all NA sequences studied were G4 structures, mostly containing the telomeric repeat sequence TTAGGG[11–13,164,166,172–175], motivated by elucidating the topology under native conditions to allow for the design of better drugs[172,176]. Short model hairpins were used for proof-of-principle experiments showing the successful detection of RNA (14 nt) and DNA (7 nt)[164,166,167], alongside other model constructs[165,177]. Additional non-canonical DNA beyond G4s were successfully observed and studied with in-cell NMR, namely i-motif[178,179], TT-mismatch[177], Z-

DNA[168] and triplex DNA[180]. Strikingly, the i-motif DNA proved more stable in the cellular context than what would have been expected based on pH-shifted.

RNA constructs are represented apart from the 14-mer model hairpin[164,166], only with an RNA G4 structure[174] and the 2'-dG-aptamer was studied[167]. Additionally, a chemically modified RNA drug, Danvatirsen, was studied with dynamic nuclear polarization (DNP) NMR[181].

Intracellular stability is a major problem for in-cell NMR, as foreign RNA and DNA are seen as threats and rapidly degraded. Many exo- and endonucleases exist are present in the cytoplasm that can degrade such RNA, such as exoribonuclease I[182–184] or the exosome[185,186] (Figure 15B). Furthermore, 5'-triphosphate-containing and duplex RNA are highly immunogenic[187]. Chemical modification greatly extents the lifetime of NAs in the cell, for example phosphorothioate, 2'-OMe or locked nucleic acids[188], as shown for Danvatirsen, which was could still be detected after 24 hours upon lysis and protein degradation[181]. As most nucleases in the cell seem to be exonucleases, the 5' and 3' ends have a key role for intracellular stability. Additionally, structures like G4 can further increase the lifetime of nucleic acids.

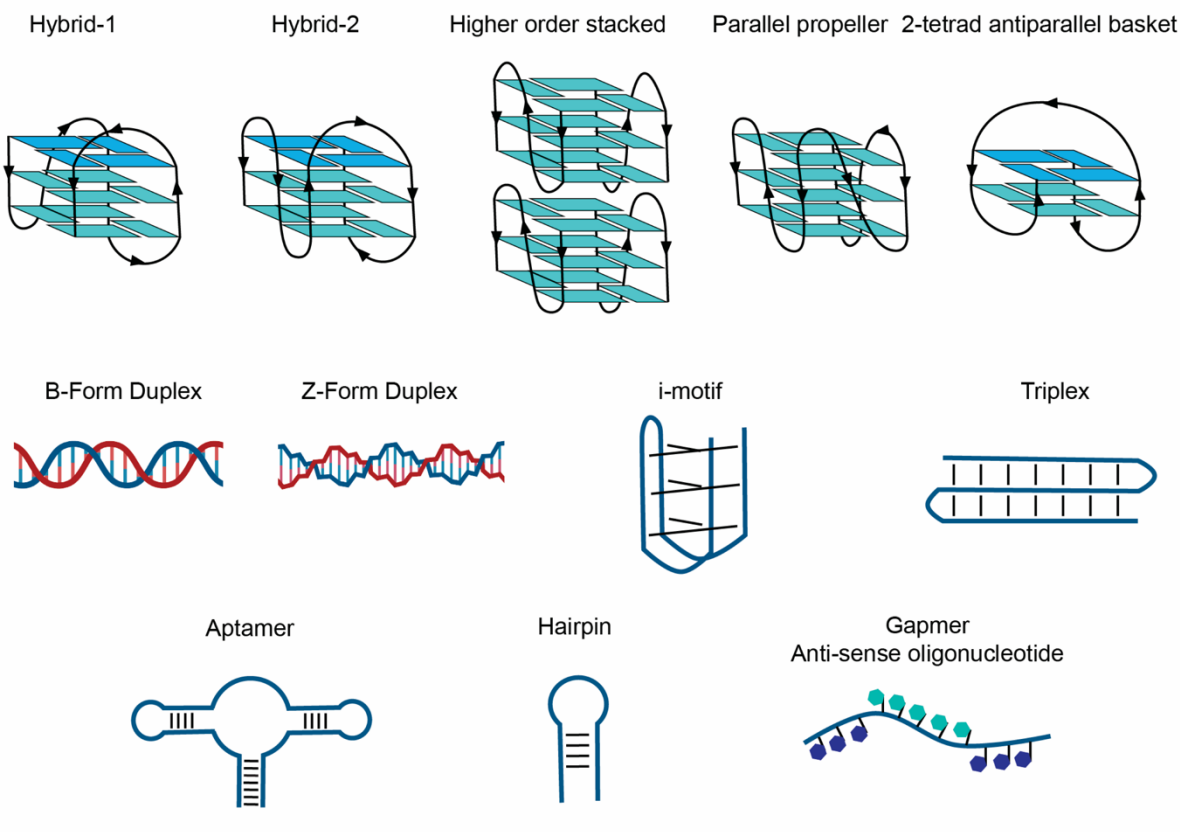

Figure 14: All DNA and RNA structures observed by in-cell NMR. For G-quadruplexes, the most common conformations are shown, even though not all might have been observed in the cell. Figure created by Henry Annecke.

Specificity

In order to make inferences about the NA structure inside the living cell, researchers needed to show specificity that (a) the signals are arising from the introduced NA, and not from the

cellular background and (b) the NA observed is in fact in the cell and not in the lumen between them.

In order to provide specificity over the cellular background, isotopic labels have been used for NMR filter steps in 1D and 2D[11,164,167,172]. When using low-abundance isotopes like ^{13}C or ^{15}N, even some cellular signals appear in control experiments, which can be subtracted from the in-cell NMR spectrum, as is common practice when proteins are overexpressed in *E.Coli* or HEK293 cells[152,162]. Magnetization transfer via ^{13}C-^{13}C or ^{13}C-^{15}N bonds can virtually remove the cellular background as shown in solid-state NMR[189]. The imino region is virtually free of endogenous signals, probably because all cellular nucleic acids are either too large (mRNA, rRNA, genomic DNA) or at too low concentration (mRNA, small RNAs). ^{19}F is an excellent isotope for in-cell NMR as it has a high gyromagnetic ratio, no cellular background, very sensitive to structural changes and easily incorporated in NA by means of SPOS. ^{31}P has strong cellular background between 10 to -20 ppm from derivates of ATP and phospholipids but is shifted to ~55 ppm in a phosphorothioate modification, which allows specific detection of those NAs[164,181].

In order to ensure local specificity of the NA in the cells, the Trantírek group transfected FAM-labeled DNA at a ratio of 1:40 with the unlabeled sample and monitored uptake and localization via flow cytometry and fluorescence microscopy[165]. Also, it was found that exogenous NA can leak from the cells and be detected in the supernatant[13], as was found for proteins previously[190]. These signals from leakage are often subtracted to yield the true in-cell signal (Figure 15A). A bioreactor setup, used to extend the lifetime of cells, also brings the advantage of flushing leaked NA from the tube, effectively removing their contribution to the signal[180].

NMR methods in cells

As mentioned earlier, the NMR measurement is hampered by various sample parameters, for example short lifetime, conformational and magnetic heterogeneity and slowed rotation in the cell. All these factors lead to increased linewidth and reduced SNR, which can be combatted with NMR methods. Potential overlap due to increased linewidth can of course be reduced by introducing a second dimension, e.g. ^{1}H, ^{13}C or ^{15}N. This has been done for heteronuclei in early experiments[11,164,167,172], but never for homonuclear correlations. Heteronuclear correlation experiments were performed either as TROSY-HSQC experiments[164] (transverse relaxation optimized spectroscopy[191]), which generates narrower lines for slowly tumbling molecules, and SOFAST-HMQC[167,172], which is a way to reduce measurement time by band-selective excitation and use of short flip angle pulses[192].

Good water suppression is crucial for successful in-cell NMR experiments when detecting ^{1}H resonances. The 1-1 jump-and-return water suppression[193] was frequently used for water suppression with retaining high sensitivity for exchangeable imino protons[167,176]. SOFAST experiments can also be recorded in a 1D fashion selective for the imino region, which greatly enhances sensitivity per time unit[180] and alleviates the need for water suppression. Schlagnitweit and colleagues used ^{1}H,^{31}P-cross polarization for solid-state NMR experiments

to detect phosphorothioate modifications on the Danvatirsen drug while retaining high sensitivity of the ^{1}H nucleus[181]. ^{19}F-spectra are generally recorded as 1D experiments, as they are very sensitive and fast.

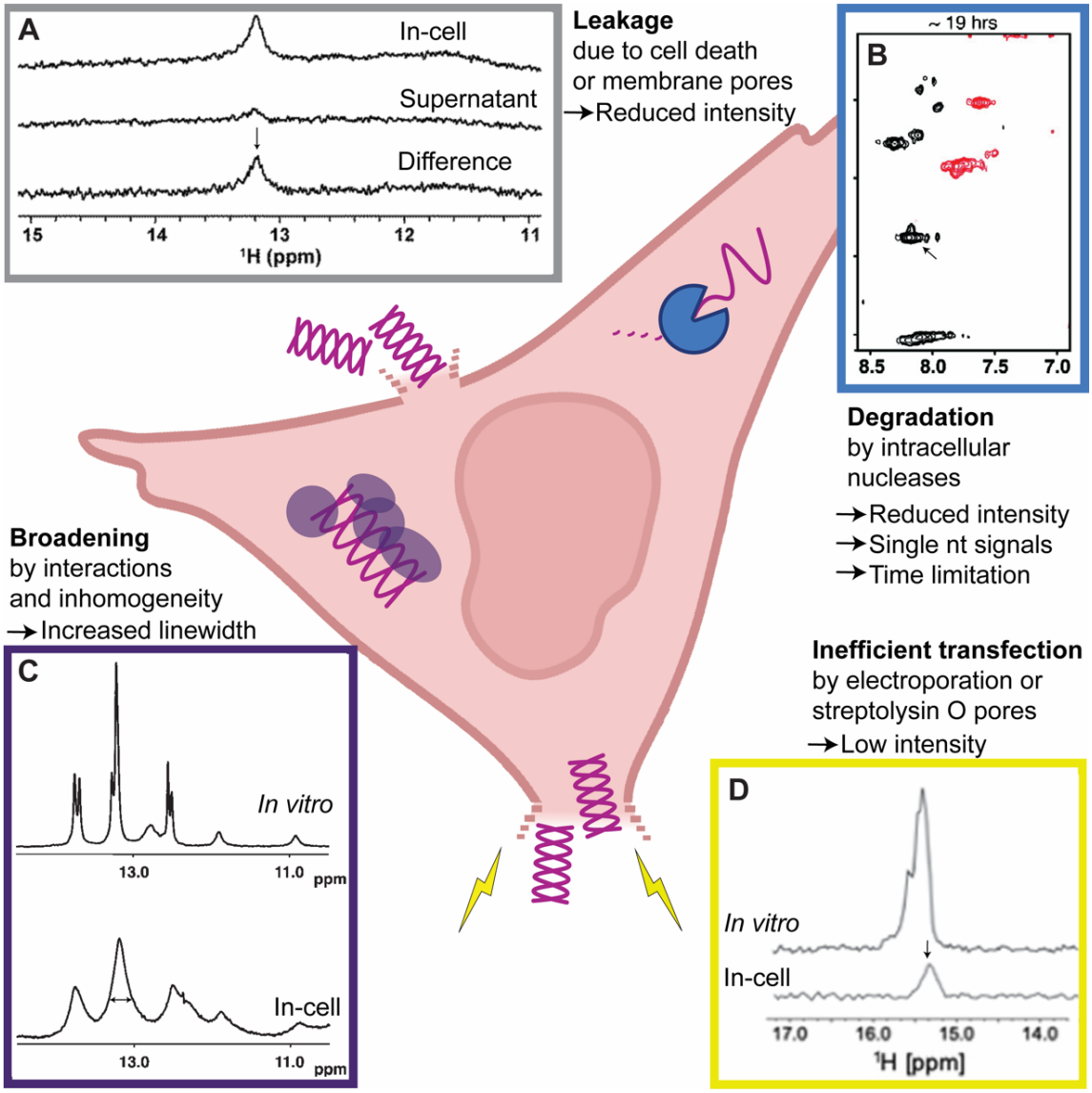

Figure 15: Challenges of an in-cell NMR experiment. A: Leaked NA can be found in the supernatant, effectively reducing the signal intensity after subtraction. B: Nucleic acids degrade in the cells, reducing signal intensity and leading to signals of single nucleotides. C: The crowded conditions in the cells are leading to considerable line-broadening, D: Transfection especially into human cells is rather inefficient, which reduces the signal intensity further.

Trends with in-cell NMR

While the field of in-cell NMR of NA developed, a shift in trend could be detected which is evidenced by which constructs were studied, and how they were designed and investigated. Initial constructs, especially G4s were approached with the methods high-resolution NMR, i.e. 2D NMR and uniform labeling[11,172], and compared to circular dichroism (CD) experiments[11]. In later studies, a simpler readout was chosen that could be obtained from a 1D spectrum. This was either ^{19}F-NMR (Z-DNA[168], G4[173–176]) or imino protons (i-motif[178,179], triplex[180], G4[176], TT-mismatch and model hairpin[177]) as both have no cellular background and are highly structure-specific. These signals were usually interpreted by comparison with *in vitro* NMR data and complementary methods like CD or fluorescence spectroscopy[12,179]. In such an

approach, in-cell NMR is closer to a tagged approach the likes of fluorescence or EPR spectroscopy than high-resolution NMR, where structures and connections can be solved *de novo*. This is also true for imino signals, as they are used as a finger print for comparison with *in vitro* experiments. It is probably the low signal intensity and severe broadening that motivated such a changed approach. Also, the possibilities of chemical synthesis allow the simple incorporation of ^{19}F-tags, which is much more difficult for proteins. In fact, most in-cell NMR studies of nucleic acid utilized ^{19}F-tags on either the 5'-end (hexyl-linked 3,5-bis(trifluoromethyl)benzyl), or a base modification (5-fluorobenzofuran-dU or 8-trifluoromethyl-dG). The frequently found fluorine modifications C2'-F, 5F-U/T or 5F-C (Or CF_3 variants thereof) have not been used in published in-cell NMR data to this point. Especially the latter two have great potential due to their cheap availability, high structural sensitivity and low perturbation of the NA structure.

Crowding agents such as PEG or Ficoll have frequently been used to mimic molecular crowding and excluded volume effects from the intracellular environment. It is evident that crowding agents can have an impact on NA structure and function, as shown for a riboswitch[194], G4[9] and i-motifs[195], but had no effect on the structure of an mRNA[196]. The in-cell NMR data shows that especially PEG is a poor mimic for intracellular crowding, especially for G-quadruplexes[11–14]. Crowding agents, alongside specific ions or dehydrating agents can help understand the effects governing molecular interaction in the cell, but have to be used in well-controlled experiments, and in-cell NMR can clearly help validating these models better.

3 RESEARCH AIMS

The overarching aims of the PhD were to develop methods which would push the boundaries of NMR studies of small RNA molecules, particularly with interest for in-cell NMR.

3.1 Increasing yield of *in vitro* transcription

Many RNA sequences are difficult to produce in high enough yield and purity for NMR samples, especially if they have a non-GG initiation sequence. An initial target of our in-cell NMR studies, miR-34a starts with a uridine nucleotide, which reduces transcription yield and purity. Hence, we were aiming to develop a method which can make sequences like miR-34a a viable construct for in-cell NMR and remove existing boundaries of sample preparation limitations of time and cost. These results were published in papers I and II.

3.2 Isotopic labeling of RNA for NMR studies

Due to low chemical complexity of RNA, resulting in chemical shift dispersion, NMR spectra of RNAs show considerable resonance overlap with increasing size. This hampers interpretation significantly, and in some instances precludes unambiguous resonance assignment. To reduce resonance overlap, we use site-specific isotopic labeling with purely enzymatic methods, which in a similar version has been published previously[197]. This work resulted in paper III, which is currently in submission.

3.3 RNA systems for in-cell NMR

In-cell NMR studies are notoriously difficult from an experimental point of view, as they require milligrams of RNA per single experiment (potentially isotope-labelled), while only yielding few hours of measurement time at low concentration and increased broadening. Hence, the aim was to develop strategies to make such measurements more feasible from an RNA, cell culture and NMR perspective. We attempt overexpression to generate in-cell NMR samples to eliminate laborious sample production and transfection protocols. To increase lifetime in the cell and allow for longer measurement times, we overexpress circular RNA using the TORNADO method[198], which should achieve intracellular RNA concentrations in the micromolar range. Results from this aim are so far unpublished and are collected in chapter 5.

4 MATERIALS AND METHODS

In this section, methods relevant to the unpublished and preliminary results presented in chapter 5 are described. Materials and methods for the papers I – III can be found in the respective manuscripts at the end of the thesis.

4.1 RNA preparation

4.1.1 IVT

Reagent preparation

Plasmid templates were transformed into *E.coli* DH5α by heat-shock, amplified in 250 mL or 500 mL culture overnight and purified using the Machery-Nagel MaxiPrep kit. Purified plasmids were linearized at 20 – 200 ng/µL using BamHI in volume from 50 µL – 1 mL. Successful linearization was confirmed on 1% an agarose gel.

Tris-Cl, $MgCl_2$ or $Mg(OAc)_2$ and DTT were prepared to 1 M in H_2O, spermidine to 0.25 M and sterile-filtered using 0.22 µm cellulose acetate syringe filters. Unlabeled NMPs and NTPs were prepared to 100 mM stocks and sterile filtered. Labeled nucleotides were shipped at 100 mM in 10 mM Tris-Cl. All IVT reagents were ordered from Sigma-Aldrich.

All in-house-produced enzymes were prepared by the protein science facility (PSF) through amplification in *E.coli* BL21(DE3) followed by His-tag and size-exclusion purification. Obtained enzymes were tested for RNase contamination by incubation with purified RNA a room temperature for 24 – 72 hours and stability confirmation with denaturing PAGE.

The wildtype T7RNAP was used, apart from the S43Y mutant[99] where indicated.

Polymerase chain reaction (PCR) was performed using the Phusion PCR kit (NEB) according to manufacturer's instruction.

IVT preparation

IVT reactions were performed at scale of 50 µL – 20 mL from different templates. Reagents were mixed in the order: H_2O, Tris-Cl, $MgCl_2$ or $Mg(OAc)_2$, DTT, Spermidine, NMP, NTPs, cleavage guide, template, iPPase, RNase H, T7RNAP, and exact concentrations for the two RNA constructs discussed in chapter 5 are shown in table 1. Reactions were incubated for 1 – 2 hours for test reactions and overnight (~16 hours) for preparative reactions at 37°C. The respective reagent concentrations for the RNAs produced in chapter 5 can be seen in table 1. Reactions were quenched by addition of EDTA to 50 mM and thorough vortexing. Quenched reactions were filtered to remove precipitated protein.

Table 1: Optimized IVT conditions for pri-racGUG and Broccoli (53 nt, tandem template).

Reagent	pri-racGUG	Broccoli (tandem template)
Tris-Cl pH 8	75 mM	80 mM
$MgCl_2$ or $Mg(OAc)_2$	20 mM ($MgCl_2$)	17 mM $Mg(OAc)_2$
DTT	10 mM	5 mM
Spermidine	20 mM	10 mM
GMP	8 mM	4 mM
NTPs	8 mM each	4 mM
Plasmid	3.5 ng/µL	10 ng/µL
Cleavage guide	-	20 µM
RNase H	-	1.2 µM
iPPase	1 µM	1 µM
T7RNAP	3 µM (wildtype)	1.4 µM (S43Y mutant)

Optimization

For a given template, reagent concentrations were varied one at a time on rational basis to generate 10 – 20 new IVT reactions. After small scale reactions for 1 – 2 hours, reaction yield was compared on a denaturing PAGE to compare the main band intensity and derive improvement of reaction conditions.

4.1.2 Gel electrophoresis

Denaturing polyacrylamide gels were cast in 12x8 cm gels and run in the Mini-Protean system (Biorad). Acrylamide solution contained 10 – 20 % acrylamide, 1X TBE (100 mM Tris, 90 mM boric acid, 1 mM EDTA) and 8 M urea and were polymerized in 5 mL upon addition of 4 µL TEMED and 40 µL ammonium persulfate. Gels were run in 1X TBE at 350 V for 20% acrylamide and 250 V for 10% acrylamide for 1 hour respectively.

1 µL sample volume was typically diluted in 9 µL loading solution (240 µM bromophenol blue, 5 mM EDTA in formamide) and heated to 95°C for 2 minutes for denaturation. 1 µL was loaded for IVT reactions, otherwise the volume was adjusted to contain 1 – 10 pmol of RNA.

Gels were stained with SYBRGold according to manufacturer instructions and imaged using a UV tray imager.

4.1.3 IE-HPLC purification

Ion-exchange (IE) HPLC purification was performed on a Dionex Ultimate 3000 system with a DNAPac200 22x250mm ion-exchange column. Buffers used were A: 20 mM ammonium acetate pH 6.5, 20 mM sodium perchlorate and B: 20 mM ammonium acetate pH 6.5, 600 mM sodium perchlorate. Flow rates were either at 4.5 mL/min or 8 mL/min. Gradients for elution were optimized with injection volumes equivalent to 1 mL IVT reactions. A typical elution gradient for a 20 – 30 nt RNA would be 20 – 30% buffer B over 30 minutes. Collected fractions were analyzed on denaturing PAGE before pooling the fractions of interest. Pooled fractions were buffer-exchanged against water using either Amicon stirred cell or Amicon centrifuge filter units. Sample concentration was determined with the Lambert-Beer law, using molar extinction coefficients from the IDT oligo analyzer tool.

4.1.4 *In vitro* circularization

Ribozyme-assisted circular GUG (racGUG) was circularized *in vitro* using recombinant RtcB ligase (uniprot: P468500. ringBroccoli was circularized using recombinant T4 RNA ligase 2 (Rnl2, uniprot: P322770). Both enzymes were produced by the Protein Science Facility at the Karolinska Institute. Both methods required extensive optimization of RNA-protein ratios, reaction time and buffer conditions to abolish intermolecular ligation products and leftover linear product. The following conditions were used for scale-up experiments.

racGUG was circularized at 15 µM RNA concentration in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 10 mM GTP, 10 mM $MnCl_2$ and 1.2 µM RtcB ligase. The reaction was incubated at 37°C for 1 h. Reactions were performed in volumes of 15 µL to 30 mL, depending on further analysis method.

ringBroccoli was circularized using the 0.1X buffer method by Chen and colleagues[199]. This entailed 5 µM RNA concentration in 5 mM Tris-Cl (pH 7.5), 0.2 mM $MgCl_2$, 0.1 mM DTT, 40 µM ATP and 0.125 µM Rnl2. Reactions were incubated at 37°C for 1 h.

4.1.5 Circularization assays

Three different methods were used to confirm circularity after intramolecular RNA ligation: site-specific RNase H digest, RNase R digest and gel shift assay on PAGE at different acrylamide concentrations. Depending on the RNA construct, one or two suitable methods were chosen, but not all. RNase R, RNase H and their respective buffers were added to circularization reactions for small-scale analysis.

Site-specific RNase H cleavage of a circular RNA creates one linear RNA band migrating at a different height compared to the circular RNA[198]. A linear product instead will be cleaved into to shorter fragments. RNase H digest was performed at cleavage guide excess of 1.2 – 2 times

in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT and 0.6 µM RNase H. The reactions were analyzed on denaturing PAGE gels.

RNase R is a 3'-5' exonuclease that cleaves linear with high tolerance for non-canonical 3' termini, but not double-stranded RNA with less than 7 overhanging nucleotides. RNase R (Lucigen) digest was performed in 20 mM Tris-HCl (pH 8.0), 100 mM KCl and 0.1 mM $MgCl_2$ at 2 u/µL RNase R and RNA concentrations between 1 and 10 µM.

Circular has altered PAGE migration properties compared to linear RNA, which can be exploited to assess circularity by comparing migration distance of circular compared to a linear RNA reference on gels with different acrylamide concentrations[198]. Acrylamide concentrations were 6% vs 10%, or 15% vs 20%, depending on RNA size. For extracted RNA, endogenous tRNA and rRNA bands served as reference bands.

4.1.6 In-gel Broccoli staining

Denaturing acrylamide gels for DFHBI staining were washed 3 times in 20 mL water for 5 minutes and stained for 30 minutes in 40 mM HEPES-KOH (pH 7.3), 100 mM KCl, 1 mM $MgCl_2$, 10 µM DFHBI and imaged on a UV imager.

4.2 NMR spectroscopy

All NMR experiments were performed on a Bruker Avance III 5 mm HCNP Cryo-probe in Shigemi tubes. The standard NMR buffer was 15 mM sodium phosphate, 25 mM NaCl, 0.1 mM EDTA pH 6.5 if not otherwise specified. Broccoli and DFHBI have been measured in 30 mM HEPES (pH 7.3), 100 mM KCl, 1 mM $MgCl_2$[41].

^{1}H-1D experiments have been recorded with excitation sculpting water suppression or as selective 1D-imino-SOFAST experiment at a bandwidth of 6 ppm and a carrier frequency of 12 ppm.

^{19}F-1D experiments have been recorded using a BBO probe with a sweep width of 492 ppm as Hahn-echo experiments without ^{1}H decoupling.

4.3 Cell culture

4.3.1 General

HEK293T and HeLa cells have been cultured in 10 cm uncoated dishes with 10 mL Dulbecco's modified Eagle's medium (DMEM, Sigma) containing 10% fetal bovine serum (FBS) at 37°C and 5% CO_2 and passaged at a ratio of 1:20 (HEK293T) and 20% (HeLa) to reach full confluency after 3-4 days. All cell work was performed in a biosafety level 2 cabinet. Cells were detached by addition of 2 mL Trypsin (0.05%, Sigma) and incubation at 37°C for 2 minutes. Trypsin was inactivated by addition of 8 mL DMEM containing 10% FBS.

Cell viability and cell count was simultaneously assessed by Trypan blue staining (Sigma) using a Countess II cell counter (Thermo) according to manufacturer's instructions.

4.3.2 TORNADO expression

Day 1: 10 dishes of HEK293T cells were seeded at a density of 2.4 million cells in 10 mL DMEM incl. 10% FBS. Day 2: Cells were transfected with the Broccoli or GUG TORNADO-plasmid (Broccoli: Addgene ID 124360, GUG: ordered from Genscript, RNA sequence in chapter 5.4.1) using FuGENE HD (Promega). Per plate, 37 µg plasmid and 112 µL FuGENE were mixed 1.7 mL Opti-MEM (Sigma), briefly vortexed, incubated for 12 minutes and dropwise added to the culture dish. Day 3: The transfection procedure from day 2 was repeated. Day 5: Cells were harvested for in-cell NMR, RNA extraction or DFHBI-stained for microscopy.

4.3.3 RNA extraction

RNA was extracted using the miRvana kit (Thermo) according to manufacturer's instructions. RNA concentration was determined using the Nanodrop Spectrophotometer (Thermo) and 10 – 30 ng total RNA extract was loaded onto a 10% denaturing PAGE for analysis.

4.3.4 In-cell NMR

Ca. 130 million harvested cells were washed in 50 mL DMEM including 10% FBS and resuspended in 600 µL Leibovitz-15 medium (L-15, Sigma) containing 10% D_2O. The cells suspension was transferred to a Shigemi tube without plunger and carefully pelleted using a hand-cranked centrifuge. Measurements were taken at 298 K using either the 1D-SOFAST sequence, a 1D experiment using excitation sculpting or 1-1 jump-and-return water suppression. After measurement, the medium and cells atop the active volume (~2 cm) were removed with a long Pasteur pipette and discarded. The cell slurry was removed from the tube and centrifuged using a small table-top centrifuge. The supernatant was collected and the equivalent in L-15 medium including 10% D_2O of the pellet volume (~ 400 µL) was added to the supernatant. This sample was again measured to assess leakage from the cells.

Cells were lysed in 100 µL PBS by 5 freeze-thaw cycles and sonication as described by Schlagnitweit, Sandoz and colleagues[181]. Cell debris was removed by centrifugation at 13,000 xg for 10 minutes at 4 degrees. The lysate was subjected to Proteinase K-treatment (Thermo) at 0.5 mg/mL for 30 minutes at 37°C and 15 minutes at 55°C. The digested lysate was measured after addition of 10% D_2O and transfer to a Shigemi tube with plunger.

5 RESULTS AND DISCUSSION

The following section will briefly sum up the constituent papers I – III, which are linked in full text at the end of the thesis, and additional unpublished data. Furthermore, preliminary results for a 4th project are shown and discussed.

5.1 One-pot production of RNase H-cleaved tandem transcripts

5.1.1 Paper I

We developed a method that allows the production of short RNAs in high yield and purity by transcribing several copies of the target RNA from a tandem template and successive site-specific cleavage with RNase H. The increased yield and purity are achieved because RNase H cleavage is more specific than the termini generated by run-off transcription of T7RNAP. Furthermore, there is no sequence strain on transcription initiation sequence or the RNase H cleavage site because the former can also be removed with RNase H cleavage, and the latter is targeted by a chimeric cleavage guide oligo with complementary over the cleavage site. We showed the production of such an approach and the high scalability of the method as required for the production of NMR samples.

5.1.2 Additional data

We explored other templates suitable for the approach of RNase H-cleaving tandem transcripts to obtain target-length RNA. As the initiation event appears to be the kinetic bottleneck in IVT, it seems advantageous to maximize the number of templates that can be transcribed. When using the tandem template from a linearized plasmid template, this is limited to the 26 repeats due to the 600 bp-restriction imposed by the manufacturer. To enhance the runtime of a single polymerase enzyme, we attempted rolling circle transcription (RCT) by circularizing the insert, which would generate very large transcripts, but can still be cleaved by RNase H as has been shown before[200].

The RCT template was generated as shown in Figure 16A. We amplified the insert by PCR and simultaneously introduced BamHI restriction sites by primer overhangs. Following PCR purification with a commercial kit, the amplicon was digested with BamHI to produce complementary sticky ends, which were ligated with T4 DNA ligase (data not shown). As shown in Figure 16B, this template can be used to generate an RNA so large it barely enters the gel. A size of the band could not be estimated. Upon addition of RNase H and cleavage guide, this band collapsed into a band of the target length, 22 nt.

This protocol to generate the RCT template suffered from very low yield. Approximately 200 µL of PCR were necessary to obtain enough template for an IVT of 50 µL. The IVT itself was also poorly optimized and yielded only little RNA. Despite the elegance of the approach, it is unfortunately not practical for production of NMR samples in its current form.

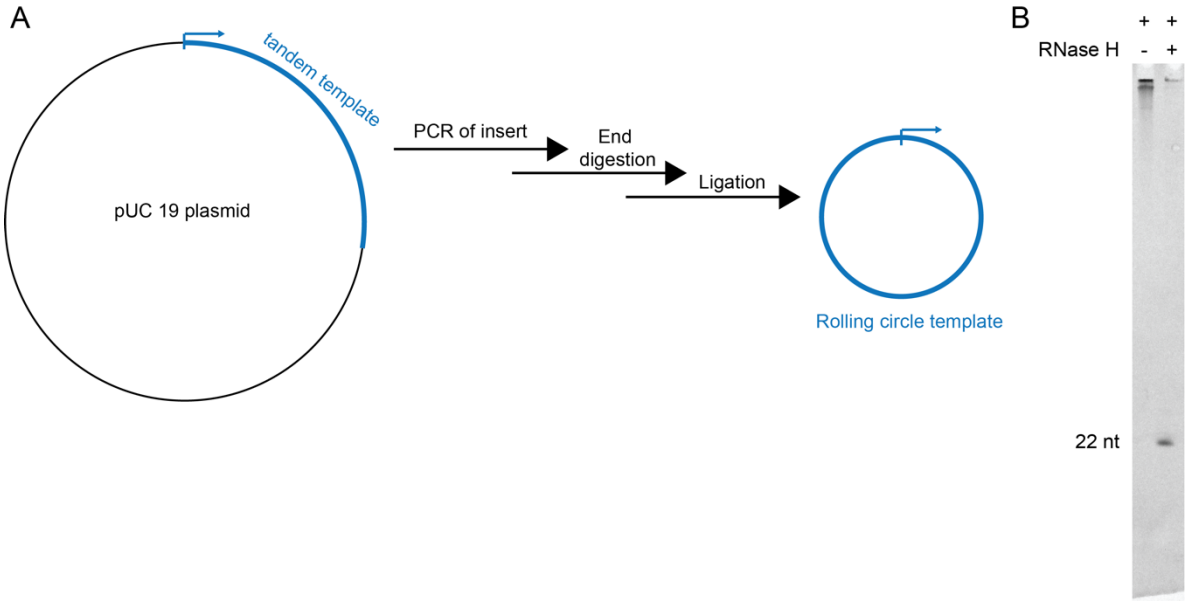

Figure 16: Rolling circle transcription from a circularized tandem template (construct 3 from paper I). A: After PCR amplification and introduction of BamHI restriction sequences, the ends were digested with BamHI, creating sticky end overhangs. The complementary sticky ends were then ligated using T4 DNA ligase. The new dsDNA circle consists only of the T7 promoter, 26 repeats of the target sequence and the ligation leftovers. B: This template can be used for T7 transcription, and such transcripts are being cleaved by RNase H to the target RNA length.

5.2 Practical aspects of sample preparation and setup of $R_{1\rho}$ relaxation dispersion experiments

Paper II was published in the journal of visual experiments (JoVE) and exemplified the communication of an experimental protocol in form of a video, in which we described the sample preparation of an RNA sample with the approach published in paper I, and successive NMR experiments to confirm RNA folding and the setup of $R_{1\rho}$ relaxation dispersion experiments. The $R_{1\rho}$ relaxation dispersion experiments were the first description of ^{1}H relaxation dispersion measured in unlabeled RNA samples and published as the SELOPE approach by my colleagues previously[57]. Since the publication is a protocol paper, no new results can be discussed at this point.

5.3 Site-specific labeling of RNA with enzymatic methods

We showed the simple product of an RNA carrying a single isotope-labeled nucleotide, namely A31 in a 46 nt-long RNA. This was achieved by transcribing two separate RNAs, relating to nucleotides 1 – 30 (5' fragment) and 31 – 46 (3' fragment) of the full-length RNA. Both were produced with IVT, and the 3' fragment was supplied with $^{13}C/^{15}N$-labeled AMP, resulting in the first nucleotide being isotope-labeled. After ligation, this label is placed at position A31 of the 46 nt RNA.

We demonstrated the successful production of this A31-labeled RNA and NMR data on the successful incorporation of the label. Surprisingly, the single nucleotide showed four

resonances in an aromatic HSQC, where only two were expected (one for C2 and one for C8). Further experiments showed that a second conformation was present in the sample, which was hidden by overlapping nucleotides in a uniformly A/U-labeled sample.

5.4 Circular RNAs as viable targets for in-cell NMR

5.4.1 Production of circular GUG and Broccoli *in vitro* and in cells

Design and expression of circular GUG

The half-life of transcripts expressed as circular RNAs is significantly higher compared to their linear counterparts, as shown by Litke and Jaffrey in the expression of TORNADO (Twister-optimized RNA for durable overexpression) method[198]. racBroccoli (ribozyme-assisted circular Broccoli) could be expressed to 15 – 20 µM intracellular concentration in HEK293 and HeLa cells. We adapted their approach to produce a racRNA construct within the size-limit of solution NMR for in-cell studies of conformational dynamics. We chose a construct called GUG, which is a previously studied and published stem-loop RNA with a single-nucleotide bulge that exhibits a strong relaxation dispersion for the nucleotides around the bulge region[57] (Figure 17A). To express racGUG in mammalian cells, a ligation stem to be recognized by RtcB ligase and flanking ribozyme sequences had to be added to the stem-loop structure (Figure 17C). Several mutations to conserve ribozyme functions had to be introduced and the ligation stem was shortened. The final racGUG construct (Figure 17B) is 40 nucleotides long after cis-cleavage of the Twister ribozymes (Figure 17C).

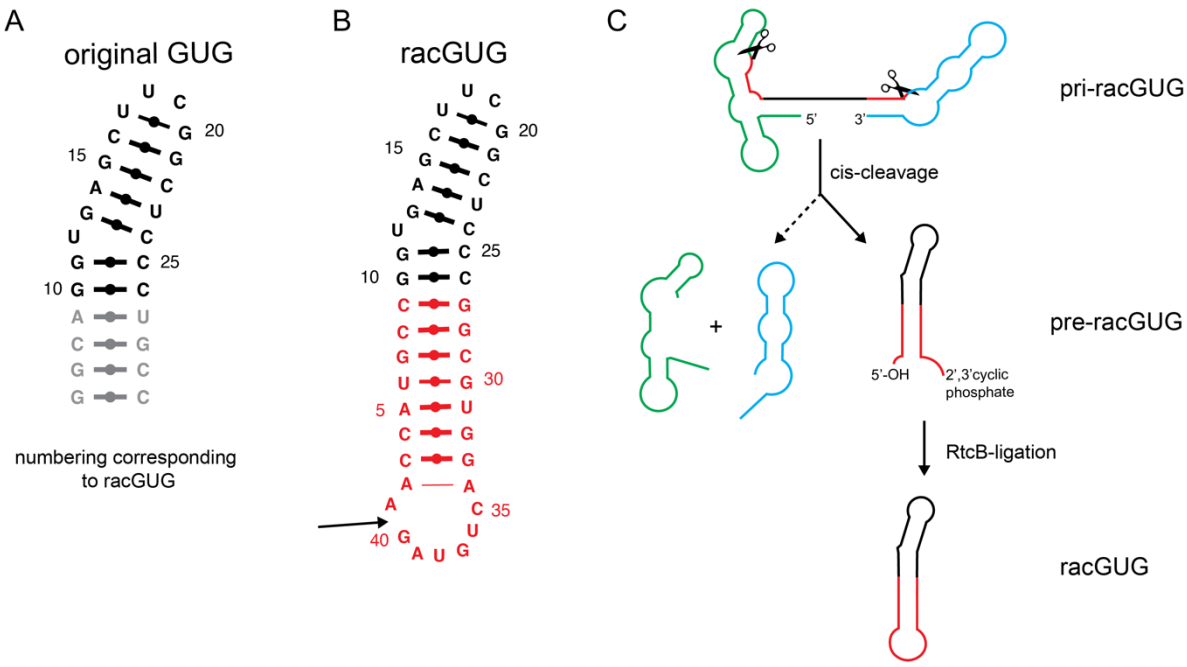

Figure 17: Design of racGUG from the original GUG sequence. A: Secondary structure representation of the original GUG construct (27 nt) and B: ribozyme-assisted circular GUG (racGUG,40 nt). Grayed out nucleotides in the original GUG have been replaced with a shortened ligation stem (red). The arrow indicates the RtcB-ligation site. Secondary structures were created with Mc-Fold. C: Scheme for the production of racGUG. A long RNA is transcribed in which the target sequence (black/red) is flanked by two Twister ribozymes (green and blue). The ribozymes cleave co-transcriptionally, leaving 5'-OH and 2',3'-cyclic phosphate termini in the resulting pre-racGUG. Finally, the RtcB ligase catalyzes the intramolecular ligation, resulting in the circular racGUG. Panel C was redrawn and adapted from Litke and Jaffrey[198].

The following sequences were created and used in the following chapter:

Original GUG (27 nt)[57]:
GGCAGGUGAGCUUCGGCUCCCUGCC

Pri-racGUG (156 nt):
GGAGAGAAGCUCACCGCCGGUCCCAAGCCCGGAUAAAAUGGGAGGGGGCGGGAA
ACCGCCUAACCAUGCCGGUGAGCUUCGGCUCCCGGCGUGGACUGUAGAACACUGC
CAAUGCCGGUCCCAAGCCCGGAUAAAAGUGGAGGGUACAGUCCACGC

racGUG (40 nt):
AACCAUGCCGGUGAGCUUCGGCUCCCGGCGUGGACUGUAG

Pri-racBroccoli (209 nt)[198]:
GGGCCGCACUCGCCGGUCCCAAGCCCGGAUAAAAUGGGAGGGGGCGGGAAACCGC
CUAACCAUGCCGAGUGCGGCCGCGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUC
GAGUAGAGUGUGGGCUCGUGGCCGCGGUCGGCGUGGACUGUAGAACACUGCCAA
UGCCGGUCCCAAGCCCGGAUAAAAGUGGAGGGU

racBroccoli (96 nt)[198]:
AACCAUGCCGAGUGCGGCCGCGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUCGA
GUAGAGUGUGGGCUCGUGGCCGCGGUCGGCGUGGACUGUAG

ringBroccoli (53 nt):
CUUCGGCUUGUCGAGUAGAGUGUGGGCUCCUUCGGGAGACGGUCGGGUCCAAG

Before expressing the adapted racGUG construct, we reproduced the protocol of Litke and Jaffrey[198] with their published vector obtained from addgene.org (#124360). In Figure 18A, fluorescence microscopy shows cells positive for DFHBI-1T fluorescence using a GFP filter block. While the fluorescence is much weaker than a GFP expressed under a CMV promoter, it is still clearly present, as untransfected cells are devoid of any fluorescent signal. After RNA extraction, the transfected cells show a strong band migrating between cellular tRNA and 5S rRNA on a denaturing PAGE, indicating the production of a 96 nt RNA (Figure 18B). To test for circularity, we used the property of circular RNA to be more sensitive to different acrylamide concentrations than linear RNA. On a 10% PAGE, the band in question migrates above the tRNA, while it migrates below the tRNA on a 6% PAGE, indicating that the band is indeed circular (Figure 18B).

Expression of the racGUG construct in HEK293T cells followed by RNA extraction shows a strong 40 nt band on a denaturing PAGE which is comparable to the endogenous RNAs at higher concentrations (tRNA, 5S and 5.8S rRNA), shown in Figure 18C, lane 3. Untransfected cells are devoid of this band (Figure 18C lane 2). Comparison with linear and circular GUG as references gives no clear indication about circularization of the RNA, which is probably due to the altered ionic strength of the RNA extract sample affecting electrophoretic mobility of GUG.

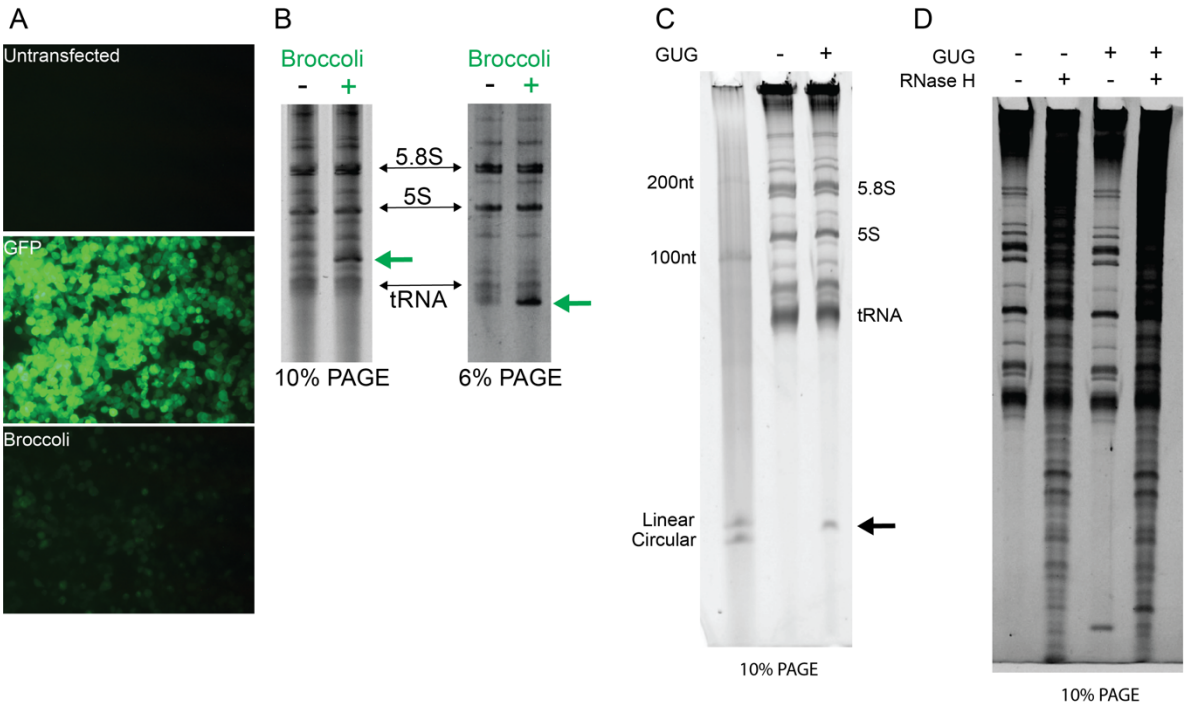

Figure 18: Expression of circular Broccoli and GUG in HEK293T cells. A: Fluorescence microscopy of DFHBI-stained untransfected cells (top), GFP expression as positive control (middle) and DFHBI-stained cells after transfection with the Broccoli plasmid (bottom). B: Total RNA extraction of Broccoli-transfected (+) and untransfected (-) cells. The altered sensitivity of the Broccoli band (green arrow) supports that the band is indeed circular. C: RNA extraction of GUG-transfected (+) and untransfected (-) cells, showing a lower band matching GUG references (arrow) on a 10% denaturing PAGE. D: RNase H-digest of the GUG-transfected RNA extract shows that the GUG band migrates upwards as expected from a linear RNA, creating a strong band which is absent in the cleavage products of the untransfected cells.

To test whether the expressed RNA is indeed circular, we performed RNase H cleavage in the total RNA extract. If racGUG is circular, a lower electrophoretic mobility is expected, while linear RNAs are cleaved into shorter products with higher electrophoretic mobility. Figure 18D shows a significant amount of cleavage products even for untransfected cells. For RNase H-cleaved transfected cells, a band slightly higher than the expressed GUG band can be seen, which is not present in the untransfected cell sample. This provides strong evidence for the expression of racGUG in HEK293T cells.

We furthermore produced racGUG by IVT and *in vitro* ligation with RtcB. Figure 19A shows that pre-racGUG is readily produced via cis-cleavage after transcription, giving rise to a 40 nt pre-racGUG band, and the 60 and 62 nt-bands of the flanking ribozymes. After ion-exchange HPLC purification and RtcB treatment, a band of slightly higher electrophoretic mobility is produced (Figure 19B). Upon RNase H cleavage, the ligated RNA produced a product of lower electrophoretic mobility, while the unligated RNA cleaves into two shorter products (Figure 19B). The cleavage guides can be seen below the 16 nt marker. The artefact band for the ligated and cleaved product (last lane) matching the longer unligated, cleaved sample cannot be explained, but does not discredit the upper band as evidence for successful circularization.

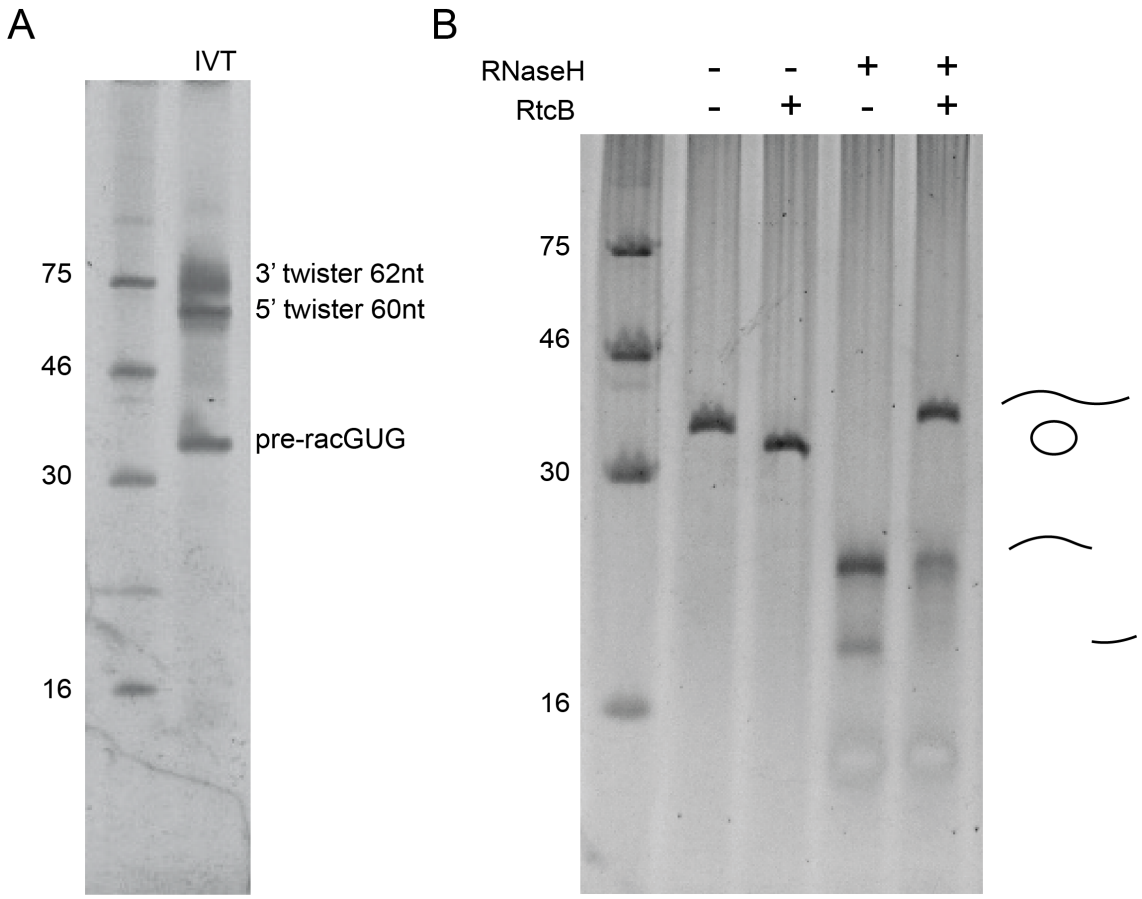

Figure 19: A: In vitro transcription of pri-racGUG, giving 3 bands by co-transcriptional cleavage of 3' and 5' ribozyme and pre-racGUG (40 nt). B: Circularization of racGUG. Treatment of pre-racGUG with RtcB ligase gives rise to a band with higher electrophoretic mobility. Internal cleavage with RNase H leads to a higher band for the ligated sample, while the unligated sample produces 2 lower bands.

In vitro production of a shorter, circular Broccoli

Next to GUG, the Broccoli RNA was also to be prepared by IVT and circularized with RtcB. As both IVT and ligation were of low yield and did not give conclusive results on whether racBroccoli was produced as a circular monomer in the correct fold, this approach was aborted (data not shown). Instead, we attempted to circularize Broccoli using T4 RNA ligase 2 (Rnl2), as shown by Chen and coworkers[199], which has been successfully executed for a shortened GUG construct (not shown). The lack of ligation stem and 5'-OH and 3'-phosphate allowed us to shorten the Broccoli sequence to 53 nt (Figure 20A) and use the RNA production method using RNase H cleavage of tandem transcripts[201] from paper I. To delineate the Rnl2-produced circular RNA from the Twister cleavage and RtcB ligation method (racRNA), we use the term 'RNA rings' as used by Chen and coworkers and thus call this construct ringBroccoli (53 nt).

It should be noted that there is no difference in principle between ringRNAs or racRNAs, apart from constraints in experimental production.

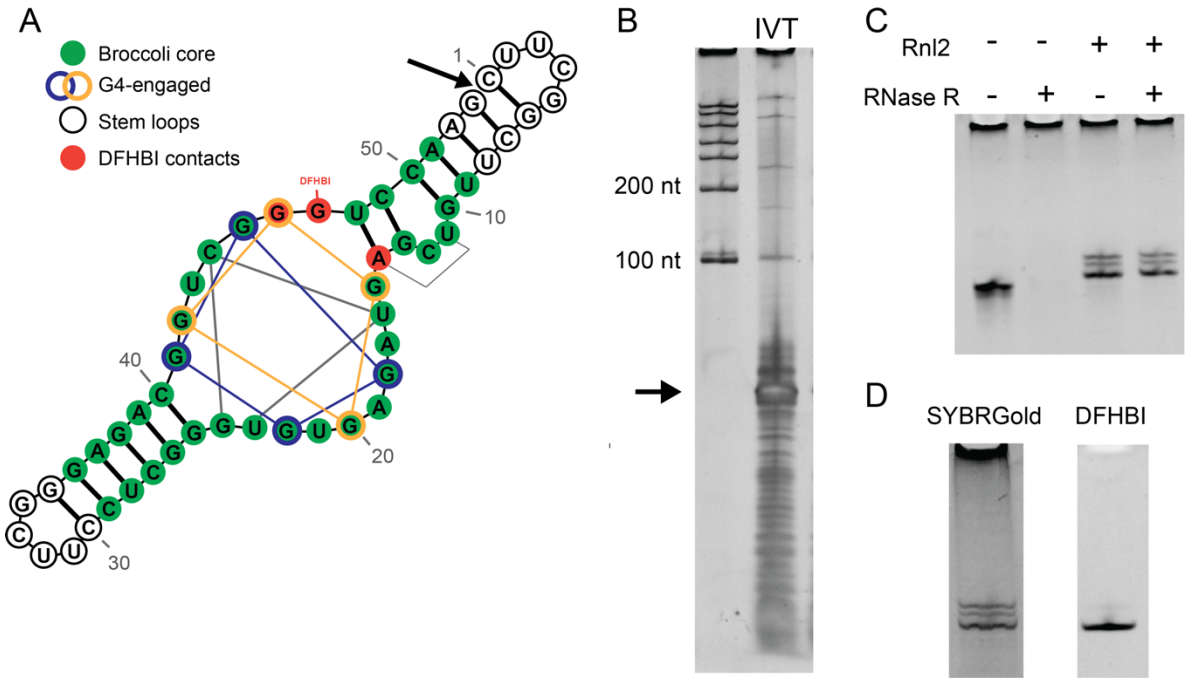

Figure 20: A: Secondary structure representation of ringBroccoli. The arrow indicates the Rnl2 ligation site. B: Denaturing PAGE analysis of cleaved transcription from tandem template, giving rise to a strong main band indicated by the arrow. C: HPLC-purified Broccoli (lane 1) yields 3 higher bands after ligation (lane 3). The unligated product is sensitive to RNase R-treatment (lane 2), while the ligated product is resistant to RNase R (lane 4), showing they are all circular. D: Only the lowest band of the 3 circular Broccoli bands stains for DFHBI.

Figure 20A shows the expected secondary structure of our Broccoli construct based on the solved structure by Warner and colleagues (pdb: 4TS2)[41]. The green nucleotides mark the crucial Broccoli nucleotides[43], white nucleotides have been added as a stem-loop scaffold. Lines outside the helix stems highlight non-canonical imino hydrogen bonds to form the tetrads, triplex lid and contact to DFHBI. G-quadruplex nucleotides from the $G4_1$ and $G4_2$ planes are highlighted with blue and yellow circles, respectively.

Figure 20B shows the successful IVT with simultaneous RNase H cleavage and a strong target band. Both low-molecular-weight impurities and a low-concentration uncleaved high-molecular-weight products can still be seen.

Treatment with Rnl2 produces 3 higher bands, which are all resistant to exonuclease RNase R, giving strong evidence of circularity (Figure 20C). In-gel staining with DFHBI shows that only the lowest band is able to bind the ligand DFHBI (Figure 20D). These 3 bands could not be collapsed into one band by reaction optimization including refolding, ligation in presence of folded or DFHBI-complexed RNA (data not shown). This indicates that the 3 bands arise in fact from 3 different molecules created in the ligation reaction.

We hypothesized these new fragments to come from nucleotide additions during the ligation reaction. To test the hypothesis, we used the RNase H cleavage guide from the IVT reaction (cleaving between G53 and C1), to probe whether the cleavage site is intact. RNase H requires sequence complementarity especially around the cleavage site, and inability to form an RNA-

DNA hybrid should render RNase H incapable of cleaving the higher bands. Figure 21 shows that indeed the lowest of the 3 bands moves to the height of the linear, unligated band, while the upper two bands stay in place. Further RNase R-digest confirms that the bands remain circular. It is thus far unclear where such base additions could come from. A possible explanation could be additions of ATP instead of the 3' end, which would add one or more adenosines into the ligation site of 5' and 3' end. It is still surprising that such a minor modification in the stem renders Broccoli incapable of binding DFHBI, as shown by the lack of fluorescence in Figure 20D. Further analysis with RNA sequencing or mass spectrometry could corroborate or challenge the conclusion based on the RNase H-cleavage experiment.

We deemed such a heterogeneous population unsuitable for in-cell NMR experiments and thus did not pursue the large-scale circularization of Broccoli any further. It would however be of general interest to elucidate the nature and mechanism of the aberrant Rnl2 activity.

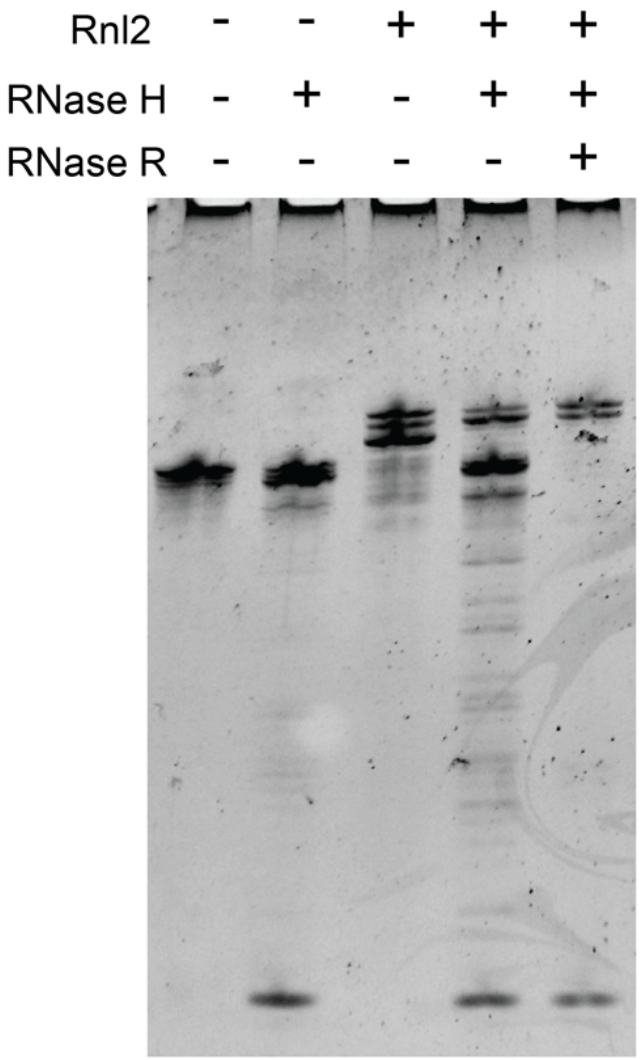

Figure 21: Assessment of intact ligation site sequence with sequence-specific RNase H digest. Linear Broccoli experiences some abberrant cleavage (lane 2). For circular Broccoli, the upper two bands resist RNase H cleavage, while the lowest band shows the same bands as RNase H-treated linear Broccoli (lane 4). The upper bands remain resistant to RNase R treatment, while the lower bands are digested, showing they are linear (lane 5). The lowest band in the gel are the RNase H cleavage oligos.

5.4.2 *In vitro* NMR of racGUG and ringBroccoli

Folding validation of racGUG by *in vitro* NMR

In order to prepare in-cell NMR experiments of racGUG, we compared the 40 nt RNAs pre-racGUG (linear, after removal of ribozymes) and racGUG (circular) with the original GUG construct of 27 nt, to validate the correct fold. As a read-out for comparable folding, we used imino protons measured by ^{1}H-1D NMR. Signals with matching same chemical shift are assumed to arise from the same sequence context. Figure 22 shows the imino region for the original GUG, pre-racGUG (linear) and racGUG (circular) and the resonance assignment for the original GUG as obtained by Katja Petzold. As expected, the conserved upper stem-loop shows good matches for the imino resonances and supports the assumption that the stem-loop is folded as expected, shown with dotted lines in Figure 22. Additionally, 8 imino signals can be counted for both pre-racGUG and racGUG, which accounts for the base pairs between C3 and C9 (G6-U30 gives rise to two imino protons). Also, the absence of new signals between linear and circular RNA indicates that no structural change occurred apart from closing the termini, rendering the loop between A34 and A2 unstructured. The broad bump-like signal between 11 and 11.5 ppm cannot be explained at this moment.

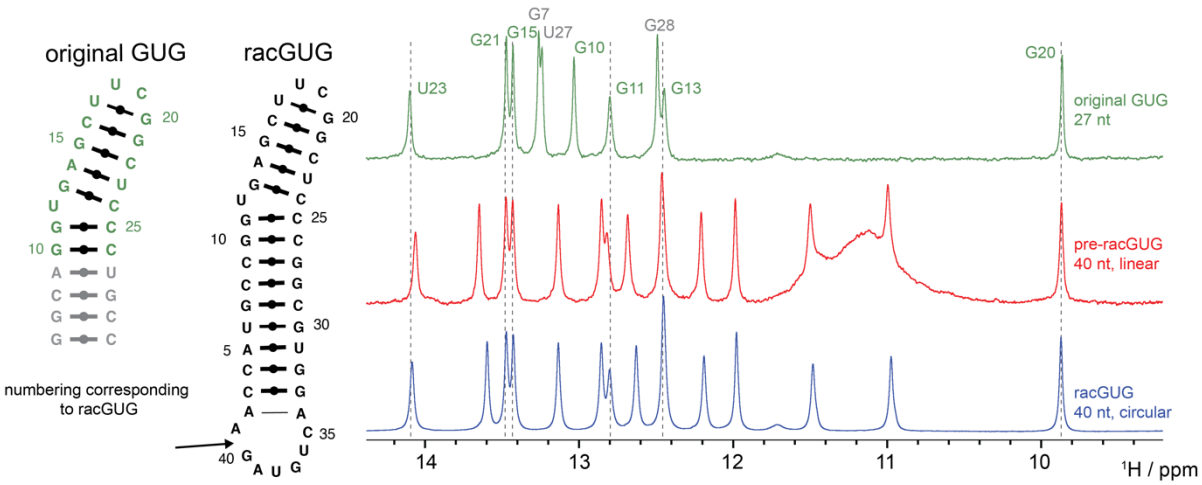

Figure 22: Imino spectra of different GUG constructs, measured at 298 K. Original GUG (27 nt), secondary structure shown on the left, next to the racGUG secondary structure. The arrow indicates where racGUG is ligated, and thus does not contain a closed loop for pre-racGUG. Assignment of the nucleotides conserved between original GUG and (pre-)racGUG are labeled green. The not conserved stem assignment is grayed. Nucleotides retaining the chemical shift are indicated with a dotted line, which are all conserved ones but G10, as expected. 8 new imino signals can be counted for pre-racGUG (red) and racGUG (blue), which does account for the entire new stem region. There is no additional signal between racGUG and pre-racGUG, indicating that the loop is unstructured between A34 and A2 also for the circular racGUG. The broad signal in the pre-racGUG signal between 11 and 12 ppm cannot be explained at this point.

Ligand-binding of Broccoli with ^{1}H-NMR and ^{19}F-NMR

The structure and ligand-binding of the Broccoli-precursor Spinach has been extensively studied with x-ray crystallography[41,45], fluorescence spectroscopy[42,47,202,203] among other methods. These results are seen as highly transferrable to Broccoli, as both contain the same core structure of 2 G-quadruplexes below the ligand-binding site, a triplex lid and a mixed tetrad below the G-quadruplexes and the few nucleotide mutations do not affects those

structures[43,44]. While the complex has been well-understood, the structure of apo-Spinach, in the absence of the fluorescent ligand, has barely been studied. Two accounts in the literature describe a barely compatible picture of apo-Spinach. Huang and colleagues found a folded, but slightly altered core structure from x-ray crystallography[45] (pdb ID 4KZE), which lacks the triplex lid and a adenosine in the mixed tetrad instead of a uridine base. Warner and colleagues performed a deuterium-exchange with ^{1}H-1D NMR, showing a significant stabilization of certain hydrogen bonds (presumably from the core structure) in presence of the ligand, but no stable imino signals in absence of the ligands[41]. Missing information on the deuterium exchange time precludes a comparison of the relative stabilities of the hydrogen bonds in the stems and core structures but suggests that the core is not folded (or at least severely destabilized) in the apo form. We wanted to use our access to NMR-scale amounts of Broccoli (linear, 53 nt) and two ligands (DFHBI and DFHBI-1T) to shed light on differences between the apo and holo states of Broccoli.

Comparison of the known Spinach crystal structure (4TS2[41]) with our 53 nt-Broccoli construct (Figure 20) allows some prediction on imino hydrogen bonds formed. Broccoli has 13 Watson-Crick base pairs on top and below the core structure (assuming a folded C1-G6 base pair, as shown in Figure 20A), which are expected to at least partially fold even in absence of a folded core, giving rise to up to 13 imino signals. Furthermore, we expect 2 imino signals in the mixed tetrad (U23 to U16 and G24 to C44), 8 imino signals for the 8 G-quadruplex-engaged guanosines, one from the triplex lid (U11 to A14) and one imino hydrogen bond from G47 to DFHBI. Furthermore, imino signals from the U31-G34 in the UUCG tetraloop (central loop) could arise and if folded despite not being ligated, also the equivalent between U2 and G4 (terminal loop). This sums up to a total of 27 imino signals, assuming a folded and DFHBI-bound core, and a formed C1-G6 base pair.

The apo structure (4KZE[45]) shows the same core, but lacks the imino hydrogen of U11 due to missing triplex lid, and the imino in the mixed tetrad (U23 to U16) due to the flipped-in adenosine. If the apo structure is indeed as shown in the crystal structure, the changes in imino signals are expected to be only minor.

We measured imino spectra with an imino-selective 1D-SOFAST experiment as a readout for overall structural similarity. All samples contained 100 mM K^{+} and 1 mM Mg^{2+} which are required for folding of Broccoli. Figure 23shows an overlay of the spectra for Broccoli in the DFHBI-1T-bound (blue) and apo state (red). The blue spectrum was scaled up by a factor of 5 to account for differences of concentration and number of scans. The holo state (blue) counts a total of 33 signals, including weak shoulder signals and excluding the signals at 9 ppm. 10 of those are stronger in the apo state (red arrows), indicating that they might arise from a

population in which DFHBI-1T is not bound to Broccoli, despite 2-fold excess of the ligand and known high on-binding rates and long residence time (>40 s)[43,48].

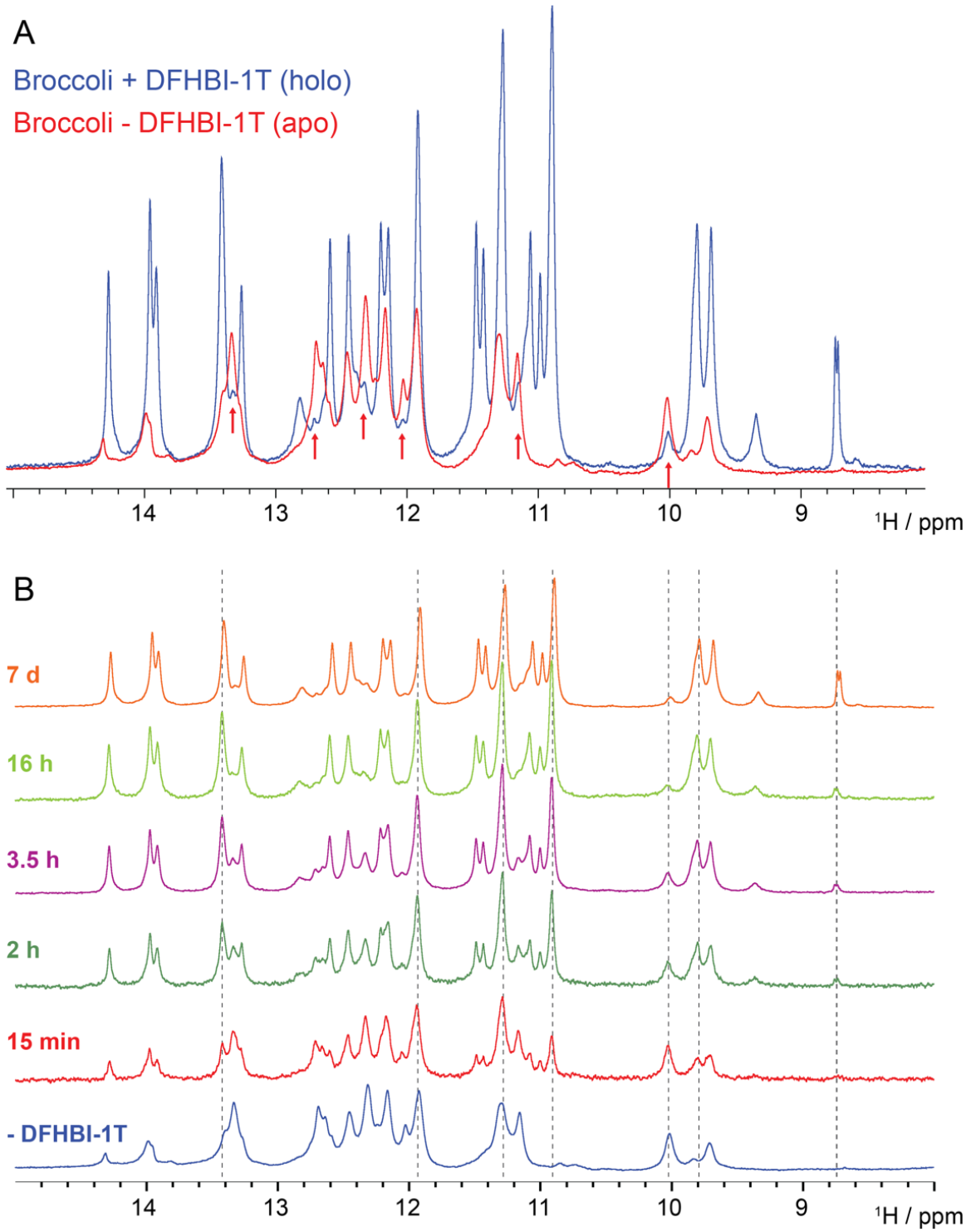

Figure 23: Impact of DFHBI-1T-binding on imino signals of Broccoli. A: Imino spectra for Broccoli with DFHBI-1T bound (blue) and without DFHBI-1T (red). The spectra were scaled based on number of scans and concentration. Red arrows indicate signals of the DFHBI-1T-containing sample which might arise from apo structures. B: Time course of DFHBI-1T binding. Clear changes can be seen already after 15 minutes of incubation. All spectra were measured at 298 K.

14 signals are unique to the holo state and do not arise in the apo state. The other 9 signals are stronger in the holo state (blue) but do have a weaker counterpart in the apo state (red).Furthermore, the signals can be clustered into structural context by known chemical shift values for helical context[60,72] and G-quadruplexes[204,205]. It can be seen, that much more signal

shared between apo and holo forms can be found in the higher chemical shift region above 12 ppm, where helical structures are expected. The G-quadruplex region between 10.5 and 12 ppm contains fewer apo signals, indicating that the core structure is severely destabilized in the absence of DFHBI-1T. Due to the general broader lines and lower intensity of the apo state, it can be concluded that the helical regions are also forming less stable hydrogen bonds, but appear to be formed.

Signals between 9 and 10 ppm indicate the formation both the central and terminal UUCG loops, which is where the G-U imino hydrogen is expected (determined at 10°C[33]). Next to the imino hydrogen bond, other unusual guanosine imino protons could resonate there, for example G26-A39, and G47-DFHBI-1T. The absence in the apo state makes the signal at 9.34 ppm a promising candidate for the DFHBI-1T-engaged hydrogen bond to G47. The two signals at ~8.7 ppm (holo state, blue) could arise from amino resonances, and their absence in the apo structure suggests that they are the N2-N7 amino hydrogen bond in G-quadruplexes. For a more detailed analysis of the structure, a resonance assignment would be needed.

A time-course experiment of DFHBI-1T binding, shown in Figure 23B shows how signals from the apo form decrease in intensity, supporting that the weaker signals in the sample with DFHBI-1T arise from a residual apo-Broccoli population. A few signals where clear changes can be seen have been highlighted with dashed lines.

We recorded ^{19}F-NMR of DFHBI-1T with and without the RNA, to assess whether the shifted resonances upon binding might serve as a good model system for ligand-detected in-cell studies. DFHBI-1T is a modified version of the ligand which exhibits brighter fluorescence and excitation and emission wavelengths closer to GFP and is thus advantageous for optical applications[202].

DFHBI-1T (Figure 24A) shows two signals, of which the triplet at -70.6 ppm originates from the trifluoroethyl group, while the aromatic fluorine atoms resonate as a doublet at -135.2 ppm (Figure 24A). The splitting arises probably from 2-bond ^{19}F-^{1}H couplings. In complex with Broccoli, a broadened signal at -70.3 ppm can be seen, which corresponds most likely to the shifted trifluoroethyl group (F_{Et}). No shift can be seen for the aromatic fluorine atoms (F_{aro}). The experiment with DFHBI-1T in excess (290 µM Broccoli, 600 µM DFHBI-1T) cannot show if the F_{aro} signals merely do not shift or disappear upon binding. To confirm the shifted F_{Et} signals and probe the effect on F_{aro}, we performed the same experiment with the ligand DFHBI, which lacks the trifluoroethyl group in substoichiometric amounts (150 µM Broccoli, 100 µM DFHBI). Figure 24B shows that indeed the F_{Et} and the broadened signal is missing for the Broccoli-DFHBI complex and the F_{aro} signal vanishes upon binding to Broccoli.

This finding is somewhat surprising, as rotational constriction due to the increased molecular size alone does not warrant such an expectation. Broccoli of 53 nucleotides is well within the size limit of solution NMR and gives clear proton signals, as can be seen in Figure 23. The F_{Et} signals shift and broadened, which is the expected effect such binding should have. The ligand is intercalated between a G-quadruplex and the triplex lid, with the phenyl ring facing into the

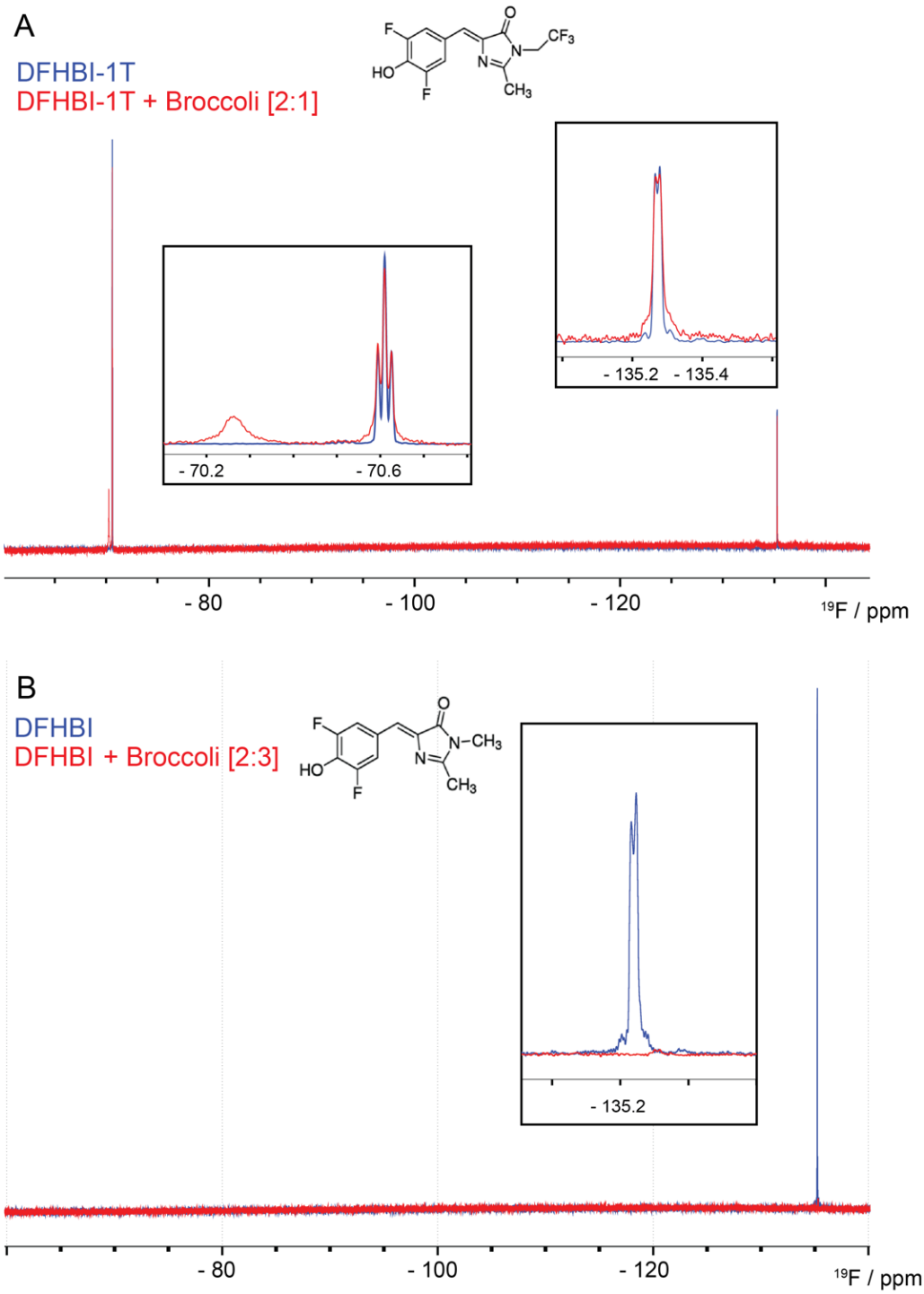

Figure 24: ^{19}F-NMR of two Broccoli-bound ligands. A: DFHBI-1T (structure shown) without (blue) and with Broccoli (red) and zooms of the respective spectral regions. DFHBI-1T is in excess over Broccoli. The zoom shows that the signal of the trifluoroethyl group (left) shifts and broadens when bound to DFHBI-1T. The signal of the aromatic fluorine atoms does not shift, which could be because it is broadened beyond detection or is not affected by the binding pocket. B: DFHBI lacking the trifluoroethyl group (structure shown) without (blue) and with Broccoli (red) in substoichiometric amounts reveals that no signal of the aromatic fluorine atoms in DFHBI can be detected when binding to Broccoli. The lacking trifluoroethyl group also provides an unambiguous resonance assignment for the fluorine groups.

pocket and the imidazolone group (including trifluoroethyl moiety) towards the solvent. Thus, a stronger shift was expected for the F_{aro} signals. Interestingly, Truong and colleagues also found the F_{aro} signal of DFHBI-1T to disappear as well when binding to the recently found Squash aptamer, while the F_{Et} group also moves downfield by ~0.5 ppm[206]. This is somewhat

surprising, because Spinach/Broccoli and Squash have very low sequence identity and share none of the defining structural features like G-quadruplexes or the triplex lid. The binding pocket lacks the general planar structure of Spinach/Broccoli but is composed of four strands of a fork-like structure, and polar contacts to surrounding nucleotides are very dissimilar.

5.4.3 Overexpressed racGUG as a target for in-cell NMR studies

The use of circular RNAs for in-cell NMR should mainly increase their stability, allowing for longer measurement times, as this was the bottleneck in previous experiments[164,167]. To test whether circular RNAs indeed have increased stability, we incubated racGUG and a shorter, linear GUG construct in HEK293T lysate for 0, 1 and 6 hours and extracted the RNA using the miRvana kit (Thermo) before comparison of RNA stability with denaturing PAGE. Figure 25 shows that the main band of racGUG is still largely present after 6 hours, while the linear construct essentially disappeared after 6 hours. As expected from endonucleotytic cleavage, a higher band is created in the degradation process of racGUG. These results agree with findings from Jaffrey and Litke on the increased intracellular stability of racBroccoli from Actinomycin D assays[198].

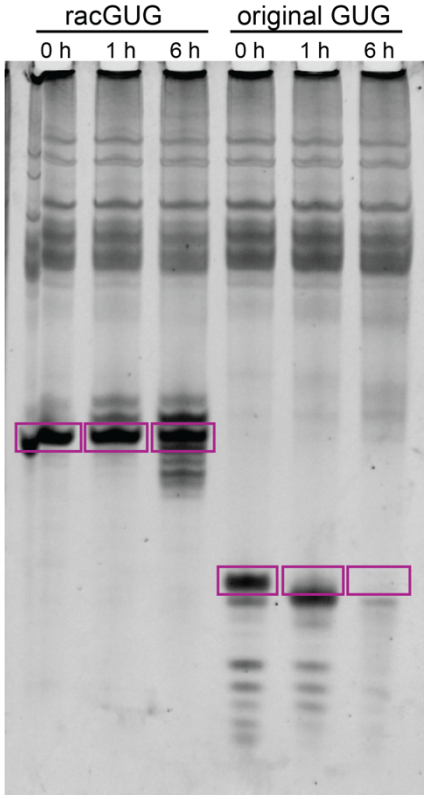

Figure 25: Stability of circular racGUG in HEK293T lysate compared to the original GUG (linear). Analysis on a denaturing PAGE shows rapid decay of the linear RNA (right), while the circular RNA remains stable over 6 hours, but shows first signs of degradation.

After assessing the increased stability, we expressed racGUG in HEK293T cells in large scale for in-cell NMR experiments. Several experiments measured with ^{1}H-1D NMR different water-suppression schemes (excitation sculpting, 1-1 jump-and-return, imino-selective SOFAST) did not yield discernible imino resonances. We hypothesized that racGUG, even if expressed to detectable concentrations, might be rotationally restricted and therefore broadened

out beyond detection. This is supported by Litke and Jaffrey, who found racBroccoli to localize both to the cytoplasm, and to droplet-like Cajal bodies in the nucleus[198].

To test whether stable racGUG could be freed from such restriction by lysis and protein digest, we followed the protocol by Schlagnitweit, Sandoz and colleagues which allowed them to detect a chemically modified RNA by ^{31}P NMR[181]. The 1D-SOFAST imino spectra (8192 scans, 40 minutes) of the in-cell sample (blue) contains no discernible signals, which could belong to racGUG (green) (Figure 26). After lysis and proteinase K-digestions, clear broadened and overlapping signals appear after 8192 scans, but no specificity to racGUG can be concluded. The signals could arise from endogenous RNA which was released from bound protein by protein digestion and thus has shorter correlation times to be detectable by solution NMR.

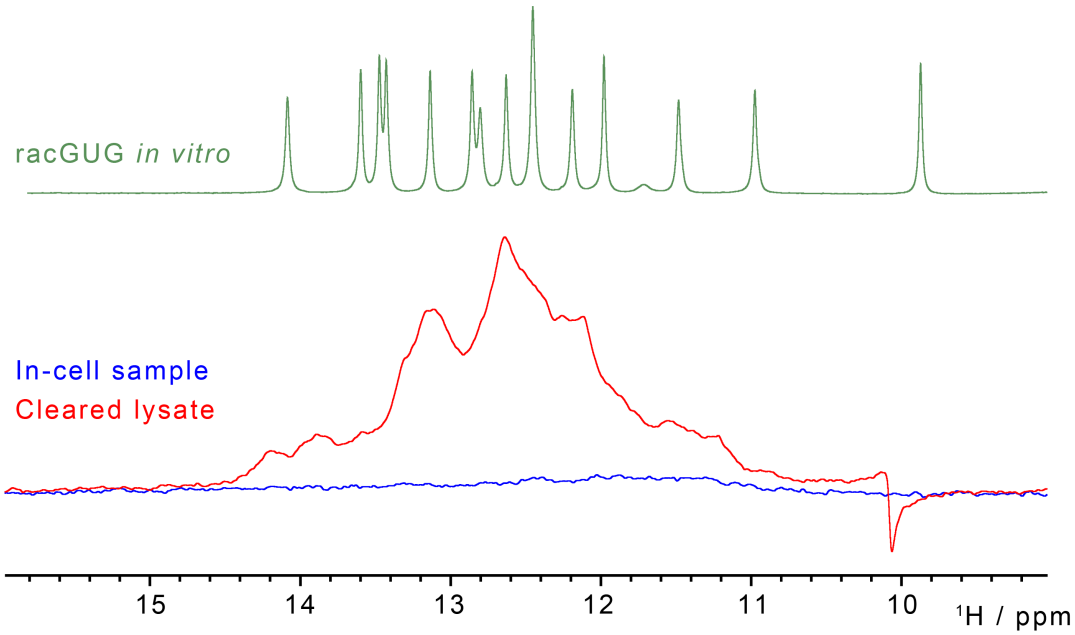

Figure 26: In-cell NMR of racGUG expressed in HEK293T cells and an in vitro spectrum for comparison. Neither in the in-cell sample (blue) nor in the digested lysate (green) can clear signals for racGUG (blue) be identified. The blue and red 1D-SOFAST spectra have both been recorded with 8192 scans.

Based on these results, we conclude that overexpression of RNA with the TORNADO method in its current form is not a feasible alternative to exogenous introduction like electroporation or SLO transfection. The intracellular concentration of the resulting RNA could be enhanced by increasing the number of racGUG repeats from a single plasmid, similar to the RNase H method shown in paper I[201]. It is possible that the promoter could be engineered in a way that it would produce less RNA in Cajal bodies or other phase-separated droplets. Also, sensitivity-enhanced NMR like in-cell DNP NMR[181,207,208] or similar methods could still give access to endogenously expressed RNA. It will however depend on the research question whether the reduced resolution of solid-state DNP will be compensated with high sensitivity. Current developments towards lower temperatures and faster spinning frequencies are promising tools for to expand the range of possible targets for in-cell NMR, including intracellular expressed RNAs at low concentrations.

6 CONCLUSIONS

6.1 Paper I – RNA production by tandem transcription and RNase H cleavage

We showed the preparation of RNA in large scale and high purity by transcribing tandem templates from a linearized plasmid and successive RNase H cleavage to generate target-length RNA between 20 and 75 nucleotides. The yield and purity of the obtained RNA products are superior to those of linear templates because RNase H produces more defined cleavage site termini than T7RNAP during run-off transcription.

In essence, the method is highly similar to cis-cleavage with flanking ribozymes, which is also the basis for the TORNADO method shown in the preliminary data of chapter 5.4. Both allow for a custom initiation sequence and benefit from high specificity of enzymatic (or ribozymatic) cleavage. As the initiation event appears to be the kinetic bottleneck, our approach produces several copies of the target RNA, instead just the one copy which is flanked between two ribozymes. The relevant difference here is the requirement of an additional enzyme and the cleavage guide, while these are included in the RNA sequence in form of the ribozymes for the cis-cleavage approach. Fortunately, RNase H can be expressed at high yield in *E.coli* and shows unprecedented stability and activity even after years of storage at -80°C. To give an example, we have used only ~20% of a 1.5 L expression over 3 years of developing the method, quantification experiments for the publication and several hundred mL of IVT for NMR sample preparation. The cost of the cleavage guide is not negligible but can be recovered during HPLC and is dwarfed by the cost of isotope-labeled nucleotide for the production of NMR samples.

Even though the method has not been cited in the production of RNA samples as of now, we have been contacted by several researchers wanting to establish the method in their laboratory. Together with the abundant application in our own group, this showcases the potential of further IVT development for specific applications, for example the production of NMR samples as opposed to RNA-interference experiments in cell culture.

An important case must be made for the relevance of any IVT approach, especially when it comes to scalability of the method. RNA of high yield and purity can be made through creative approaches using different templates, cleavage and ligation methods. Not all of them are easily scalable to the amount needed for NMR samples. For example, Wang and colleagues used rolling circle transcription (RCT) from a single-stranded DNA circle and simultaneous RNase H cleavage, highly similar to our own method[200]. The preparation of such circular templates however demands another step of ligation which itself requires a splint and ligase, further complicating the scale-up of the protocol. Plasmid DNA instead can robustly be amplified and linearized for hundreds of mL worth of IVT in a single day. Similarly, we showed that our method does work when using circularized dsDNA as rolling-circle template, but ultimately it fails to be useful for NMR sample preparation due to the very low yield of RCT template. It is a good example for a method that does not only have to be ‘elegant’ but must pass higher standards of efficiency and scalability, and our methods presented in paper I does fulfil those criteria well.

6.2 Paper II – Publication of a video protocol

Even though papers are published with rigorous detail, it can often be difficult to reproduce them in practice, which would preclude of course any meaningful impact of such publications. In order to make our work more accessible, we decided to explore the new video publications format of the journal of visual experiments (JoVE) by showing one of the group's typical experiments from RNA preparation to acquisition of $R_{1\rho}$ relaxation dispersion experiments. As this is essentially a protocol publication with a supplementary video file, there are no new results and conclusions that have not previously been published respective under the lead of my colleague Judith Schlagnitweit[57] and myself[201].

The focus of the filmed segments were the non-standard steps of the IVT, especially heating the large-scale sample in the microwave (which is generally received with some reservations by fellow researchers), bubble-free filling of a Shigemi tube and setting up the relaxation dispersion measurements. The natural shortcomings of such an extensive experiment become apparent when in the protocol, as many steps have to be expected as prerequisite knowledge of the reader. For example, amplification of plasmid DNA is expected to be known, but may come with many implicit steps for the NMR spectroscopist lacking biochemistry training. Similarly, resonance assignment of the RNA must exist already for meaningful studies of conformational dynamics to be performed, which implies that another NMR sample of the same RNA already exists. The modularity of the protocol is a strength here: RNA prepared with the presented IVT method can be used for other analyses, and the relaxation dispersion experiments can be performed with NMR samples of any source, given they are well-folded and of sufficient purity.

6.3 Paper III – site-specific labeling of RNA

Resonance overlap is one of the main problems while solution NMR spectroscopy keeps pushing the size-limit to ever larger RNA molecules. This can only partially be solved with higher field-strengths and better resolution, because the low chemical variability of RNA results in low chemical shift dispersion and thus resonances of different nucleotides can share the same chemical shift value. Selective labeling schemes are a very effective method to combat this, and nucleotide-specific labeling trumps most other methods, as it eliminates the need for any resonance assignment (but comes at the obvious shortcoming of information on one nucleotide only).

We have shown here another instance of simple site-specific labeling of RNA with only standard enzymatic methods, which makes it accessible to the standard biochemistry laboratory without the need of custom solid-phase synthesis. Such methods have been published in very similar approaches before but seem to be finding only little application in the scientific literature. Our study shows how such a simple approach can solve assignment problems due to resonance overlap and serendipitously revealed a second conformation that was previously hidden. We hope to make such an approach more attractive to fellow researchers and help them overcome assignment problems.

6.4 Circular RNA for in-cell NMR experiments

In-cell NMR with RNA in human cells have been rarely reported in the literature, and apart from the 14 nt model hairpin were limited to G-quadruplexes[173] and the 2'-dG-binding aptamer[167], which could only be detected via the isotope-labeled ligand. There are not many conceptual differences between DNA and RNA, that would justify the expectation that RNA would be inherently more difficult to perform in-cell NMR with, than with DNA, other than potentially reduced intracellular stability and significantly higher cost for solid-phase synthesis of RNA.

We attempted to create the first system of RNA overexpression for in-cell NMR experiments and deemed the TORNADO method a well-suited and robust system for the task. While it was in fact straight-forward to reproduce and adapt to the racGUG construct, the expected results, i.e., clear imino signals after overexpression could not be detected. Lysis and protein digestion did still not yield an NMR signal, which could be evidence for the RNA not reaching high enough concentrations with the overexpression method but could potentially be masked by the strong signal from endogenous RNA after lysis and protein digestion. Further experiments for absolute quantification of the in-cell NMR sample could clarify where exactly the issue with racRNA lies.

In its current form, the TORNADO method seems not be a promising approach to produce in-cell NMR samples in liquid state, as no signal could be detected. Further engineering of expression strategies could potentially make them feasible targets for solution NMR. DNP applications could have the sensitivity and tolerance to large RNA molecules that solution NMR lacks, and thus be the right analysis method for in-cell NMR samples produced with TORNADO. The low sample volume needed for DNP NMR does also significantly decrease the price tag on isotope-labeled expression.

The *in vitro* production of circular RNA would allow the exogenous introduction via electroporation or SLO transfection and have the advantage of an increased lifetime. This could serve as a promising model system for not only stem-loop RNA, but also highly structured RNA fragments, here exemplified with the Broccoli aptamer. The production of both racGUG and ringBroccoli was successfully shown with RtcB ligase and T4 RNA ligase 2, respectively. However, the demands on scalability and yield that were discussed in chapter 6.1 are only more relevant when it comes to the production of in-cell NMR samples, as they require several times the amount of RNA than *in vitro* NMR. RtcB-ligated racGUG could be produced at in-cell NMR quantities, but it took several weeks to prepare the amount necessary, as intermediate HPLC steps were required. This project, with the aim to measure conformational dynamics in living cells, was put on hold to focus on different approaches.

Rnl2-ligated ringBroccoli did produce misfolded side products, which renders the circular construct unsuitable for in-cell NMR studies. Of course, Broccoli could still be studied as a linear construct, to shed more light on whether the core structure shows stronger folding in the crowded intracellular environment, since the apo and holo structure show such clear differences

in their imino regions. We found that there the stem structures seem to be folded also in the absence of the ligand, while the core seems severely destabilized, due to the lack of most of the imino protons. Our data adds to the hydrogen-deuterium-exchange experiment by Warner and colleagues[41], which found that the Spinach core structure is much more stable than the stems. Clearer conclusions cannot be made at this point, as no resonance assignment is available. With a resonance assignment, time-resolved hydrogen-exchange experiments could shed light on the relative stabilities of stems and core structures in presence and absence of DFHBI. It would be of high interest to compare the stabilities of especially the apo structure to conditions of molecular crowding and ultimately the cellular interior. Beyond Spinach and Broccoli, the folding and stability of such small molecule-binding pockets are of high interest to understand the druggability of RNA fragments[209].

Furthermore, we found that ligand-detected ^{19}F-NMR of the Broccoli-DFHBI complex in cells does not seem feasible but should be performed with a derivate like DFHBI-1T, since the aromatic fluorine resonances disappear upon binding, while the trifluoroethyl group of DFHBI-1T shows a clear shift (Figure 24).

Lastly, we unexpectedly found an aberrant function of Rnl2, in which it modified the ligation site to the extent that the Broccoli structure was unable to bind DFHBI. It would be interesting to investigate (a) which properties in the Broccoli construct leads to such aberrant protein function and (b) how such a loop modification impacts the Broccoli core structure while this should be outside the conserved region (supposed it is the only affected site). This could shed light on unknown mechanisms of such a commonly used ligase and highlight issues in other applications.

To conclude, even though no successful in-cell NMR of RNA has been carried out in this thesis, we have made important steps towards new RNA systems and targets for such studies, that can hopefully soon be solidified and explored further.

7 POINTS OF PERSPECTIVE

In a very crude way, the work of my PhD studies can be divided in developing RNA sample preparation methods for NMR spectroscopy (and in principle biophysical analysis in general), and in-cell NMR-related experiments ranging from cell culture, TORNADO expression to the NMR measurements and many complementary methods. As the reader may have noticed by now, all publications relate to the sample preparation part, and none to the in-cell NMR experiments. I briefly want to discuss the latter part with respect to what the field hoped to gain from in-cell NMR, and where I think it will be going in the future.

In the PhD studies, we luckily do not only learn from our own mistakes and successes, but also the ones from our peers and the literature. Even though the failures may often not be communicated in publications, we can get a clearer idea from talking to fellow scientists on conferences and courses. At first glance, in-cell NMR seems like a rather straight-forward approach, with consists basically of growing cells, transfecting them with the nucleic acid sample and measuring the collected cells with NMR spectroscopy. The only 18 published papers over 13 years already indicate that either there is not much interest in the community, or it is not as simple as it seems. I have been fortunate to attend many conferences between 2017 and 2019 (i.e., before the COVID-19 pandemic), and have met many fellow PhD students pursuing in-cell NMR projects, and none of them has published their work by now, March 2022. As discussed in chapter 2.3, certain strategies for in-cell NMR experiments seemed particularly successful, which is summed up with: DNA samples, ^{19}F-labels, and G-quadruplexes. These 3 keywords make up most of the published literature. In retrospect, I can see that with focusing on RNA, and potentially their ^{1}H-detected conformational dynamics, we faced an uphill battle. Despite not succeeding in the attempts discussed in this thesis have learned a lot about the challenges firsthand.

It appears that we are so far limited to choosing the suitable NA system for the in-cell NMR method we have and cannot apply the in-cell NMR method to any NA system. One could extend the proverb: “If you’re a hammer, everything looks like a nail, and most walls are made of concrete.” A robust way to execute in-cell NMR experiment, especially in a way that gives biologically relevant information on NA structure has not been found yet.

If we are working under the assumption that the intracellular environment alters NA structure, can we then tolerate that our signal might come from NA located in the cytoplasm, nucleus and different organelles? Would the resulting spectrum not be the sum of all of those altered structures? Interestingly, what we have not really seen in in-cell NMR experiments are completely changed structures (apart from G-quadruplex topology) compared to the transfected samples. We did, however, learn also learn an important fact about mimics of the crowding effect: PEG repeatedly fails to reproduce the intracellular conformation and should probably no longer be used as an artificial crowding agent. Also, we should keep in mind that most cellular RNAs act through proteins in one way or another, which is something that has not yet been successfully studied with in-cell NMR.

NMR spectroscopy is a fascinating technique that sounds like it should belong to the realm of science fiction. It is certainly one of the methods that Galison described to drive scientific progress through technological advancement, rather than a paradigm shift of ideas. Hopefully, some time we will be at a point where we can execute the high-resolution experiments and *a priori* structure solving to in-cell samples in a biologically meaningful way. The groundwork has been laid by people like Robert Hänsel-Hertsch, Lukaš Trantírek and Hong-Ling Bao, and is now expanded upon by many more people. Until the community solved practical problems like intracellular location, measurement stability and more, the method will rely on many complementary methods like circular dichroism, *in vitro* NMR, and fluorescence microscopy. This is not a weakness, but necessary for the field to establish biological relevance. Such experiments will help us to better understand the inner workings of the cell and potentially, after being driven by technologies like in-cell NMR, we can witness another paradigm shift that paints the molecular basis for living systems in a different way.

8 ACKNOWLEDGEMENTS

Despite such a PhD being a lot of independent and individual work, this journey was by far not lonely, neither from a personal nor a professional sense. I want to spend a few words on people that were central to me in the last 5 years, whether they were involved in the projects or not.

First and foremost, I want to thank **Katja Petzold** for all the supervision she has given me, and for building such a fantastic research group. Second, I would like to thank my co-supervisor **Emma Andersson** who has been insightful and helpful whenever I asked for help.

I want to thank past and present members of the **Petzoldlab**, and name a few that were especially important in the last years: The original in-cell NMR gang: **Ileana**, **Judith** and **Sarah**, and the new in-cell NMR group: **Henry**, **Reiner**, **Rubin** and **Julian**. The PhD students who have paved the path for me and my successors: **Lorenzo**, **Hampus** and **Luca**. Your sacrifices and legacies have not been forgotten. My friend **Lara**, who has been a valuable companion on our shared journey. The postdocs, who have been inspiring discussion partners over the years: **Emilie**, **Maja**, **Walter** and **Christian**. The students I have supervised: **Raluca**, **Cenk**, **Henry** and **Sabrina**. You taught me more than you were aware, and I am grateful for how you challenged me, and helped me grow as a person and scientist.

Fellow coffee break and roof beer partners from the Quarter B9: **Natali**, **Igor**, **Marco**, **Ioanna**, **Ian**, **Allie**, **Sammy**, **Simon** and many more. Especially do I want to thank certain Postdocs, who have been teaching and guiding me selflessly, and were always there to brainstorm to what was going on in my mind: **Derrick**, **Giulio**, **Chris** and **Çagla**. You always showed me light when I could not see it.

I send lots of love to my Swedish friends (~50% Swedes is as good as it gets): **Miloš**, **Karen**, **Žiga**, **Johan&Johan**, **Stelia** and **Ann-Helen**. Thank you for taking me into your group with open arms, and for the many fantastic vacations and memories we created together. My FRN crew **Chris** and **Philipp**: the regular conversations and discussion with you mean the world to me. Friends from around KI have been invaluable: **Gabo**, **Javi**, **Gabi**, **Katharina**, **Dörte&Alvaro**, **Hannes**, **Leonie**, **Jess**, **Martina** and especially **Ashley** and **Carol**, both of you are an inspiration every day.

Lea and **Philipp**, your friendship was a source of stability through rough patches, for happy times, when playing board games or deep diving into science and society. These years would have not been the same without you.

Unzählbarer Dank geht zurück nach Deutschland, zu Freunden und Familie die 'nach Hause fahren' wirklich zu einem 'Zuhause' gemacht haben. Besonders meinen **lieben Eltern** möchte ich danken, denn ohne Eure jahrzehntelange Ermutigung meiner Neugier und meines Wunsches nach Wissen und höherer Ausbildung wäre ich nie so weit gekommen. Großer Dank gilt auch Freunden für andauernde Freundschaft, besonders **Patrick**, **Sebastian**, **Max M.**, **Max H**. samt Entourage und meiner lieben Schwester und wertvollen Freundin **Steffi mitsamt Familie**.

Sist vill jag tacka **Lena**, för att hon är en trygg hamn i vilken situation som helst och en aldrig sinande källa av lugn och frid.

9 REFERENCES

1. Kuhn, T. S. *The Structure of Scientific Revolutions (Foundations of the Unity of Science, Vol. 2, No. 2)*. 210 (University of Chicago Press, 1970).

2. Potthast, T. Paradigm shifts versus fashion shifts? Systems and synthetic biology as new epistemic entities in understanding and making "life". *EMBO Rep.* **10**, S42-5 (2009).

3. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

4. Dyson, F. J. History of science. Is science mostly driven by ideas or by tools? *Science* **338**, 1426–1427 (2012).

5. Mauger, D. M. *et al.* mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci USA* **116**, 24075–24083 (2019).

6. Jonas, H. *Technik, Medizin und Ethik. Zur Praxis des Prinzips Verantwortung.* (Suhrkamp, 1987).

7. Schlesinger, A. P., Wang, Y., Tadeo, X., Millet, O. & Pielak, G. J. Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.* **133**, 8082–8085 (2011).

8. Benton, L. A., Smith, A. E., Young, G. B. & Pielak, G. J. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry* **51**, 9773–9775 (2012).

9. Petraccone, L., Malafronte, A., Amato, J. & Giancola, C. G-quadruplexes from human telomeric DNA: how many conformations in PEG containing solutions? *J. Phys. Chem. B* **116**, 2294–2305 (2012).

10. Miller, M. C., Buscaglia, R., Chaires, J. B., Lane, A. N. & Trent, J. O. Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.* **132**, 17105–17107 (2010).

11. Hänsel, R., Löhr, F., Trantirek, L. & Dötsch, V. High-resolution insight into G-overhang architecture. *J. Am. Chem. Soc.* **135**, 2816–2824 (2013).

12. Manna, S., Sarkar, D. & Srivatsan, S. G. A Dual-App Nucleoside Probe Provides Structural Insights into the Human Telomeric Overhang in Live Cells. *J. Am. Chem. Soc.* **140**, 12622–12633 (2018).

13. Bao, H.-L., Liu, H.-S. & Xu, Y. Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.* **47**, 4940–4947 (2019).

14. Hänsel, R. *et al.* The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **39**, 5768–5775 (2011).

15. Keeler, J. *Understanding NMR spectroscopy*. (2004).

16. Bloch, F. Nuclear Induction. *Phys. Rev.* **70**, 460–474 (1946).

17. Besghini, D., Mauri, M. & Simonutti, R. Time domain NMR in polymer science: from the laboratory to the industry. *Appl. Sci.* **9**, 1801 (2019).

18. Koos, M. R. M. & Luy, B. Polarization recovery during ASAP and SOFAST/ALSOFAST-type experiments. *J. Magn. Reson.* **300**, 61–75 (2019).

19. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33**, 199–211 (2005).

20. Schulze-Sünninghausen, D., Becker, J. & Luy, B. Rapid heteronuclear single quantum correlation NMR spectra at natural abundance. *J. Am. Chem. Soc.* **136**, 1242–1245 (2014).

21. Schnieders, R., Keyhani, S., Schwalbe, H. & Fürtig, B. More than Proton Detection-New Avenues for NMR Spectroscopy of RNA. *Chem. Eur. J* **26**, 102–113 (2020).

22. Nußbaumer, F., Plangger, R., Roeck, M. & Kreutz, C. Aromatic 19 F-13 C TROSY-[19 F, 13 C]-Pyrimidine Labeling for NMR Spectroscopy of RNA. *Angew Chem Int Ed Engl* **59**, 17062–17069 (2020).

23. Zangger, K. Pure shift NMR. *Prog. nucl. magn. reson. spectrosc.* **86–87**, 1–20 (2015).

24. LeBlanc, R. M., Longhini, A. P., Le Grice, S. F. J., Johnson, B. A. & Dayie, T. K. Combining asymmetric 13C-labeling and isotopic filter/edit NOESY: a novel strategy for rapid and logical RNA resonance assignment. *Nucleic Acids Res.* **45**, e146 (2017).

25. Fürtig, B. *et al.* Time-resolved NMR studies of RNA folding. *Biopolymers* **86**, 360–383 (2007).

26. *Chemical biology of nucleic acids*. (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-54452-1.

27. Steitz, T. A. A mechanism for all polymerases. *Nature* **391**, 231–232 (1998).

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

29. Saenger, W. *Principles of nucleic acid structure*. (Springer New York, 1984). doi:10.1007/978-1-4612-5190-3.

30. Bramsen, J. B. & Kjems, J. Development of Therapeutic-Grade Small Interfering RNAs by Chemical Engineering. *Front. Genet.* **3**, 154 (2012).

31. Vester, B. & Wengel, J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* **43**, 13233–13241 (2004).

32. Noronha, A. M. *et al.* Synthesis and biophysical properties of arabinonucleic acids (ANA): circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA.RNA hybrid duplexes. *Biochemistry* **39**, 7050–7062 (2000).

33. Fürtig, B., Richter, C., Bermel, W. & Schwalbe, H. New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *J. Biomol. NMR* **28**, 69–79 (2004).

34. Borkar, A. N., Vallurupalli, P., Camilloni, C., Kay, L. E. & Vendruscolo, M. Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop. *Phys. Chem. Chem. Phys.* **19**, 2797–2804 (2017).

35. Baronti, L. Conformational dynamics in microRNAs: The example of miR-34a targeting Sirt1 mRNA. (Karolinska Institutet, 2020).

36. Sigler, P. An analysis of the structure of tRNA. *Annu. Rev. Biophys. Bioeng.* **4**, 381–527 (1975).

37. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).

38. 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).

39. Guiset Miserachs, H., Donghi, D., Börner, R., Johannsen, S. & Sigel, R. K. O. Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *J. Biol. Inorg. Chem.* **21**, 975–986 (2016).

40. Staple, D. W. & Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **3**, e213 (2005).

41. Warner, K. D. *et al.* Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol.* **21**, 658–663 (2014).

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

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

44. Savage, J. C., Davare, M. A. & Shinde, U. Subtle sequence variations alter tripartite complex kinetics and G-quadruplex dynamics in RNA aptamer Broccoli. *Chem. Commun.* **56**, 2634–2637 (2020).

45. Huang, H. *et al.* A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **10**, 686–691 (2014).

46. Neubacher, S. & Hennig, S. RNA Structure and Cellular Applications of Fluorescent Light-Up Aptamers. *Angew Chem Int Ed Engl* **58**, 1266–1279 (2019).

47. Dao, N. T. *et al.* Photophysics of DFHBI bound to RNA aptamer Baby Spinach. *Sci. Rep.* **11**, 7356 (2021).

48. Han, K. Y., Leslie, B. J., Fei, J., Zhang, J. & Ha, T. Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. *J. Am. Chem. Soc.* **135**, 19033–19038 (2013).

49. Li, X., Kim, H., Litke, J. L., Wu, J. & Jaffrey, S. R. Fluorophore-Promoted RNA Folding and Photostability Enables Imaging of Single Broccoli-Tagged mRNAs in Live Mammalian Cells. *Angew Chem Int Ed Engl* **59**, 4511–4518 (2020).

50. Li, X., Wu, J. & Jaffrey, S. R. Engineering fluorophore recycling in a fluorogenic RNA aptamer. *Angew Chem Int Ed Engl* **60**, 24153–24161 (2021).

51. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55 (2008).

52. Zok, T. *et al.* RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res.* **46**, W30–W35 (2018).

53. Antczak, M. *et al.* RNAvista: a webserver to assess RNA secondary structures with non-canonical base pairs. *Bioinformatics* **35**, 152–155 (2019).

54. Barnwal, R. P., Yang, F. & Varani, G. Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.* **628**, 42–56 (2017).

55. Zheng, G. & Price, W. S. Solvent signal suppression in NMR. *Prog. nucl. magn. reson. spectrosc.* **56**, 267–288 (2010).

56. Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **112**, 275–279 (1995).

57. Schlagnitweit, J., Steiner, E., Karlsson, H. & Petzold, K. Efficient detection of structure and dynamics in unlabeled RNAs: the SELOPE approach. *Chem. Eur. J* **24**, 6067–6070 (2018).

58. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *Chembiochem* **4**, 936–962 (2003).

59. Baronti, L. *et al.* Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. *Nature* **583**, 139–144 (2020).

60. Riad, M. *et al.* Mutate-and-chemical-shift-fingerprint (MCSF) to characterize excited states in RNA using NMR spectroscopy. *Nat. Protoc.* **16**, 5146–5170 (2021).

61. Roques, B. P., Rao, R. & Marion, D. Use of nuclear Overhauser effect in the study of peptides and proteins. *Biochimie* **62**, 753–773 (1980).

62. Guerry, P. & Herrmann, T. Advances in automated NMR protein structure determination. *Q. Rev. Biophys.* **44**, 257–309 (2011).

63. Vangaveti, S., Ranganathan, S. V. & Chen, A. A. Advances in RNA molecular dynamics: a simulator's guide to RNA force fields. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

64. Shi, H. *et al.* Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nat. Commun.* **11**, 5531 (2020).

65. Ebrahimi, P., Kaur, S., Baronti, L., Petzold, K. & Chen, A. A. A two-dimensional replica-exchange molecular dynamics method for simulating RNA folding using sparse experimental restraints. *Methods* **162–163**, 96–107 (2019).

66. Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W. & Al-Hashimi, H. M. Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320 (2015).

67. Marušič, M., Schlagnitweit, J. & Petzold, K. RNA dynamics by NMR spectroscopy. *Chembiochem* **20**, 2685–2710 (2019).

68. Wenter, P., Bodenhausen, G., Dittmer, J. & Pitsch, S. Kinetics of RNA refolding in dynamic equilibrium by 1H-detected 15N exchange NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 7579–7587 (2006).

69. Hwang, T. L., van Zijl, P. C. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**, 221–226 (1998).

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

71. Zhao, B., Baisden, J. T. & Zhang, Q. Probing excited conformational states of nucleic acids by nitrogen CEST NMR spectroscopy. *J. Magn. Reson.* **310**, 106642 (2020).

72. Xue, Y. *et al.* Characterizing RNA excited states using NMR relaxation dispersion. *Meth. Enzymol.* **558**, 39–73 (2015).

73. Kloiber, K., Spitzer, R., Tollinger, M., Konrat, R. & Kreutz, C. Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label. *Nucleic Acids Res.* **39**, 4340–4351 (2011).

74. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **491**, 724–728 (2012).

75. Barraud, P. *et al.* Time-resolved NMR monitoring of tRNA maturation. *Nat. Commun.* **10**, 3373 (2019).

76. Catala, M., Gato, A., Tisné, C. & Barraud, P. 1H, 15N chemical shift assignments of the imino groups of yeast tRNAPhe: influence of the post-transcriptional modifications. *Biomol. NMR Assign.* **14**, 169–174 (2020).

77. Beaucage, S. L. Solid-phase synthesis of siRNA oligonucleotides. *Curr. Opin. Drug Discov. Devel.* **11**, 203–216 (2008).

78. Lönnberg, H. Synthesis of oligonucleotides on a soluble support. *Beilstein J. Org. Chem.* **13**, 1368–1387 (2017).

79. Becette, O., Olenginski, L. T. & Dayie, T. K. Solid-Phase Chemical Synthesis of Stable Isotope-Labeled RNA to Aid Structure and Dynamics Studies by NMR Spectroscopy. *Molecules* **24**, (2019).

80. Flamme, M., McKenzie, L. K., Sarac, I. & Hollenstein, M. Chemical methods for the modification of RNA. *Methods* **161**, 64–82 (2019).

81. Röthlisberger, P., Berk, C. & Hall, J. RNA chemistry for RNA biology. *Chimia (Aarau)* **73**, 368–373 (2019).

82. Höbartner, C. & Wachowius, F. Chemical synthesis of modified RNA. in *The chemical biology of nucleic acids* (ed. Mayer, G.) 1–37 (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470664001.ch1.

83. Wojczewski, C., Stolze, K. & Engels, J. W. Fluorescent Oligonucleotides - Versatile Tools as Probes and Primers for DNA and RNA Analysis. *Synlett* **1999**, 1667–1678 (1999).

84. Sigurdsson, S. Th. Nitroxides and nucleic acids: Chemistry and electron paramagnetic resonance (EPR) spectroscopy. *Pure Appl. Chem.* **83**, 677–686 (2011).

85. Martínez-Montero, S. *et al.* Locked 2'-Deoxy-2',4'-Difluororibo Modified Nucleic Acids: Thermal Stability, Structural Studies, and siRNA Activity. *ACS Chem. Biol.* **10**, 2016–2023 (2015).

86. Deleavey, G. F. & Damha, M. J. Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.* **19**, 937–954 (2012).

87. Kurreck, J. Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.* **270**, 1628–1644 (2003).

88. Smith, C. I. E. & Zain, R. Therapeutic oligonucleotides: state of the art. *Annu. Rev. Pharmacol. Toxicol.* **59**, 605–630 (2019).

89. Neuner, S., Santner, T., Kreutz, C. & Micura, R. The "Speedy" Synthesis of Atom-Specific (15)N Imino/Amido-Labeled RNA. *Chem. Eur. J* **21**, 11634–11643 (2015).

90. Wunderlich, C. H. *et al.* Stable isotope-labeled RNA phosphoramidites to facilitate dynamics by NMR. *Meth. Enzymol.* **565**, 461–494 (2015).

91. LeBlanc, R. M., Longhini, A. P., Tugarinov, V. & Dayie, T. K. NMR probing of invisible excited states using selectively labeled RNAs. *J. Biomol. NMR* **71**, 165–172 (2018).

92. Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int. J. Biol. Macromol.* **118**, 49–56 (2018).

93. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).

94. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).

95. Zhu, B., Tabor, S. & Richardson, C. C. Syn5 RNA polymerase synthesizes precise run-off RNA products. *Nucleic Acids Res.* **42**, e33 (2014).

96. Zhu, B. *et al.* Synthesis of 2'-Fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Res.* **43**, e94 (2015).

97. Dietz, A., Weisser, H. J., Kössel, H. & Hausmann, R. The gene for Klebsiella bacteriophage K11 RNA polymerase: sequence and comparison with the homologous genes of phages T7, T3, and SP6. *Mol. Gen. Genet.* **221**, 283–286 (1990).

98. Jorgensen, E. D., Durbin, R. K., Risman, S. S. & McAllister, W. T. Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *J. Biol. Chem.* **266**, 645–651 (1991).

99. Wu, H. *et al.* A single mutation attenuates both the transcription termination and RNA-dependent RNA polymerase activity of T7 RNA polymerase. *RNA Biol.* **18**, 451–466 (2021).

100. Brunelle, J. L. & Green, R. In vitro transcription from plasmid or PCR-amplified DNA. *Meth. Enzymol.* **530**, 101–114 (2013).

101. Cunningham, P. R. & Ofengand, J. Use of inorganic pyrophosphatase to improve the yield of in vitro transcription reactions catalyzed by T7 RNA polymerase. *BioTechniques* **9**, 713–714 (1990).

102. Arnold, S. *et al.* Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnol. Bioeng.* **72**, 548–561 (2001).

103. Jasinski, D. L., Binzel, D. W. & Guo, P. One-Pot Production of RNA Nanoparticles via Automated Processing and Self-Assembly. *ACS Nano* **13**, 4603–4612 (2019).

104. Lee, J. H. *et al.* Rolling circle transcription-based polymeric siRNA nanoparticles for tumor-targeted delivery. *J. Control. Release* **263**, 29–38 (2017).

105. Macdonald, L. E., Zhou, Y. & McAllister, W. T. Termination and slippage by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **232**, 1030–1047 (1993).

106. Helm, M., Brulé, H., Giegé, R. & Florentz, C. More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5**, 618–621 (1999).

107. Pleiss, J. A., Derrick, M. L. & Uhlenbeck, O. C. T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates. *RNA* **4**, 1313–1317 (1998).

108. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).

109. Ikeda, R. A., Lin, A. C. & Clarke, J. Initiation of transcription by T7 RNA polymerase as its natural promoters. *J. Biol. Chem.* **267**, 2640–2649 (1992).

110. Ikeda, R. A. The efficiency of promoter clearance distinguishes T7 class II and class III promoters. *J. Biol. Chem.* **267**, 11322–11328 (1992).

111. Kuzmine, I., Gottlieb, P. A. & Martin, C. T. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J. Biol. Chem.* **278**, 2819–2823 (2003).

112. Conrad, T., Plumbom, I., Alcobendas, M., Vidal, R. & Sauer, S. Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun. Biol.* **3**, 439 (2020).

113. Gholamalipour, Y., Karunanayake Mudiyanselage, A. & Martin, C. T. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses. *Nucleic Acids Res.* **46**, 9253–9263 (2018).

114. Lapham, J. & Crothers, D. M. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA* **2**, 289–296 (1996).

115. Duss, O., Maris, C., von Schroetter, C. & Allain, F. H.-T. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res.* **38**, e188 (2010).

116. Shields, T. P., Mollova, E., Ste Marie, L., Hansen, M. R. & Pardi, A. High-performance liquid chromatography purification of homogenous-length RNA produced by trans cleavage with a hammerhead ribozyme. *RNA* **5**, 1259–1267 (1999).

117. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249**, 398–408 (1995).

118. Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268–1272 (1999).

119. Chen, Z. & Zhang, Y. Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. *Biochem. Biophys. Res. Commun.* **333**, 664–670 (2005).

120. Gong, P. & Martin, C. T. Mechanism of instability in abortive cycling by T7 RNA polymerase. *J. Biol. Chem.* **281**, 23533–23544 (2006).

121. Gong, P., Esposito, E. A. & Martin, C. T. Initial bubble collapse plays a key role in the transition to elongation in T7 RNA polymerase. *J. Biol. Chem.* **279**, 44277–44285 (2004).

122. Liu, C. & Martin, C. T. Promoter clearance by T7 RNA polymerase. Initial bubble collapse and transcript dissociation monitored by base analog fluorescence. *J. Biol. Chem.* **277**, 2725–2731 (2002).

123. Ramírez-Tapia, L. E. & Martin, C. T. New insights into the mechanism of initial transcription: the T7 RNA polymerase mutant P266L transitions to elongation at longer RNA lengths than wild type. *J. Biol. Chem.* **287**, 37352–37361 (2012).

124. Turingan, R. S., Liu, C., Hawkins, M. E. & Martin, C. T. Structural confirmation of a bent and open model for the initiation complex of T7 RNA polymerase. *Biochemistry* **46**, 1714–1723 (2007).

125. Steitz, T. A. The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase. *Curr. Opin. Struct. Biol.* **14**, 4–9 (2004).

126. Xu, J., Lapham, J. & Crothers, D. M. Determining RNA solution structure by segmental isotopic labeling and NMR: application to Caenorhabditis elegans spliced leader RNA 1. *Proc Natl Acad Sci USA* **93**, 44–48 (1996).

127. Kim, I., Lukavsky, P. J. & Puglisi, J. D. NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J. Am. Chem. Soc.* **124**, 9338–9339 (2002).

128. Nelissen, F. H. T. *et al.* Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res.* **36**, e89 (2008).

129. Duss, O., Diarra Dit Konté, N. & Allain, F. H.-T. Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. *Meth. Enzymol.* **565**, 537–562 (2015).

130. Brown, J. D. *et al.* Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* **368**, 413–417 (2020).

131. Kawahara, I. *et al.* Site-specific isotope labeling of long RNA for structural and mechanistic studies. *Nucleic Acids Res.* **40**, e7 (2012).

132. Keyhani, S., Goldau, T., Blümler, A., Heckel, A. & Schwalbe, H. Chemo-Enzymatic Synthesis of Position-Specifically Modified RNA for Biophysical Studies including Light Control and NMR Spectroscopy. *Angew Chem Int Ed Engl* **57**, 12017–12021 (2018).

133. Zhang, X., Li, M. & Liu, Y. Optimization and characterization of position-selective labelling of RNA (PLOR) for diverse RNA and DNA sequences. *RNA Biol.* **17**, 1009–1017 (2020).

134. Minton, A. P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* **20**, 2093–2120 (1981).

135. Lee, J. C. & Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **256**, 7193–7201 (1981).

136. Winzor, D. J. & Wills, P. R. Effects of thermodynamic nonideality on protein interactions. *Biophys. Chem.* **25**, 243–251 (1986).

137. Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37**, 375–397 (2008).

138. Rivas, G. & Minton, A. P. Macromolecular crowding in vitro, in vivo, and in between. *Trends Biochem. Sci.* **41**, 970–981 (2016).

139. Ignatova, Z. & Gierasch, L. M. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci USA* **101**, 523–528 (2004).

140. Dauty, E. & Verkman, A. S. Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: measurement by fluorescence correlation spectroscopy. *J. Mol. Recognit.* **17**, 441–447 (2004).

141. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* **158**, 244–253 (2007).

142. Hänsel, R., Luh, L. M., Corbeski, I., Trantirek, L. & Dötsch, V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew Chem Int Ed Engl* **53**, 10300–10314 (2014).

143. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. In-cell NMR spectroscopy. *Meth. Enzymol.* **394**, 17–41 (2005).

144. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 8895–8901 (2001).

145. Mu, X. *et al.* Physicochemical code for quinary protein interactions in Escherichia coli. *Proc Natl Acad Sci USA* **114**, E4556–E4563 (2017).

146. Waudby, C. A. *et al.* In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within E. coli cells. *PLoS ONE* **8**, e72286 (2013).

147. Croke, R. L., Sallum, C. O., Watson, E., Watt, E. D. & Alexandrescu, A. T. Hydrogen exchange of monomeric alpha-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. *Protein Sci.* **17**, 1434–1445 (2008).

148. Thongwichian, R. & Selenko, P. In-cell NMR in Xenopus laevis oocytes. *Methods Mol. Biol.* **895**, 33–41 (2012).

149. Selenko, P., Serber, Z., Gadea, B., Ruderman, J. & Wagner, G. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc Natl Acad Sci USA* **103**, 11904–11909 (2006).

150. Ye, Y. *et al.* Labeling strategy and signal broadening mechanism of Protein NMR spectroscopy in Xenopus laevis oocytes. *Chem. Eur. J* **21**, 8686–8690 (2015).

151. Luchinat, E. *et al.* Intracellular Binding/Unbinding Kinetics of Approved Drugs to Carbonic Anhydrase II Observed by in-Cell NMR. *ACS Chem. Biol.* **15**, 2792–2800 (2020).

152. Luchinat, E. & Banci, L. In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Acc. Chem. Res.* **51**, 1550–1557 (2018).

153. Theillet, F.-X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).

154. Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105 (2009).

155. Ikeya, T. *et al.* Improved in-cell structure determination of proteins at near-physiological concentration. *Sci. Rep.* **6**, 38312 (2016).

156. Takaoka, Y. *et al.* Quantitative comparison of protein dynamics in live cells and in vitro by in-cell (19)F-NMR. *Chem. Commun.* **49**, 2801–2803 (2013).

157. Latham, M. P. & Kay, L. E. Is buffer a good proxy for a crowded cell-like environment? A comparative NMR study of calmodulin side-chain dynamics in buffer and E. coli lysate. *PLoS ONE* **7**, e48226 (2012).

158. Burz, D. S. & Shekhtman, A. The STINT-NMR method for studying in-cell protein-protein interactions. *Curr. Protoc. Protein Sci.* **Chapter 17**, Unit 17.11 (2010).

159. Burz, D. S., Dutta, K., Cowburn, D. & Shekhtman, A. In-cell NMR for protein-protein interactions (STINT-NMR). *Nat. Protoc.* **1**, 146–152 (2006).

160. Burz, D. S., Dutta, K., Cowburn, D. & Shekhtman, A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods* **3**, 91–93 (2006).

161. Xie, J., Thapa, R., Reverdatto, S., Burz, D. S. & Shekhtman, A. Screening of small molecule interactor library by using in-cell NMR spectroscopy (SMILI-NMR). *J. Med. Chem.* **52**, 3516–3522 (2009).

162. Barbieri, L., Luchinat, E. & Banci, L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat. Protoc.* **11**, 1101–1111 (2016).

163. Sakai, T. *et al.* In-cell NMR spectroscopy of proteins inside Xenopus laevis oocytes. *J. Biomol. NMR* **36**, 179–188 (2006).

164. Hänsel, R. *et al.* Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 15761–15768 (2009).

165. Viskova, P., Krafcik, D., Trantirek, L. & Foldynova-Trantirkova, S. In-Cell NMR Spectroscopy of Nucleic Acids in Human Cells. *Curr. Protoc. Nucleic Acid Chem.* **76**, e71 (2019).

166. Yamaoki, Y. *et al.* The first successful observation of in-cell NMR signals of DNA and RNA in living human cells. *Phys. Chem. Chem. Phys.* **20**, 2982–2985 (2018).

167. Broft, P. *et al.* In-Cell NMR Spectroscopy of Functional Riboswitch Aptamers in Eukaryotic Cells. *Angew Chem Int Ed Engl* **60**, 865–872 (2021).

168. Bao, H.-L., Masuzawa, T., Oyoshi, T. & Xu, Y. Oligonucleotides DNA containing 8-trifluoromethyl-2'-deoxyguanosine for observing Z-DNA structure. *Nucleic Acids Res.* **48**, 7041–7051 (2020).

169. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184-9 (2016).

170. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).

171. Reichholf, B. *et al.* Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis. *Mol. Cell* **75**, 756-768.e7 (2019).

172. Salgado, G. F., Cazenave, C., Kerkour, A. & Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **6**, 3314–3320 (2015).

173. Bao, H.-L. & Xu, Y. Telomeric DNA-RNA-hybrid G-quadruplex exists in environmental conditions of HeLa cells. *Chem. Commun.* **56**, 6547–6550 (2020).

174. Bao, H.-L. *et al.* Characterization of human telomere RNA G-quadruplex structures in vitro and in living cells using 19F NMR spectroscopy. *Nucleic Acids Res.* **45**, 5501–5511 (2017).

175. Bao, H.-L. & Xu, Y. Investigation of higher-order RNA G-quadruplex structures in vitro and in living cells by 19F NMR spectroscopy. *Nat. Protoc.* **13**, 652–665 (2018).

176. Krafčík, D. *et al.* Towards Profiling of the G-Quadruplex Targeting Drugs in the Living Human Cells Using NMR Spectroscopy. *Int. J. Mol. Sci.* **22**, (2021).

177. Krafcikova, M. *et al.* Monitoring DNA-Ligand Interactions in Living Human Cells Using NMR Spectroscopy. *J. Am. Chem. Soc.* **141**, 13281–13285 (2019).

178. Dzatko, S. *et al.* Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew Chem Int Ed Engl* **57**, 2165–2169 (2018).

179. Cheng, M. *et al.* Thermal and pH Stabilities of i-DNA: Confronting in vitro Experiments with Models and In-Cell NMR Data. *Angew Chem Int Ed Engl* **60**, 10286–10294 (2021).

180. Sakamoto, T., Yamaoki, Y., Nagata, T. & Katahira, M. Detection of parallel and antiparallel DNA triplex structures in living human cells using in-cell NMR. *Chem. Commun.* **57**, 6364–6367 (2021).

181. Schlagnitweit, J. *et al.* Observing an Antisense Drug Complex in Intact Human Cells by in-Cell NMR Spectroscopy. *Chembiochem* **20**, 2474–2478 (2019).

182. Pellegrini, O., Mathy, N., Condon, C. & Bénard, L. Chapter 9 In Vitro Assays of 5′ to 3′-Exoribonuclease Activity. in *RNA Turnover in Eukaryotes: Nucleases, Pathways and Analysis of mRNA Decay* vol. 448 167–183 (Elsevier, 2008).

183. Jones, C. I., Zabolotskaya, M. V. & Newbury, S. F. The 5' → 3' exoribonuclease XRN1/Pacman and its functions in cellular processes and development. *Wiley Interdiscip. Rev. RNA* **3**, 455–468 (2012).

184. Stevensj, A. & Poole, T. 5' -Exonuclease-2 of Saccharomyces cerevisiae.

185. Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: the exosome. *Biochim. Biophys. Acta* **1829**, 552–560 (2013).

186. Łabno, A., Tomecki, R. & Dziembowski, A. Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochim. Biophys. Acta* **1863**, 3125–3147 (2016).

187. Mu, X., Greenwald, E., Ahmad, S. & Hur, S. An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic Acids Res.* **46**, 5239–5249 (2018).

188. Ku, S. H., Jo, S. D., Lee, Y. K., Kim, K. & Kim, S. H. Chemical and structural modifications of RNAi therapeutics. *Adv. Drug Deliv. Rev.* **104**, 16–28 (2016).

189. Costello, W. N., Xiao, Y. & Frederick, K. K. DNP-Assisted NMR Investigation of Proteins at Endogenous Levels in Cellular Milieu. *Meth. Enzymol.* **615**, 373–406 (2019).

190. Barnes, C. O. & Pielak, G. J. In-cell protein NMR and protein leakage. *Proteins* **79**, 347–351 (2011).

191. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* **94**, 12366–12371 (1997).

192. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).

193. Sklenář, V. & Bax, A. Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson* **74**, 469–479 (1987).

194. Paudel, B. P. & Rueda, D. Molecular crowding accelerates ribozyme docking and catalysis. *J. Am. Chem. Soc.* **136**, 16700–16703 (2014).

195. Cui, J., Waltman, P., Le, V. H. & Lewis, E. A. The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules* **18**, 12751–12767 (2013).

196. Mahen, E. M., Watson, P. Y., Cottrell, J. W. & Fedor, M. J. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol.* **8**, e1000307 (2010).

197. Lebars, I. *et al.* A fully enzymatic method for site-directed spin labeling of long RNA. *Nucleic Acids Res.* **42**, e117 (2014).

198. Litke, J. L. & Jaffrey, S. R. Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat. Biotechnol.* **37**, 667–675 (2019).

199. Chen, H. *et al.* Preferential production of RNA rings by T4 RNA ligase 2 without any splint through rational design of precursor strand. *Nucleic Acids Res.* **48**, e54 (2020).

200. Wang, X., Li, C., Gao, X., Wang, J. & Liang, X. Preparation of Small RNAs Using Rolling Circle Transcription and Site-Specific RNA Disconnection. *Mol. Ther. Nucleic Acids* **4**, e215 (2015).

201. Feyrer, H., Munteanu, R., Baronti, L. & Petzold, K. One-Pot Production of RNA in High Yield and Purity Through Cleaving Tandem Transcripts. *Molecules* **25**, (2020).

202. Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).

203. Ketterer, S., Fuchs, D., Weber, W. & Meier, M. Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. *Nucleic Acids Res.* **43**, 9564–9572 (2015).

204. Kulikov, K., Nozinovic, S. & Kath-Schorr, S. The 5'-AG5 CC-3' Fragment from the Human CPEB3 Ribozyme Forms an Ultrastable Parallel RNA G-Quadruplex. *Chembiochem* **18**, 969–973 (2017).

205. Dickerhoff, J., Appel, B., Müller, S. & Weisz, K. Sugar-Edge Interactions in a DNA-RNA G-Quadruplex: Evidence of Sequential C-H···O Hydrogen Bonds Contributing to RNA Quadruplex Folding. *Angew Chem Int Ed Engl* **55**, 15162–15165 (2016).

206. Truong, L. *et al.* The fluorescent aptamer Squash extensively repurposes the adenine riboswitch fold. *Nat. Chem. Biol.* **18**, 191–198 (2022).

207. Ghosh, R., Xiao, Y., Kragelj, J. & Frederick, K. K. In-Cell Sensitivity-Enhanced NMR of Intact Viable Mammalian Cells. *J. Am. Chem. Soc.* **143**, 18454–18466 (2021).

208. Overall, S. A. & Barnes, A. B. Biomolecular Perturbations in In-Cell Dynamic Nuclear Polarization Experiments. *Front. Mol. Biosci.* **8**, 743829 (2021).

209. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).