Probing Nucleic Acid Structure and Recognition in Cell-Free and Cellular Environments Using Environment-Sensitive Nucleoside Probes

A thesis submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

by

Sudeshna Manna

ID: 20133235

Indian Institute of Science Education and Research, Pune

2019

Dedicated to

my parents and husband

Dr. Seergazhi G. Srivatsan
Associate Professor, Chemistry

Dr. Seergazhi G. Srivatsan
Associate Professor
Department of Chemistry
IISER, Pune

CERTIFICATE

Certified that the work incorporated in the thesis entitled "*Probing Nucleic Acid Structure and Recognition in Cell-Free and Cellular Environments Using Environment-Sensitive Nucleoside Probes*" submitted by Ms. Sudeshna Manna was carried out by the candidate under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other university or institution.

Date: 15/01/2019
Place: Pune

Dr. Seergazhi G. Srivatsan

DECLARATION

I declare that this written submission represents my ideas in my own words and wherever other's ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in my submission. I understand that violation of the above will cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Sudeshna Manna

Date: 15.01.2019   **Sudeshna Manna**

Place: Pune   Reg No: 20133235

Acknowledgments

Firstly and most importantly I want to express my heartfelt gratitude to my supervisor Dr. S. G. Srivatsan for giving me the opportunity to work in his group and guiding me in my PhD thesis work. His enthusiasm towards science, motivational words and broader goal in research always provided me the positive energy to do well in research. His support and thoughtful insights helped me to cross the difficulties appeared during my research work. I wish to thank him for providing me the training for designing experiments, performing them meticulously and working independently in research projects. Besides, I am grateful to him for his critical evaluation of my presentation and writing skills. I also thank him for his considerable help in my thesis writing. The training I got from him has made me confident for my future research endeavors. I always feel fortunate to have him as my PhD mentor.

It has been a great opportunity for me to work in IISER Pune which provides an excellent research environment with world-class research facilities. My sincere gratitude goes to former Director of IISER Pune, Prof. K. N. Ganesh for establishing such an excellent institute which has made many students' dream come true. I would like to thank my Research Advisory Committee (RAC) members, Dr. G. J. Sanjayan and Dr. Jeetender Chugh for their valuable comments and suggestions during RAC meetings. A special thank goes to Dr. Jeetender Chugh for helping me in peroforming my NMR experiments. The productive discussions with him were very useful for my research work. I am thankful to Dr. Jeet Kalia for generously providing the oocytes for my experiments and allowing me to use his frog oocytes facility at IISER Pune. I want to thank Prof. M. Jayakannan, Chair Chemistry for all departmental facilities. I thank all non-teaching and technical staffs particularly Nitin, Mahesh, Swati, Mayuresh, Tushar and Nayana for their help. I am thankful to University Grants Commission (UGC) for research fellowship and DBT India and IISER Pune Infosys foundation for travel support to attend international conference.

It's my great pleasure to acknowledge my former and present labmates Dr. Maroti, Dr. Anupam, Dr. Arun, Dr. Pramod, Dr. Ashok, Dr. Vyankat, Dr. Cornelia, Jerrin, Manisha, Pankaj, Saddam, Pulak, Akanksha, Samiksha, Uddhav and Sangamesh who played important roles in my PhD in many ways. Their help during my work, making presentations and writing thesis made my PhD journey smoother. Importantly, the fun times with them in lab, lab

Chapter 1 is a reprint of: **Manna, S.**; Srivatsan, S. G. Fluorescence-Based Tools to Probe G-Quadruplexes in Cell-Free and Cellular Environments. *RSC Adv.* **2018**, *8*, 25673–25694.

The dissertation author is the main author for this article.

Chapter 2 is a reprint of: **Manna, S.**; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model by Using Responsive Fluorescent Nucleoside Probes. *ChemBioChem* **2017**, *18*, 1604–1615.

The dissertation author is the main author and researcher for this work.

Chapter 3 is a reprint of: **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.* **2018**, *140*, 12622–12633.

The dissertation author is the main author and researcher for this work.

Table of Contents

Chapter 1: Tools to Investigate Nucleic acid Structure, Function and Recognition in Cell-Free and Cellular Environments

Chapter 2: Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model

Chapter 3: A Dual-App Nucleoside Probe Provides Structural Insights into the Human Telomeric Overhang in Live Cells

Chapter 4: Enzymatic Incorporation and Utility of Fluorobenzofuran-Modified Uridine in Probing Conformational Change in Viral RNA

Chapter 5: An Environment-Sensitive Fluorescent Nucleoside Probe Detects Metal-Mediated Base Pairing in Duplexes

<u>Abbreviations</u>

2-AP	2-Aminopurine
A	Adenosine
ACN	Acetonitrile
AOT	Dioctyl Sodium Sulfosuccinate
ATP	Adenosine Triphosphate
C	Cytosine
CD	Circular Dichroism
COSY	Correlated Spectroscopy
CTP	Cytidine Triphosphate
3D	Three Dimensional
DEER	Double Electron-Electron Resonance
DMAP	4-Dimethylaminopyridine
DMSO	N, N-dimethyl sulfoxide
DMF	Dimethylformamide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
ds	Double-stranded
EDTA	Ethylenediaminetetraacetic Acid
em	Emission
ex	Excitation
EPR	Electron Paramagnetic Resonance
FLIM	Fluorescence Lifetime Imaging Microscopy
FRET	Fluorescence Resonance Energy Transfer
G	Guanosine
GTP	Guanosine Triphosphate
GQ	G-quadruplex
HCV	Hepatitis C Virus

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBC	Heteronuclear Multiple Bond Correlation
HPA	Hydroxylpiccolinic Acid
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
H-Telo	Human Telomeric
in vitro	Outside living organism
in vivo	Inside living organism
IRES	Internal Ribosome Entry Site
K_d	Dissociation Constant
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time of flight
max	Maximum
MeOH	Methanol
mg	Milligram
MHz	Megahertz
6-MI	6-Methyl Isoxanthopterin
μM	Micromolar
μL	Microliter
mM	Milimolar
ε	Molar extinction coefficient
nm	Nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NTP	Nucleoside Triphosphate
ON	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank

PEG	Polyethylene Glycol
PELDOR	Pulsed Electron Double Resonance
Pd	Palladium
PDS	Pyridostatin
ppm	Parts per million
RM	Reverse Micelle
RNA	Ribonucleic acid
RNase	Ribonuclease
SNP	Single Nucleotide Polymorphism
ss	Single-stranded
T	Thymine
TBAF	Tetrabutylammonium Fluoride
TBDMS	Tert-Butyldimethylsilyl
TCA	Trichloroacetic acid
TCSPC	Time Correlated Single Photon Counting
TEAA	Triethylammonium Acetate
TERRA	Telomeric Repeat-Containing RNA
THF	Tetrahydrofuron
TLC	Thin layer chromatography
T_m	Thermal melting
TMEDA	N,N,N',N'-Tetramethylethylenediamine
TOCSY	Total Correlation Spectroscopy
Tris	Tris (hydroxymethyl) Amino Methane
U	Uridine/Uracil
UTP	Uridine Triphosphate
UV	Ultraviolet
WC	Watson-Crick

Synopsis

Probing Nucleic Acid Structure and Recognition in Cell-Free and Cellular Environments Using Environment-Sensitive Nucleoside Probes

---

Background and Aim

Biological functions of nucleic acids are directly associated with the structures they adopt in cellular environment. Apart from classical double helical structure, nucleic acids adopt complex yet defined functional architectures, which manifest in their diverse cellular functions including storage and transfer of genetic information, catalysis and regulation of protein synthesis.[1] Aberrations in structure due to mutations, lesions and etc., can lead to nucleic acid dysfunction thereby resulting in disease states.[2] Hence, fundamental understanding of nucleic acid structure and dynamics and its interaction with proteins and small molecules is very important in biomedical research. In this regard, several biophysical tools based on fluorescence, NMR, EPR and X-ray crystallography techniques, to name a few, have been employed in studying nucleic acids structures, dynamics and recognition. As nucleic acids do not contain intrinsic labels that are compatible with biophysical techniques, researchers greatly rely on custom-labeled nucleic acid sequences.[3] For example, nucleoside analogs containing appropriate labels are very useful in studying nucleic acid structure, metal-nucleic acid, small molecule-nucleic acid and nucleic acid-protein interactions, and in developing screening platforms to identify efficient small molecule binders for different nucleic acid structures.

While these approaches have arguably provided valuable information, our current understanding of how nucleic acids fold and perform their function, particularly in cellular environment is limited.[4] For example, structure, dynamics and recognition of nucleic acid are greatly influenced by its surrounding environments like ionic conditions, pH, viscosity, molecular crowding and confinement.[5] Hence the properties of nucleic acids *in vitro* and in native cellular environment need not be same. Recently, nucleic acid structure-specific antibodies and light-up probes have been developed to visualize structural motifs in cells.[6] Additionally, in-cell NMR and EPR techniques have been employed to understand the preferred conformation of nucleic acid structural motifs in the crowded cellular environment.[4] In this direction, we hypothesized that development of environment-sensitive nucleoside probes, which (i) are minimally invasive, (ii) detect and report different conformations of a

nucleic acid sequence, (iii) can be used in estimating the ligand binding to different nucleic conformations, and (iv) importantly, are compatible will cell-free and cellular biophysical assays, will be highly beneficial in the comprehensive analysis of nucleic acids.

This thesis describes the design, synthesis and applications of responsive nucleoside analog probes containing one or two labels (fluorophore and or ^{19}F NMR isotope) in the investigation of nucleic acid structure and recognition in cell-free and cellular environments. Nucleoside probes are derived by attaching heterobicycles (benzofuran, 5-fluoro-benzofuran and 5-methoxybenzofuran) at the 5 position of uridine and 2'-deoxyuridine (Figure 1). The probes are highly sensitive to changes to their microenvironment and are structurally minimally perturbing upon incorporation into DNA and RNA ONs. The conformation-sensitive fluorescence properties of benzofuran-modified uridine and deoxyuridine analogs enabled the probing of GQ structure adopted by human telomeric (H-Telo) DNA and RNA repeats and their ligand binding affinty in aqueous buffer and cell-like confined environment of reverse micelles (RM, Figure 1A). 5-fluorobenzofuran-modified nucleoside analogs containing a fluorophore and ^{19}F NMR label act as a dual-app probe. By performing in-cell ^{19}F NMR analysis using this dual-app probe we successfully determined the preferred GQ structure of H-Telo DNA oligonucleotide repeat in live cells (Figure 1B). Further, this probe was used in probing the metal ion-induced conformational change in a hepatitis C viral RNA domain. The fluorescent methoxybenzofuran-modified uridine analog enabled the detection of dT-U and U-U pyrimidine-pyrimidine mispairing and dT-Hg-U, U-Hg-U metallo basepairs in nucleic acid duplex structures (Figure 1C). Taken together, the utility of these nucleoside probes should complement existing biophysical tools for the analysis of nucleic acids.
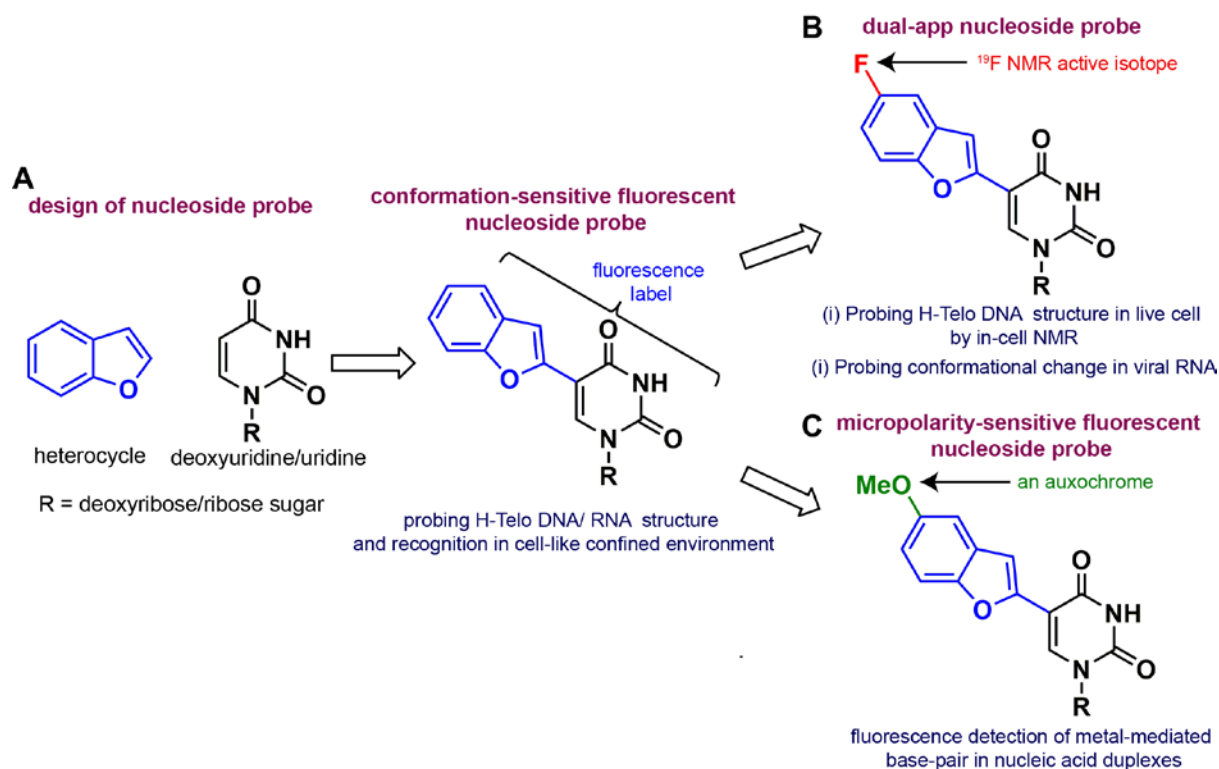
Figure 1. Design of responsive nucleoside probes and their applications in studying nucleic acid structure and recognition in cell-free and cellular environments.

The thesis is organized as follows.

Chapter 1: Tools to Investigate Nucleic acid Structure, Function and Recognition in Cell-Free and Cellular Environments

Various non-canonical nucleic acid structures and their biological significance have been briefly discussed in this chapter. An overview of how different biophysical tools like CD, NMR, EPR and fluorescence have been employed in nucleic acid structural analysis is presented. The usefulness of spin-, isotope- and heavy atom-labeled nucleoside probes in EPR, NMR and X-ray crystallography is described shortly. A detailed discussion on the use of environment-sensitive fluorescent nucleoside probes in studying non-canonical nucleic acid structures and ligand binding is provided. Following this, the utility of structure-specific antibodies, fluorescent light-up probes, in-cell NMR and in-cell EPR techniques for *in vivo* nucleic acid structural analysis is discussed. In the last section, limitations of existing biophysical tools and motivation for our probe design approach are illustrated.

Chapter 2: Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model

G-rich human telomeric (H-Telo) DNA and RNA repeat which can form stable GQ structures are integral parts of telomere structures.[2a] Stabilization of these structures by small molecules is viewed as a therapeutic strategy for cancer. H-Telo DNA can form multiple topologies like parallel, antiparallel and hybrid structures depending on the ionic conditions and crowding environment.[2a] On the other hand, H-Telo RNA repeat forms only parallel conformation irrespective of ionic conditions and crowded environment. Despite of flurry of research on GQs, distinguishing different conformations of GQs and understanding the preferred GQ topology in native cellular environment remains a challenge. Crowding agent like polyethylene glycol (PEG) has been used majorly to mimic the crowding environment of a native cell.[7] But dehydrating nature of PEG forces the G-rich sequences to form a particular topology. Therefore PEG fails to mimic the actual crowding environment of a cell.[8] Hence development of a method, which will enable the probing of H-Telo DNA and RNA structures and recognition properties in a cell-like confined environment will be highly advantageous in GQ-directed drug discovery.

In this chapter, the establishment of a fluorescence-based assay to detect different GQ conformations in cell-like confined environment is described.[9] The assay uses the conformation sensitivity of fluorescent benzofuran-modified uridine and deoxyuridine analogs and a widely used cellular model, AOT reverse micelles (RM). The benzofuran-modified uridine analog reports the microenvironment of the water core encapsulated in RMs with increasing molar ratio of water (w_0 value). With increasing the size of the water core, the probe showed quenching in fluorescence intensity with red shift in emission maximum (Figure 2A). As different GQ topologies have distinct loop conformation, the probe was incorporated into different loop positions of H-Telo DNA and RNA (Figure 2B). The probe was found to be minimally perturbing and did not hamper the formation of respective GQs in different ionic conditions. The nucleoside probe photophysically reported the formation of different GQ topologies in aqueous buffer and in AOT RM (Figure 2C). Additionally, this system provided a platform to directly compare the ligand binding affinity of human telomeric DNA and RNA repeat in aqueous buffer and confined environment of RM (Figure 2D). It is expected that this tool would be useful in identifying small molecules, which will bind efficiently to a particular GQ structure in confined environment of a cell.
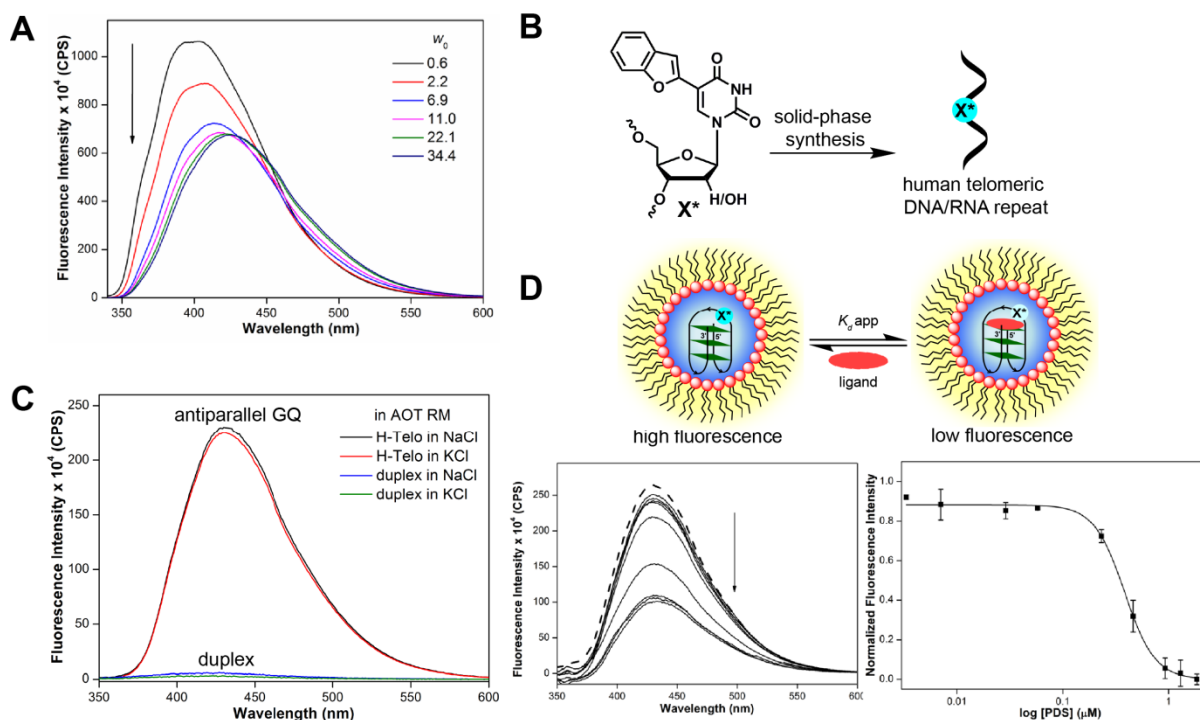
Figure 2. (A) Emission spectra of nucleoside **X***(uridine analog) in AOT RM at different w_0 values. (**B**) Nucleoside probes are incorporated into human telomeric DNA and RNA. (**C**) The nucleoside probes photophysically detect the formation of particular GQ topology in AOT RM. (**D**) The nucleoside probes also help in determining the binding affinity of a ligand to GQ structures in confined environment of RM.[9]

Chapter 3: A Dual-App Nucleoside Probe Provides Structural Insights into the Human Telomeric Overhang in Live Cells

As H-Telo DNA can form multiple conformations *in vitro* depending on the surrounding conditions, it has remained most debated till date that which structure actually is formed in native cellular environment. Studies using GQ topology-specific antibody, in-cell NMR and in-cell EPR showed contradicting results about H-Telo GQ topology in cells.[9] Hence the development of a method to understand GQ topology adopted by H-Telo DNA in cell will be highly beneficial in GQ-directed therapeutics. In order to investigate preferred H-Telo GQ topology in cell, we designed a dual-app probe 5-fluorobenzofuran-modified deoxyuridine analog (^{F}dU) (Figure 3A).[10] The probe contains an environment-sensitive fluorophore and an in-cell NMR compatible ^{19}F isotope. Both fluorescence and ^{19}F NMR properties of the probe is sensitive to solvent polarity and viscosity. CD and thermal melting analysis of modified and control unmodified H-Telo DNA GQ revealed that the probe minimally hampers the H-Telo DNA GQ topology. The dual-app probe differentiates the GQ structure from duplex and distinguishes different GQ topologies by showing changes in its fluorescence intensity and ^{19}F NMR signal (Figure 3B and 3C). The fluorophore component of the dual-purpose probe

provides a platform to estimate topology-specific binding of ligands to GQs (Figure 3D). The responsiveness of ^{19}F labeled nucleoside **1** to subtle changes in the conformation produces a distinct ^{19}F NMR signature for different GQ conformations. A comparison of ^{19}F NMR signal in the *in vitro* (buffer), *ex vivo* (egg extract), and *in vivo* (live oocytes) conditions provides a new structural insights into the GQ topologies adopted by H-Telo DNA overhang in cells (Figure 3E). Previous studies using synthetic cell models, immunostaining on fixed cells, and crystallization conditions proposed that parallel topology is the preferred topology of H-Telo DNA in cellular environment. However, our findings indicate that H-Telo can form multiple topologies like parallel and hybrid in native cellular environment.[10]
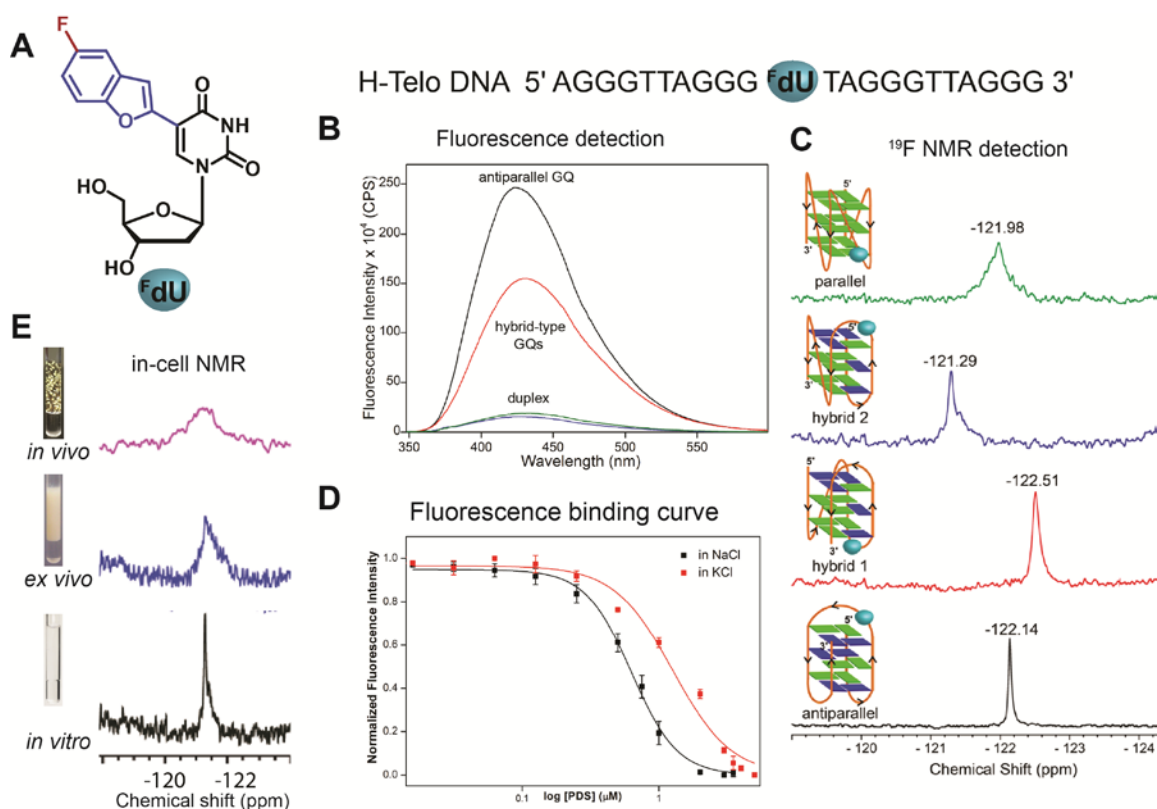
Figure 3. (**A**) Dual-app nucleoside probe composed of a fluorophore and a ^{19}F label. (**B**) Fluorescence detection of various GQ topologies. (**C**) The dual-app probe provides distinct ^{19}F NMR signal for each GQ topology. (**D**) The probe reports topology-specific binding of GQ ligands. (**E**)^{19}F NMR analysis in egg extract (ex-vivo) and in live oocyte (in vivo) enabled the detection of preferred GQ topology in cellular environment.[10]

Chapter 4: Enzymatic Incorporation and Utility of Fluorobenzofuran-Modified Uridine in Probing Conformational Change in Viral RNA

The scope of the dual-app probe is further explored in the context of RNA. We have synthesized 5-fluorobenzofuran-modified uridine analog (^{F}U) and the nucleoside shows similar fluorescence and ^{19}F NMR sensitivity like its deoxyuridine analog (^{F}dU). Further, we have synthesized corresponding triphosphate (^{F}UTP) and performed transcription reaction using T7 RNA polymerase. By this enzymatic method the probe is incorporated into RNA ONs with good to moderate efficiency (Figure 4). Interestingly, the probe when placed between different flanking bases in RNA signals the changes in its neighbouring base environment by exhibiting significant change in its fluorescence and ^{19}F NMR properties. These results have prompted us to incorporate it into a therapeutically important RNA ON, IRES (Internal Ribosomal Entry Site) element of Hepatitis C virus to study its conformation and recognition properties. The subdomain IIa of HCV IRES adopts bent shape structure in presence of metal ions Mg^{II} and Mn^{II} and this conformation is crucial for the protein synthesis in virus.[11] Aberration of this bent structure using small molecule can hamper the viral protein synthesis and is considered as a new approach in antiviral drug discovery.[12] The probe is incorporated site-specifically into two different positions (U56 and U106) of HCV IRES subdomain (Figure 5).[11a] When the fluorescence studies of modified subdomain IIa constructs is performed in the presence of increasing concentration of metal ions Mg^{II} and Mn^{II}, it shows a dose-dependent fluorescence quenching which helps us to get apparent binding constant for each metal ion to a particular site of the structural RNA. Utilization of the probe in the screening of the efficient small molecule binders for this structural element will be explored in future.
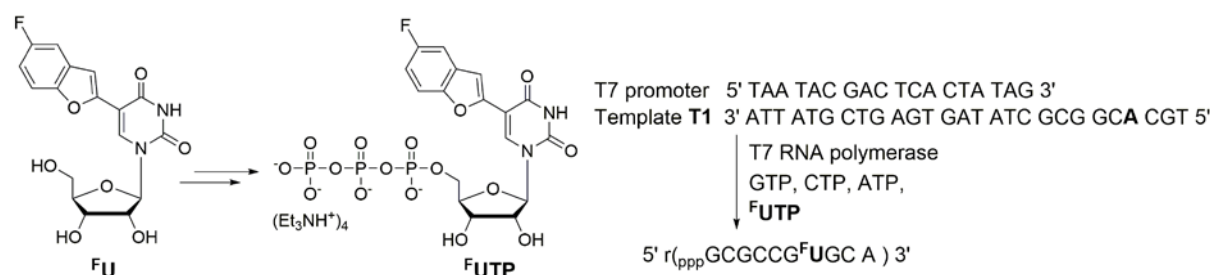
Figure 4. Incorporation of triphosphate of ^{F}U into RNA ONs by transcription reaction.
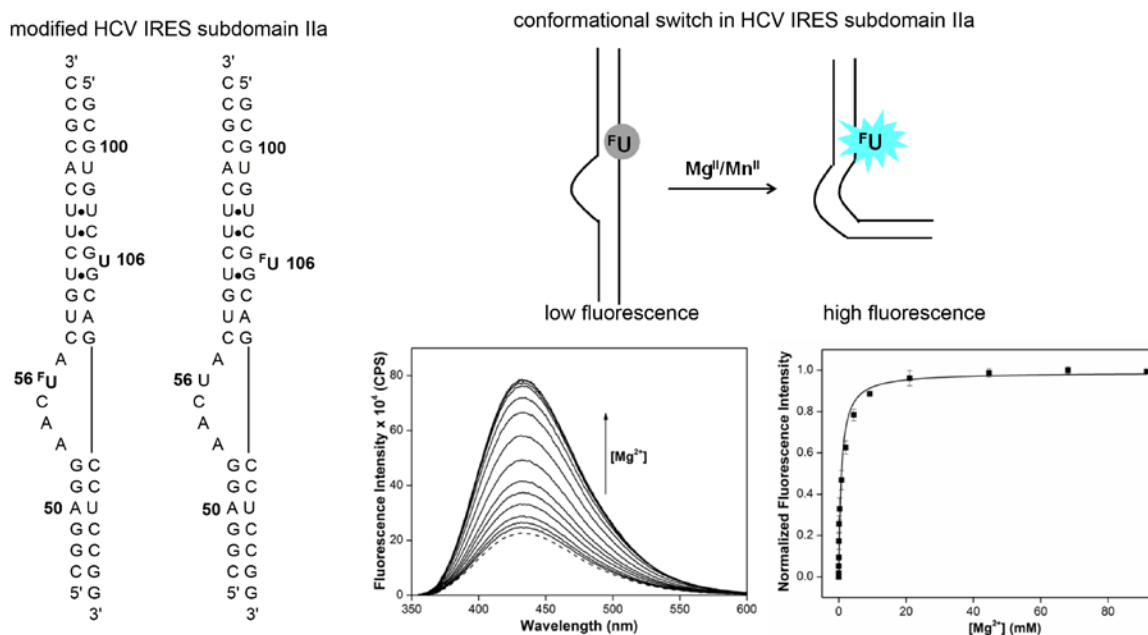
Figure 5. ^FU is incorporated site-specifically at U106 and U56 positions in IRES subdomain IIa of Hepatisis C virus (HCV). Nucleoside probe ^FU reports the metal-ion stabilized bent conformation of IRES subdomain IIa by showing turn-on fluorescence and helps to determine the apparent K_d for metal ion binding to a particular site of the structural RNA.

Chapter 5: An Environment-Sensitive Fluorescent Nucleoside Probe Detects Metal-Mediated Base Pairing in Duplexes

As a part of the ongoing effort in our lab to design responsive fluorescent nucleoside probes, here in this chapter we have developed 5-methoxybenzofuran-modified uridine (^{OMe}U) analog. Compared to previously-established fluorescent nucleoside probes in our group, ^{OMe}U is highly sensitive to its microenvironment. Particularly the fluorescence properties of the probe are sensitive to small changes in micropolarity (Figure 6A). This probe is successfully incorporated into RNA ONs by transcription reaction with good efficiency (Figure 6B). Importantly, after incorporation into RNA, it selectively reported thymine-uracil (dT-U) and uracil-uracil (U-U) mismatches in model DNA-RNA and RNA-RNA duplexes, respectively, with the significant enhancement in fluorescence intensity (Figure 7A). This property of the nucleoside analog is aptly utilized further in establishing a diagnostic method to detect single mercury (Hg) mediated base paring in DNA-RNA and RNA-RNA duplexes. When fluorescence studies are performed with DNA-RNA duplex containing dT-^{OMe}U mispair and RNA-RNA duplex containing U-^{OMe}U mispair, they show significant quenching in the fluorescence intensity due to the formation of dT-Hg-^{OMe}U and U-Hg-^{OMe}U base-pairs (Figure 7B). Further, by monitoring the fluorescence profile of 5-methoxybenzofuran-modified duplexes, containing a particular mismatch, as a function of Hg^{II} concentration we

are able to compare the binding affinity of Hg^{II} to dT-^{OMe}U and U-^{OMe}U mismatches. (Figure 7C). Hg^{II} shows strong binding affinity to both the mispairs with a comparatively higher affinity towards dT-^{OMe}U mismatch.
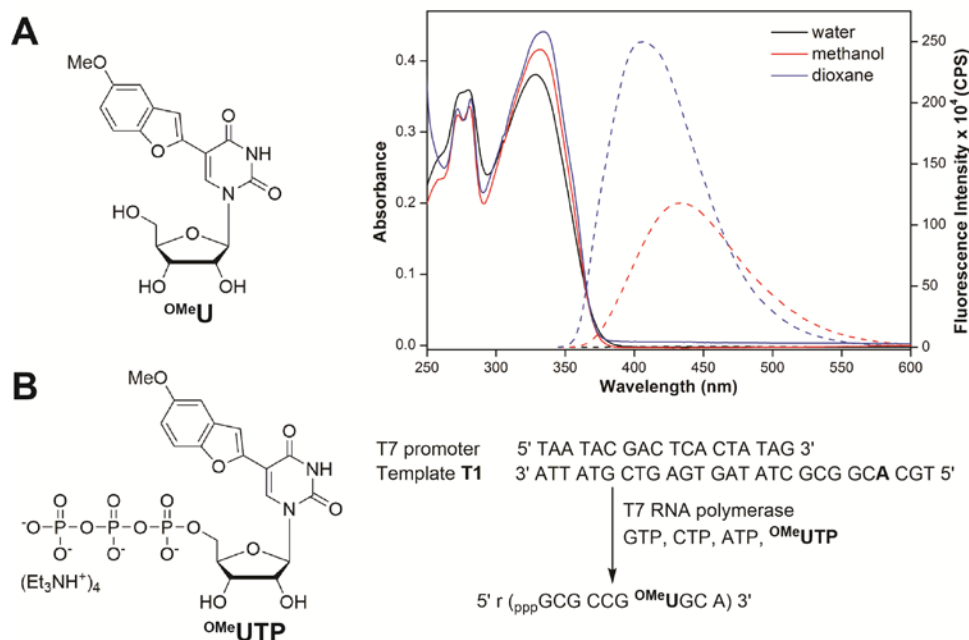
Figure 6. (A) ^{OMe}U is highly sensitive to changes in micropolarity of the surrounding environment.**(B)** Incorporation of ^{OMe}UTP into RNA ON by transcription reaction.
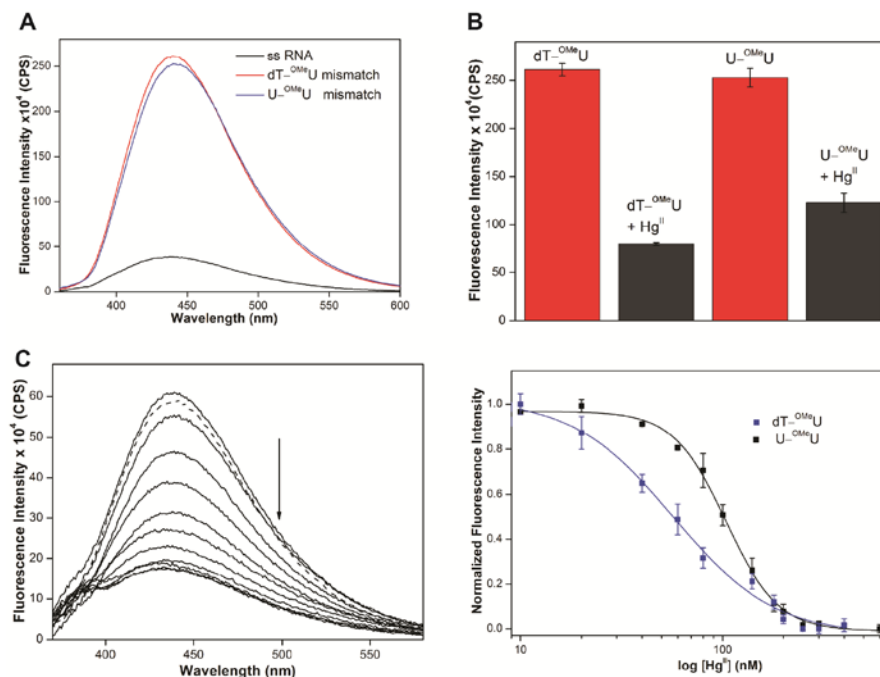
Figure 7. (A) ^{OMe}U reports the presence of dT-U and U-U mismatches in nucleic acid duplexes by showing turn-on fluorescence **(B)** ^{OMe}U detects the Hg^{II} mediated base-pair formation in duplexes by showing quenching in fluorescence intensity. **(C)** The probe helps to determine the apparent binding constant of Hg^{II} to nucleic acid duplexes with respective mispair.

References

1. (a) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1013–1023. (b) Majima, T. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909.
2. (a) Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867–5892. (b) Cammas, A.; Millevoi, S. *Nucleic Acids Res.* **2017**, *45*, 1653–1668. (c) Theisen, A.; Shaffer, L. G. *Appl Clin Genet.* **2010**, *3*, 159–174.
3. Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.
4. Giassa, I.-C.; Rynes, J.; Fessl, T.; Foldynova-Trantirkova, S.; Trantirek, L. *FEBS Letters* **2018**, *592*, 1997–2011.
5. (a) Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. *FEBS Letters* **2002** *526*, 77–81. (b) Nakano, S.; Miyoshi, D.; Sugimoto, N. *Chem. Rev.* **2014**, *114*, 2733. (c) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res*. **2018**, *46*, 8038–8056.
6. Manna, S.; Srivatsan, S. G. *RSC Adv.* **2018**, *8*, 25673–25694.
7. Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.
8. Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2013**, *41*, 7934–7946.
9. Manna, S.; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. *ChemBioChem* **2017**, *18*, 1604–1615.
10. Manna, S.; Sarkar, D.; Srivatsan, S. G. *J. Am. Chem. Soc.* **2018**, *140*, 12622–12633.
11. (a) Dibrov, S. M.; Johnston-Cox, H.; Weng, Y. H.; Hermann, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 226–229. (b) Otto, G. A.; Puglisi, J. D. *Cell* **2004**, *119*, 369–380.
12. Dibrov, S. M.; Parsons, J.; Carnevali, M.; Zhou, S.; Rynearson, K. D.; Ding, K.; Sega, E. G.; Brunn, N. D.; Boerneke, M. A.; Castaldi, M. P.; Hermann, T. *J. Med. Chem.* **2014**, *57*, 1694–1707.

List of Publications

1. **Manna, S.**; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model by Using Responsive Fluorescent Nucleoside Probes. *ChemBioChem* **2017**, *18*, 1604–1615.

2. **Manna, S.**; Srivatsan, S. G. Fluorescence-Based Tools to Probe G-Quadruplexes in Cell-Free and Cellular Environments. *RSC Adv.* **2018**, *8*, 25673–25694.

3. **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.* **2018**, *140*, 12622–12633.

4. **Manna, S.**; Sontakke, V. A.; Srivatsan, S. G. Enzymatic Incorporation and Utility of Fluorobenzofuran-Modified Uridine in Probing Conformational Change in Viral RNA. (Manuscript under preparation)

5. **Manna, S.**; Srivatsan, S. G. An Environment-Sensitive Fluorescent Nucleoside Probe Detects Metal-Mediated Base Pairing in Duplexes. (Manuscript under preparation)

Chapter 1

Tools to Investigate Nucleic acid Structure, Function and Recognition in Cell-Free and Cellular Environments

1.1 Introduction

The ability of nucleic acids to function as a genetic material, catalyst and regulatory element emanates from their ability to adopt complex secondary and tertiary structures. In addition to the classical double helical structure of nucleic acids, they form structural motifs like hairpin, pseudoknot, bulge, internal loop, cruciform, triplex, tetraplex (G-quadruplex (GQ) and i-motif (iM), Z-DNA, etc.[1] Unlike the chemical diversity in proteins, nucleic acids use only four nucleosides (A, G C, and T/U) to form these structures. Formation of such diverse structures is aided by canonical and non-canonical base pairing interactions, stacking interaction and interactions of nucleobase, sugar and phosphate groups with various metal ions.[2] The structure thus formed has intrinsic conformation dynamics, which interconverts between different conformations depending on the environment and cognate protein factors and small molecule metabolites.[3] This structural reorganization not only controls the ensuing function of the nucleic acid sequence but also has therapeutic implication as aberrations in the structure often lead to disease states.[4]

Among the various structural motifs, four-stranded structural motifs GQ and iM have attained much of the attention owing to their (i) abundance in the genome,[5,1c] (ii) important cellular functions[6] and (iii) direct evidence of formation of these structure in cell.[7] It was observed that the functions of these structures are associated with their location in the genome. The tetraplex structures located in the telomeric region of the chromosomes, promoter DNA regions and untranslated regions of mRNA of several proto-oncogenes play a crucial role in telomere maintenance, transcriptional and translational regulation, respectively.[8] Additionally, other noncanonical nucleic structures like Z-DNA, triplex helix was discovered long before.[9] Although the exact biological role of these structures is not well understood, recent studies have indicated that the formation of these structures could be connected to different disease states.[10] Along with these structural motifs, metal ion-mediated base pairs have been considered as non-canonical structural components in nucleic acids.[11] The interactions of metal ion with nucleic acid have been observed to be biologically important and helpful in devising functional materials.[12] Apart from the vital biological functions of all these structural motifs, they are structurally quite diverse and their folding topology and stability largely depend on the sequence, ionic conditions, pH and surrounding environment like molecular crowding and confinement.[13,10b,6c] Owing to the structural diversity and potential role in diseases, it is imperative to distinguish different topologies of a

specific nucleic acid structural motif and understand their structure-function relationship both *in vitro* and *in vivo.*

Several biophysical techniques like circular dichroism (CD), NMR, EPR, X-ray crystallography and fluorescence have been employed in the understanding of the structure-function relationship of nucleic acids *in vitro.*[14] Majority of these techniques rely on labeling of nucleic acid with appropriate tags. For example, NMR, EPR, X-ray crystallography use isotope-, spin-, heavy atom- and fluorescently labeled nucleic acid seqeunces, respectively.[15] Notably, responsive nucleoside probes, incorporated site-specifically into nucleic acids have been successful in providing valuable information about nucleic acid structure, dynamics and interaction with ligands, metal ions and proteins. However it is not mandatory that the structures formed *in vitro* conditions can also be maintained in cell. Hence the idea of studying nucleic acid structures in native cellular environment is getting more importance in recent days.[16] Studying nucleic acid structure-function relationship in complex cellular environment is not very straightforward. In this context, structure-specific antibody and fluorescent light-up probes have greatly advanced our understanding of nucleic acid structure in cellular environments. Additionally, in-cell NMR and EPR have emerged as potential tools in investigating the effect of native cellular environment on the stability, conformational and binding event of various nucleic acid structures.[17] Although these methods are quite efficient, each of them has some disadvantages, which limit their application.

In this chapter, different nucleic acid structures, particularly non-canonical motifs and their roles in important cellular process are discussed. The development and use of various biophysical tools that have aided the understanding of structure, dynamics and recognition properties of nucleic acid are presented. In particular, design and applications of base-modified nucleoside analogs in probing noncanonical nucleic acid structures are discussed in detail. The last part of this chapter is focused on the currently available methods used for structural studies of nucleic acid in the complex cellular environment. Specially, the utilization of structure-specific antibodies, small molecule probes, in-cell NMR and EPR to study nucleic acid in cells is provided. Finally, the current challenges in probing nucleic acids in cell-free and cellular environments and motivation for the work presented in this thesis are discussed.

1.2 Non-canonical nucleic acid structures

1.2.1 G-quadruplex

The four-stranded GQ assembly is composed of stacks of two or more G-tetrads formed by multiple H-bonds and stabilized by metal ions (e.g., K^+, Na^+, etc.) (Figure 1).[18,13a] In general, intramolecular GQ is formed by sequences having four or more G-tracts, which involve in tetrad formation. The intervening nucleotides, which form the loops, connect the tetrads. While tetrad stacks are almost exclusively formed by guanines, the bases involved in loop formation are not limited. GQ exhibits wide variety of folding topologies, which can be broadly classified as parallel stranded, antiparallel stranded and hybrid type mixed parallel-antiparallel stranded conformations (Fig. 1B). These structures differ in the (i) strand orientation, (ii) loop type (propeller, diagonal and lateral) and length, (iii) groove width (wide, medium and narrow), and (iv) guanosine glycosidic conformation (*syn* and *anti*) of the tetrads. The ability of a given G-rich region to adopt one or more GQ structures primarily depends on the nucleotide sequence itself.[13a] In addition, surrounding environment and assay conditions can considerably influence the conformation of the GQs.[13c]

DNA GQs

Human telomeric (H-Telo) DNA overhang composed of a $(TTAGGG)_n$ repeat is the most studied GQ-forming sequence. It forms an antiparallel structure in Na^+ solution and multiple conformations in the presence of K^+ ions (Figure 1).[19] In a buffer solution containing K^+ ions, telomeric repeat mainly adopts hybrid 1 and hybrid 2 conformations in which the double chain reversal loop is located at the 5′- and 3′-end, respectively.[20] However, under crystallization conditions in the presence of K^+ ions and polyethylene glycol (PEG), the same sequence adopts a parallel GQ conformation.[21,13d] Structural analysis performed in the presence of crowding agents (PEG), highly viscous cosolutes (deep eutectic solvents) and polysaccharides, which are commonly used synthetic models mimicking the confinement and molecular crowding of cells, show that parallel GQ structure is the preferred structure of the telomeric repeat.[13c,22] Further, studies in these systems reveal a slower folding kinetics with reduced stability and ligand binding ability of GQ as compared to in dilute aqueous buffers. On the contrary, a recent study using optical tweezers indicates that telomeric DNA repeat in the confined space of DNA nanocages forms hybrid type or anti-parallel structure, depending on the size of the confinement.[23]
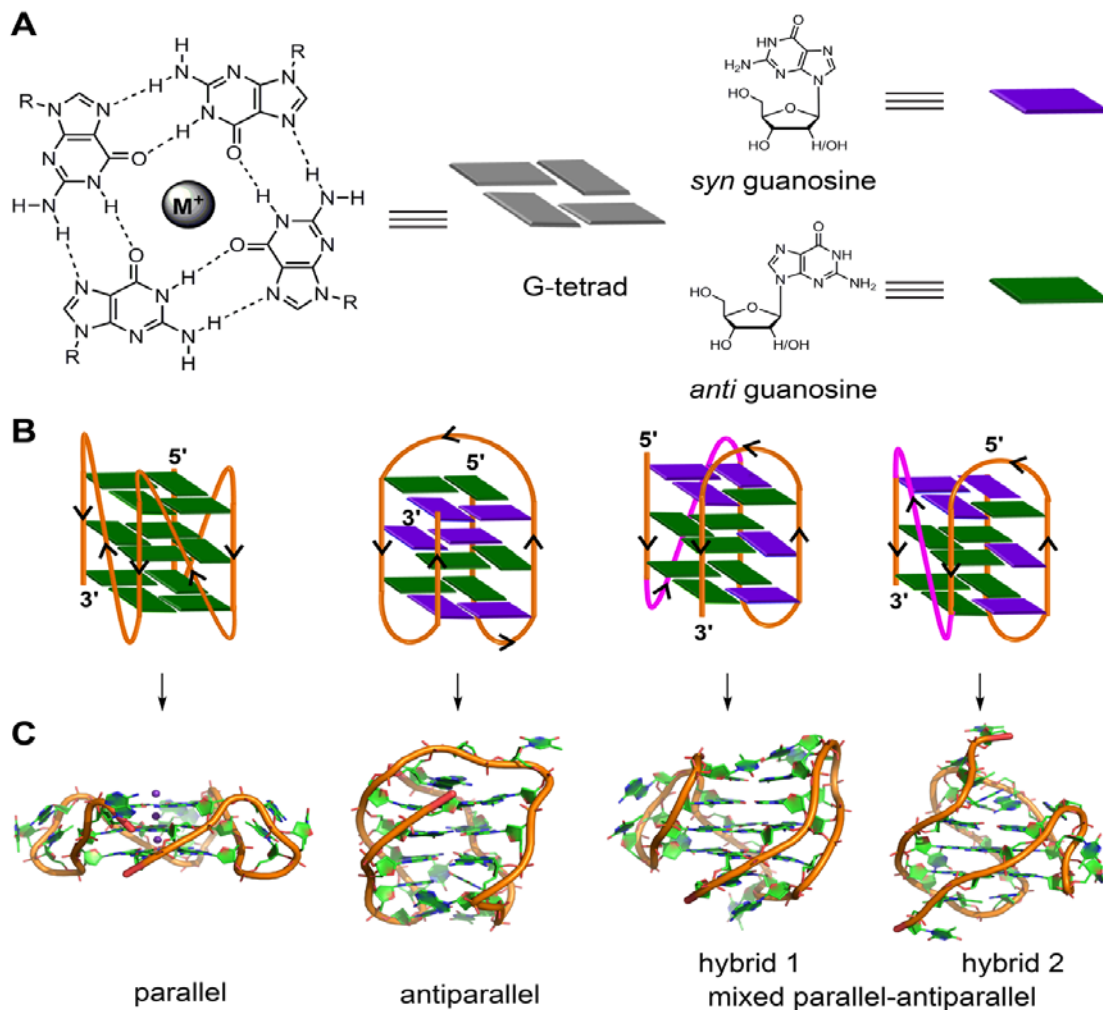
Figure 1. (**A**) Schematic representation of a G-tetrad where four coplanar guanines are involved in Hoogsteen and Watson-Crick hydrogen bonding. Metal ions stabilize the tetrad. For representation, guanines are shown as rectangular box. *syn* and *anti* glycosidic conformations of guanosine are represented in violet and green colour, respectively. (**B**) Loop orientation and tetrads of different GQ structures, namely parallel, antiparallel, hybrid 1 and hybrid 2 are shown. (**C**) Crystal/NMR structure of parallel (PDB ID: 1KF1),[21] antiparallel (PDB ID: 143D),[19] hybrid 1 (PDB ID: 2GKU)[20a] and hybrid 2 (PDB ID: 2JPZ)[20b] conformations of H-Telo DNA ON.

Unlike telomeric DNA repeat, the GQ-forming sequences present in the promoter region of oncogenes have varying numbers of G-tracts with unequal number of guanines and intervening loop residues.[13a] Hence, these motifs can potentially form multiple GQ structures by using different combinations of the G-tracts and loop residues. G-rich motif (NHE III$_1$) present in the promoter region of the *c-MYC* gene has five tracts of guanine residues each separated by a nucleotide residue. The I–IV tracts and II–V tracts both form parallel GQ structure with 1:2:1 loop-size arrangement (Figure 2A).[24] Other promoter sequences, which form parallel GQ structure are *VEGF*, *HIF-1α*, *c-Kit*, *RET* and *PDGF-A*.[25]

Upstream of P1 promoter region of the human *BCL-2* gene is a 39-bp GC rich sequence containing six G-tracts.[26] Inhibiting the expression of this gene reduces cell proliferation, thereby improving chemotherapy efficacy.[27] Notably, the P1 promoter region, including the GC rich segment, has been shown to play a significant role in regulating the transcription of *BCL-2* gene.[28] This G-rich sequence can form three GQs involving different G-tracts (Figure. 2B). The middle segment (four consecutive G-tracts) forms a stable mixed parallel-antiparallel-stranded hybrid type GQ structure with two lateral loops (3 nucleotides and 7 nucleotides, respectively) and one single nucleotide double chain reversal loop.[26] However, recent studies indicate that four nonsuccessive G-tracts (I, II, IV and V) fold into a parallel GQ structure containing three loops, which is unexpectedly more stable than the GQ of middle segment.[29]
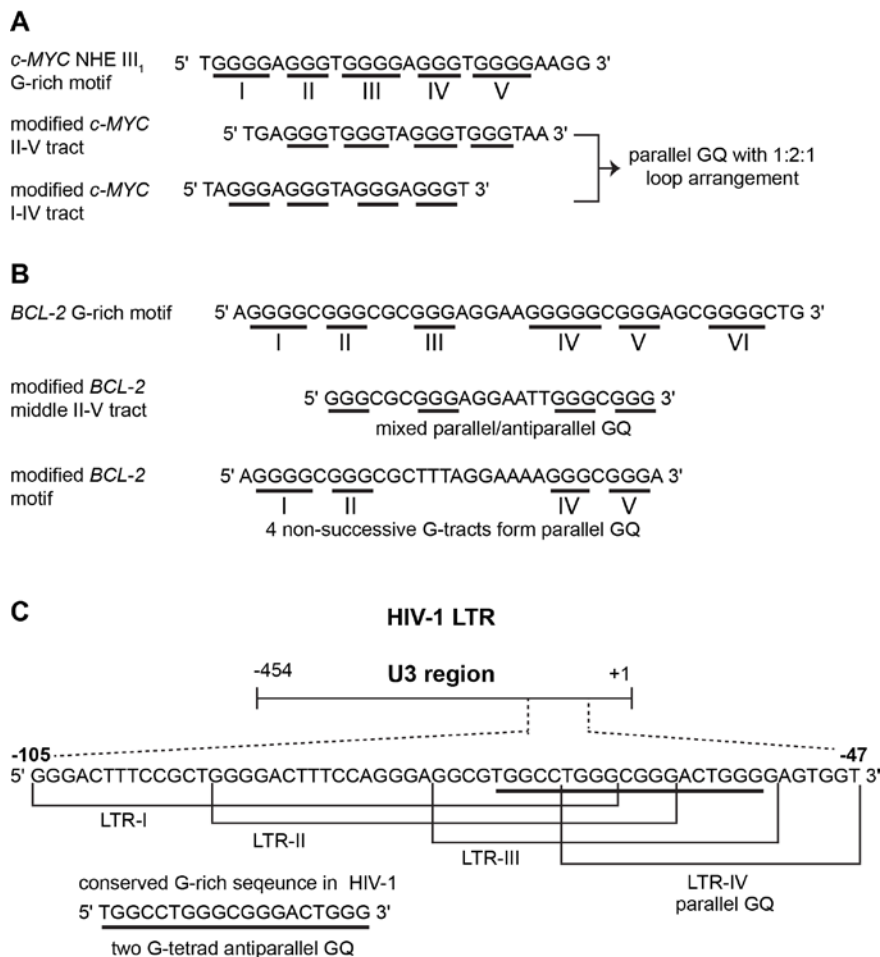
Figure 2. (**A**) The G-rich NHE III$_1$ sequence with five G-tracts in the promoter region of *c-MYC* gene. G-tracts are underlined. Modified *c-MYC* sequences, which are used for structural analysis is shown. (**B**) The sequence of GQ-forming *BCL-2* promoter with six G-tracts is shown. Different GQs are formed by different G-tract combinations. (**C**) The dynamic GQ-forming sequence in the U3 region of 5′ LTR of HIV-1 is shown. It has four overlapping GQ-forming segments.

Recently, a dynamic GQ forming segment has been identified in the U3 domain of the 5′ long terminal repeat (LTR) of proviral HIV-1 genome (Figure 2C).[30] The U3 region contains a stretch of eight G-tracts upstream of the transcription start site and overlaps with the binding site of important transcription activators, Sp1 and NF-κB. This G-rich sequence (-105 to -47) is sub-divided into four overlapping G-tract segments (LTR I–IV), each containing four G-tracts. When individually studied, LTR-II and LTR-III produced circular dichroism (CD) signatures resembling a parallel-type GQ structure, whereas, CD profile of LTR-I did not match with any of the GQ forms. Similarly, LTR-IV formed a parallel-stranded GQ structure containing a single-thymidine bulge.[31] However, studies with full-length G-rich sequence showed the presence of multiple GQs. In another study, the central part of the G-rich region, which shows high level of conservation, adopted a two G-tetrad antiparallel GQ structure.[32] Further, mutations that disrupt GQ formation and by using ligands that stabilize GQ structures, it was inferred that the LTR of HIV-1 could serve as a new target for anit-HIV-1 drug screening.

RNA GQs

Unlike the structural diversity exhibited by DNA GQs, G-rich RNA sequences generally form all-parallel GQ structure, which is thermodynamically more stable and less hydrated than equivalent DNA GQs.[33] For example, telomeric repeat-containing RNA (TERRA) $(UUAGGG)_n$, an equivalent sequence of the human telomeric DNA repeat, adopts a parallel GQ structure irrespective of the type of metal ions and surrounding environment.[34] GQ-forming motif has been identified in different positions of UTR of mRNAs, which includes *Zic-1*, *NRAS*, *BCL-2*, *TRF2*, and IRES element of human of *VEGF*.[35,33a] These motifs form a parallel GQ structure with different loop configurations *in vitro*.[36]

1.2.2 i-motif DNA

Analogous to the GQ structures formed by G-rich DNA sequences, C-rich sequences form four-stranded structures called iMs under slightly acidic or even at near neutral pH by intercalation of hemiprotonated C•CH$^+$ base-paired strands (Figure 3A).[37] Depending on the spatial arrangements of C:CH$^+$, the iM structure can be classified into two different intercalation topologies: 3'E, where the outmost C:CH$^+$ base pair is at the 3'-end and 5'E where the outmost C:CH$^+$ base pairs is at 5' end.[6c] The stability of the iM structure largely depends on C-tract length, ionic strength, molecular crowding, superhelicity and pH. While

putative GQ- and iM forming sequences mostly coexist as partners in the genome, the majority of studies have been focused on GQs. This is because there is clear evidence for the presence and function of GQs *in vivo*,[7a-c] while the existence of iM structure in cellular environment has been remained ambiguous until a very recent report, which provides the evidence of iM formation in human cell.[7d] Moreover, the protein factors have been identified, which preferentially bind to C-rich sequences and transcriptionally activate the expression of certain oncogenes.[38] In one such example, hnRNP LL has been shown to bind to *BCL2* promoter iM-forming sequence and activate *BCL2* expression.[39] The role of iM as a molecular switch in controlling the expression of an oncogene has been shown using small molecule binders.[39,40] These findings suggest that iM and GQ structures could be cooperatively targeted by using structure-specific small molecule binders.[41] The evidence of iM structure in cell as well as its regulatory role in gene expression have established iM as a novel potential therapeutic target in recent years.[42]

1.2.3 Triple helix DNA

Three stranded triple helix structure (Figure 3C) is one of the important non-canonical DNA structures and has been suggested to have critical biological roles in gene regulation.[3d,10a] Triplex DNA is mainly formed by a polypurine·polypyrimidine tract, and a third single-stranded DNA, which binds to the major groove of the duplex by Hoogsteen or reverse Hoogsteen base pairing in a parallel or an antiparallel orientation. DNA triplex can be intramolecular, called as H-DNA, whether the third stand comes from the mirror repeat of other half of the same sequence or can be intermolecular where the third strand is a different oligonucleotide (ON), called as triplex-forming ON (TFO). The triplets, which are formed in triplex structures are T*A•T, G*G•C, A*A•T, C$^+$*G•C. etc. Some of the triplets have been depicted in Figure 3C. This structure formation has been indicated to induce different biological consequence like blockage of replication, transcriptional repression, mutagenesis and genetic recombination.[43]

1.2.4 Z-DNA

Z-DNA is a unique left-handed double helical DNA structure, which is formed by alternating pyrimidine-purine sequence e. g. $(CG)_n$ (Figure 3B).[3d] Zig-zag arrangements of the backbone, alternating *syn* and *anti* conformation of nucleobases, deep single groove and placement of the bases away from the axis are the characteristic features of the Z-DNA structure.[9b,44]

Stimuli like high ionic strength, negative supercoiling, protein binding, and chemical modification have been observed to induce the structural transformation from B-DNA to Z-DNA.[10b] Moreover, presence of Z-DNA forming sequence near several promoter regions and discovery of Z-DNA binding protein have strengthened the possible role of Z-DNA in gene regulation. Recently C8-guanine adducts, which induce Z-DNA formation are being utilized to understand the probable role Z-DNA in carcinogenesis.[10b]
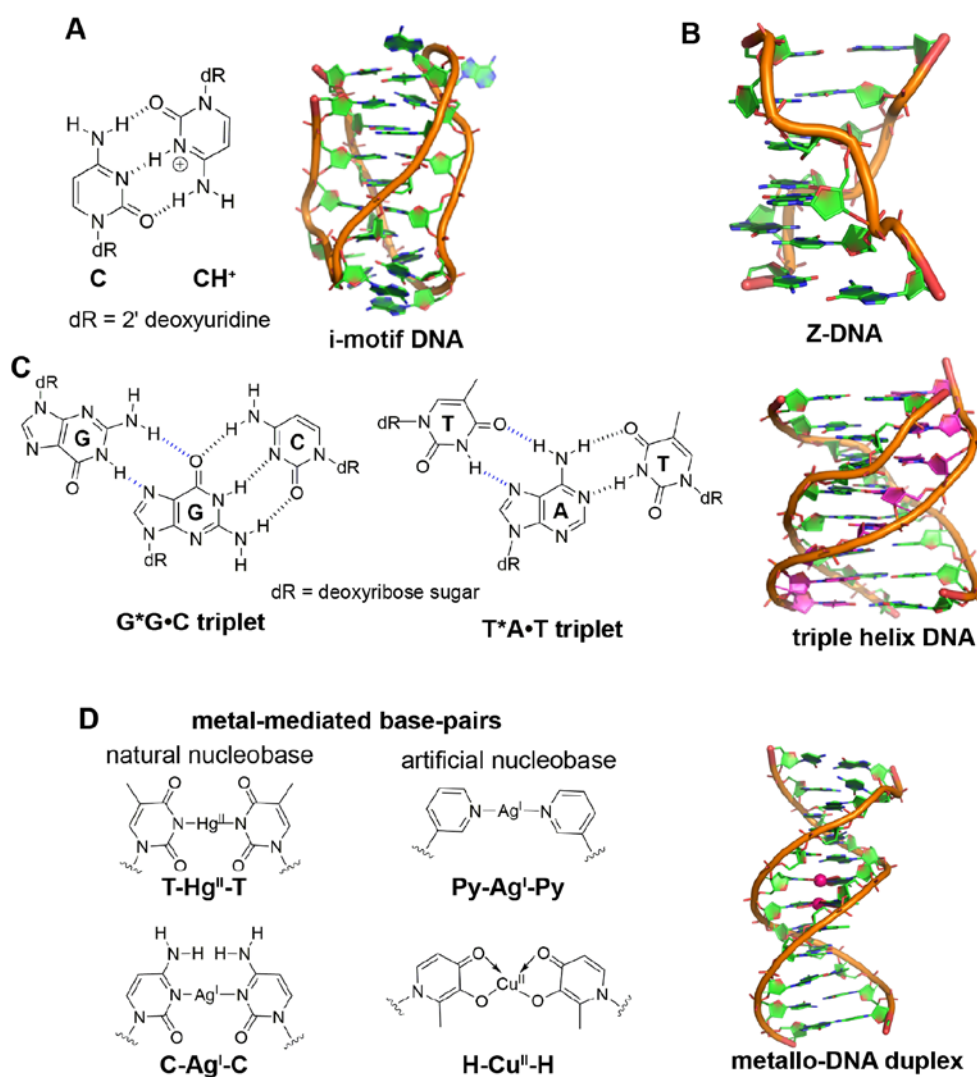
Figure 3. (**A**) Schematic representation of C:CH$^+$ base-pair and three-dimensional (3D) NMR structure of an iM DNA (PDB ID: 1EL2).[45a] (**B**) 3D crystal structure of a Z-DNA (PDB ID: 1I0T).[45b] (**C**) Schematic representation of selective base triplets observed in triplex DNA structure. Hoogsteen hydrogen bonding is represented in blue and Watson-Crick hydrogen bonding is represented in black dotted line. Three-dimensional NMR structure of a triple helix DNA, where the bases in third strand are represented in pink colour (PDB ID: 1D3X).[45c] (**D**) Schematic representation of some metal-mediated base-pairs involving either natural nucleobases or artificial nucleobases and crystal structure of a metallo DNA duplex containing T-Hg^{II}-T base pairs. Hg^{II} ions are represented in magenta color.[45d]

1.2.5 Non-canonical metal-mediated base-pair

To expand the scope of the nucleic acid, apart from regular Watson-crick hydrogen bonding, base pairing by coordination to metal ions, has appeared as an attractive approach.[11] Non-canonical metal-mediated base-pairings have been observed to stabilize duplexes similarly or more than the regular Watson-Crick base pairs.[12] Metal mediated base-pairing can be two types: (i) base pairs with natural nucleobases for example, T-Hg^{II}-T, C-Ag^{I}-C etc and (ii) base pairs with artificial nucleosides for example, pyridine-based Py-Ag^{I}-Py, hydroxyquinoline-based H-Cu^{II}-H etc (Figure 3D). These metal-mediated base-pairs mainly have been utilized to develop ON-based toxic transition metal sensor, conducting material and detection of single nucleotide polymorphism.[46] Interestingly some of these metallated base-pairs were successfully accepted by enzymes like DNA polymerases.[47] Researchers are putting efforts to find out the biological impact of these metallo base pairs as well as to develop new metal induced functional DNA molecules.

1.3 Techniques to study nucleic acid structure and ligand binding *in vitro*

1.3.1 CD and UV absorption spectroscopy

The structure, stability, folding dynamics and ligand binding ability of ON sequences are commonly studied *in vitro* by using CD, UV-absorption, fluorescence, NMR, EPR and X-ray crystallography techniques. Among them, CD provides a qualitative understanding of the various nucleic acid structures, and can be used to distinguish them as well as various conformation of a particular nucleic acid structure.[48] This is because the CD signal is highly sensitive to strand orientation and *anti/syn*-glycosidic conformation of the nucleosides in a particular structure.[48] A B-DNA duplex shows a typical CD with a positive band at ~260–280 nm and a negative band at ~245 nm. All topology of GQ structure has characteristic CD profile.[49] Parallel GQ structures of DNA and RNA ONs typically exhibit a strong positive band at ~260 nm and a shallow negative band at ~240 nm.[49] A CD spectrum containing a positive peak at ~295 nm, a negative peak at ~265 nm and a smaller positive peak at ~245 nm is characteristic of a chair- or basket-type antiparallel GQ structure.[19a,49] Hybrid type mixed parallel-antiparallel GQ structures show a positive band (~290 nm) with a shoulder near 265 nm and a smaller negative band at ~240 nm. On the other hand, iMs typically display a positive band at ~288 nm and a negative band at ~256 nm.[50] The Z DNA showed almost an inverse CD spectra compared to B-DNA; it displays a negative band ~ 290 nm, a positive

band ~ 260 nm and a characteristics deep negative band ~ 203 nm.[48] While CD analysis is highly useful in studying individual structures, the spectrum is not easy to interpret when multiple conformations are present. For example, telomeric DNA repeat is known to fold into multiple GQs in K^+ ionic conditions. However, the CD spectrum resembles more like hybrid type structures.[19b] Determination of conformation of the sequences containing longer G-tracts, which can fold into multiple structures, is also not straight forward by CD.[51]

Thermal melting is another convenient method to evaluate the stability of different nucleic acid structures.[52] The melting temperature of most of the nucleic acid structure by UV-absorption method is mainly calculated by monitoring the changes in absorbance at 260 nm.[52,53] As the structured nucleic acid melts, unstacking of the bases and exposure of them to solvent enhance the absorbance at 260 nm. However the melting temperature of GQ is mainly calculated by monitoring the change in absorbance at 295 nm due to larger variations (50−80%) in absorbance at 295 nm compared to at 260 nm (4%) upon GQ melting.[54] As GQ structures have higher absorbance at 295 nm, the GQ to random coil transition shows a typical inverted melting profile as compared to the classical duplex to single-strand transition profile. Apart from UV absorption, temperature-dependent CD and FRET assays have also been used to study the formation as well as binding of ligands to particular GQ structure.[54] However, sequences showing multiple conformations pose challenges in deconvoluting the signal from individual structures. Hence, CD and thermal melting are used as complementing techniques along with fluorescence, NMR or X-ray crystallography in determining the structure and ligand binding properties of nucleic acid.

1.3.2 NMR spectroscopy

NMR spectroscopy is a powerful tool to analyze nucleic acid structures in solution at atomic resolution.[55,14g] NMR techniques have also been used in studying the folding kinetics, dynamics, and in determining the binding of ligands to various domains of different nucleic acid secondary structures. Simple 1D proton NMR analysis has been used to study the formation of different nucleic acid structures. For example, GQs and iMs structures give rise to characteristic NMR signals for imino protons as compared to that of the duplex structure.[56] Chemical shift of guanine imino protons of GQs comes within the range of 10−12 ppm compared to the duplex, which falls in the range of 13−14 ppm. The iM structures formed by C•CH+ base-paired strands give rise to imino proton signals in the range of 15−16 ppm. In triplex structure, imino proton resonance corresponding to Watson-Crick and Hoogsteen

base-pairing appeared between 12–16 ppm. However, when sequences exhibit multiple conformations, signal overlap makes the structural analysis very difficult. In this context, slight changes in the sequences have been made to favour the formation of a particular conformation, thereby allowing detailed structural analysis of that nucleic acid conformation.[20,55] For detailed structural analysis nucleic acid has been labeled with ^{13}C, ^{15}N either by enzymatic incorporation of uniformly or partially labeled ^{13}C and ^{15}N enriched rNTPs using T7 RNA polymerase or by site-specific incorporation of isotope labeled nucleoside phosphoramidite using solid phase ON synthesis.[15] Few examples of ^{15}N and ^{13}C labeled nucleoside which have been used for nucleic acid structural analysis are listed in figure 4A. However ^{15}N and ^{13}C labeled precursors for both solid phase and enzymatic synthesis are expensive and difficult to synthesize. Structure determination of various nucleic acid structures is achieved by performing a combination of NOESY, COSY, TOCSY, ^{13}C-^{1}H HMBC and ^{13}C-^{1}H and ^{15}N-^{1}H HSQC experiments using either native or isotope-labeled ON sequences.[20,57,55] Moreover, the mode of binding of ligands to various secondary structures has been studied by using NMR.[58] Mainly the information about nucleic acid structure, dynamics and its recognition can be acquired from these NMR methods up to 100 nucleotides.[55a] Due to complex resonance overlap, it is very difficult to determine the structure of the large nucleic acid molecules.

Recently ^{19}F NMR has appeared as a new approach to investigate conformational change of nucleic acids and their interactions with small molecule and proteins.[59] ^{19}F-labeled nucleoside probes have been introduced into oligonucleotide to investigate equilibrium between alternative RNA structures,[60a] DNA and RNA secondary structures,[60b] structural transitions,[60c,60d] RNA-ligand[60e] and nucleic acids-protein[60f] interactions by ^{19}F NMR analysis (few [19]F -labeled nucleoside have been depicted in figure 4A). In other reports by Xu and co-worker, 3,5-bis(trifluoromethyl) benzene modified deoxyguanosine (**7**) and deoxyuridine (**8**) residues distinguished G-quadruplex structure formed by G-rich H-Telo DNA repeat and thrombin binding aptamer from their respective single-stranded ON by displaying distinct ^{19}F NMR signal in quadruplex and single-stranded state (Figure 4A).[61] In another report, 5-trifluoromethyl cytidine ([TF]C, **9**) and 5-trifluoromethyl thymidine ([TF]T, **10**) -modified ONs exhibited distinct ^{19}F NMR chemical shift in single-stranded DNA, B-DNA and Z-DNA and detected the B-DNA to Z-DNA transition at a high Salt concentration (Figure 4A).[62]

1.3.3 EPR spectroscopy

EPR is another spectroscopic technique, useful in investigating nucleic acid structure and dynamics.[14g] EPR technique relies on the magnetic moments of unpaired electrons. As nucleic acids are diamagnetic, spin-centers have to be incorporated in nucleic acids to perform EPR experiments.[15] In EPR analysis nitroxide radical is used widely due to its thermodynamic stability. Few selective nitroxide radical-labeled nucleoside probes are depicted in Figure 4B. The spin labels are introduced into ON by solid-phase chemical synthesis or by post synthetic modification. Two approaches are used in EPR, one is the attachment of the spin label at site of interest and monitoring the changes in its surrounding environment and another one is measuring the distance between two spin labels.[14g] A particular EPR experiment PELDOR (pulsed electron double resonance) was used to probe analyte binding to aptamers,[63a] protein induced structural change in nucleic acid structure,[63b] RNA dynamics,[63c] metal ion-RNA,[63d] small molecule–RNA inrteractions[63e] by using different nitroxide-labeled nucleoside probes. EPR techniques were exploited further to investigate non-canonical structures like G-quadruplex. Drescher and co-worker investigated G-quadruplex conformations adopted by H-Telo DNA repeat in K^+ ionic solution by using double electron-electron resonance (DEER) spectroscopy.[64] They doubly labeled H-Telo DNA with nitroxide spin in such a way that the distance between the two spin labels would be different in different GQ topologies (parallel, antiparallel and (3+1) hybrid structure. Distance measurements by EPR in K^+ solution reveal the existence of two species, parallel and antiparallel GQ forms. In another report, nitroxide spin label was attached at phosphate backbone in multiple positions of H-Telo DNA repeat and the distance measured between them provided the information about preferred conformation in different condition like in NaCl, KCl and in presence of glycerol and small molecule ligands.[65]

1.3.4 X-ray crystallography

X-ray crystallography has been instrumental in providing 3D structural information of various nucleic acid secondary structures sequences as well as their modes of binding (tetrad, groove or loop) to small molecule ligands.[66] Phasing techniques in X-ray crystallography relies on derivatization with heavy atoms, heavy-atom soaking, chemical modification of nucleosides with halogen or selenium atoms. Particularly, selenium gives rise to a strong anomalous scattering signal, which has facilitated the crystallographic analysis of nucleic

acid. Few selenium-modified nucleosides used for nucleic acid crystallization are shown in Figure 4C. Given the structural polymorphism exhibited by nucleic acids there is always a concern whether the structure favoured by a sequence in crystallization conditions will also be preferred under equilibrium conditions in a highly complex cellular environment.[68]
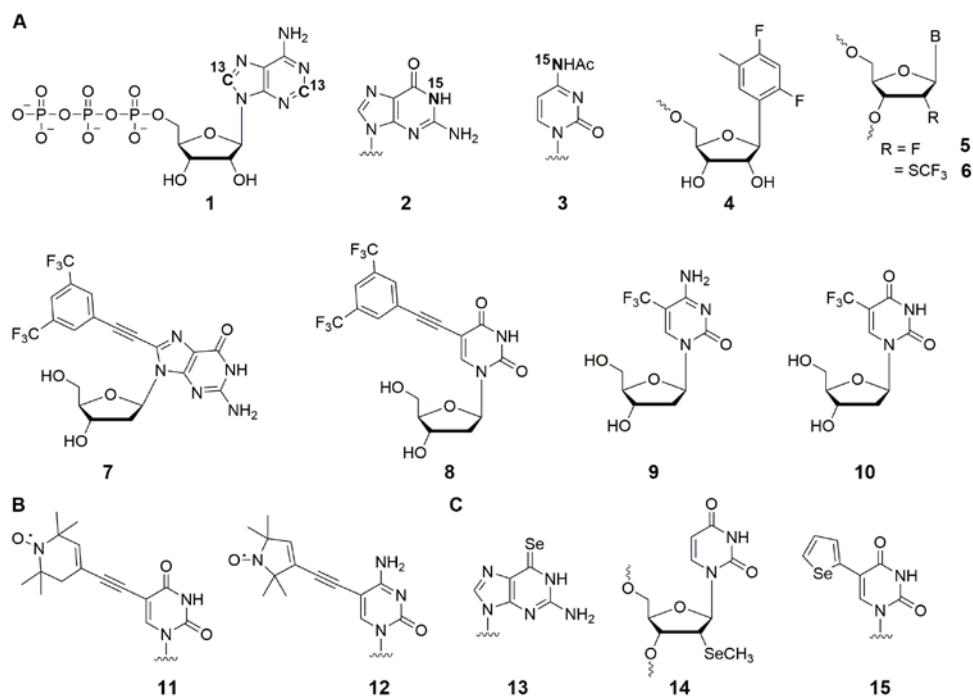
Figure 4. (**A**) Selected examples of ^{13}C, ^{15}N, and ^{19}F-labeled nucleotide or nucleoside used to tag nucleic acid for NMR analysis. Selected examples of (**B**) spin-labeled and (**C**) selenium-labeled nucleoside probes used in EPR and X-ray crystallography analysis of nucleic acid, repectively.

1.3.5 Fluorescence spectroscopy

Among the various biophysical methods, fluorescence based methods are advantageous as they provide easy tools to study the structure, folding dynamics and recognition of nucleic acid in real time with a high level of sensitivity.[14b,14c,69] Since the native nucleosides are practically non-emissive, the study of nucleic acid structures is accomplished by using non-covalent fluorescent binders or by labeling the ONs with fluorescent nucleoside reporters. These fluorescent probes show changes in their fluorescence properties like emission maximum, intensity, quantum yield, anisotropy and lifetime upon changes in nucleic acid conformation. Fluorescence Resonance Energy Transfer (FRET), structure-specific fluorescent probes and fluorescent nucleoside probes are very useful fluorescence-based methods in studying different nucleic acid structures, structural transition and recognition events.[69] Few examples of different fluorescence-based method, particularly the utility of

covalently labeled environment-sensitive fluorescent nucleoside analogues in distinguishing different nucleic acid structure and estimating ligand-binding have been presented here.

1.3.5.1 FRET and structure-specific fluorescent probes

FRET-based assays are widely used in investigating the formation, stability and dynamics of different nucleic acid structures like GQ, iM etc. at the ensemble as well as single-molecule level.[70] Moreover FRET-based method has reported the structural transition between different nucleic acid secondary structures for example qudruplex-duplex,[71a,71b] triplex-duplex,[71c] B DNA-Z DNA[71d] etc. ONs labeled with appropriate FRET pairs are used in assessing the selectivity of structure-specific binding of ligands,[72] and also in devising fluorometric methods to detect metal ions and proteins.[73]. Similarly, several structure-specific responsive fluorescent ligands and metals complexes that display changes in their photophysical properties in the bound and unbound states serve as good sensors to detect the formation of various GQ, iM structures and metal-mediated duplex formation.[74] While the utility of these tools are arguably undeniable, probing the conformational differences in different nucleic acid structures at the nucleotide level is not straightforward.

1.3.5.2 Fluorescent nucleoside probes

Along with global structural changes, distinct conformational changes, difference in interactions with neighbouring bases as well as with surrounding environment happen at nucleoside level during nucleic acid folding and recognition event. Hence a detailed understanding of the nucleic acid structures at the nucleotide level is important to understand the structural diversity of different nucleic acid structures and estimating affinity of small-molecule ligands to different domains of a particular structure. In this context, conformation-sensitive fluorescent nucleoside analogues that can report the subtle difference in nucleic structures via changes in fluorescence properties have been highly useful.[69b] Some of these probes are potentially suitable for designing screening platforms to identify structure-specific binders of therapeutic relevance.

1.3.5.2.1 Isomorphic fluorescent nucleoside probes

2-aminopurine (2-AP), a structural isomer of adenine which can form Watson-Crick type base pair, high quantum yield in water ($\Phi = 0.68$) and sensitivity to the solvent polarity has been used widely in studying different secondary structures of nucleic acid and their ligand

15

binding. Majima and coworkers demonstrated the usefulness of 2AP in reporting GQ structure formed by H-Telo DNA ON repeat AGGG(TTAGGG)$_3$ (Fig. 5A).[75a] Replacement of the loop adenine residues with 2AP was structurally non-invasive revealed and the fluorescent label reported the formation of antiparallel GQ structure by showing significantly enhanced fluorescence intensity and longer lifetime as compared to the complementary duplex. The enhancement in fluorescence intensity and longer lifetime was attributed to the more solvent exposed and less stacked 2AP in GQ structure as compared to in the duplex structure. The sensing ability of 2AP was further utilized in monitoring the binding of TMPyP4 to the antiparallel GQ.[75b] Interestingly, modification at diagonal loop produced the maximum response upon binding to the ligand, which was found to be consistent with the binding mode of ligand TMPyP4 to GQ, i.e., ligand inserted between the diagonal loop and G-tetrad (Fig. 4A).
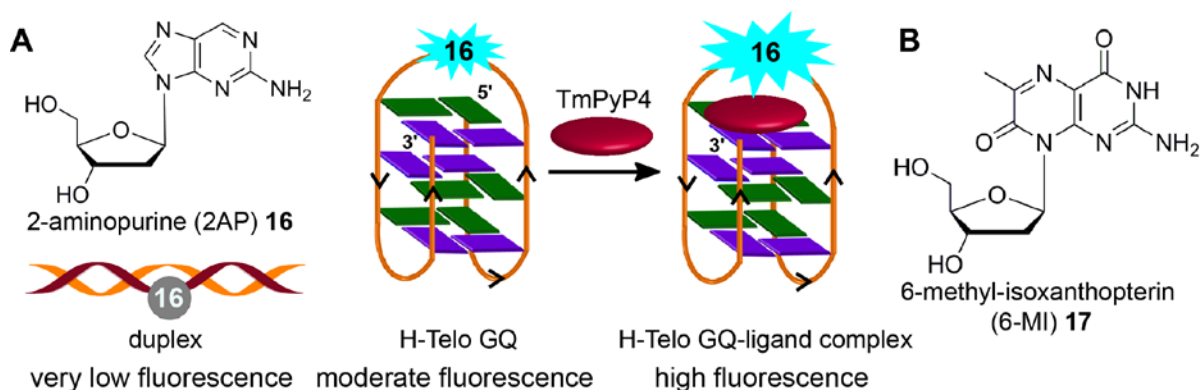
Figure 5. (**A**) 2AP displays different fluorescence properties in duplex, GQ and GQ-ligand complex.[75a] (**B**) Structure of fluorescent 6-MI 2′-deoxyribonucleoside is shown. The figure is adopted from reference 69c.

Chaires and coworker utilized 2AP-modified H-Telo DNA ONs to study the formation of GQ structures in different ionic conditions. The results obtained from these studies indicated that GQ structure adopted by telomeric repeat in solution containing K$^+$ ions and in solid-state are not same.[76] In another study conducted by Sugiyama and Xu, 2AP-substituted G-rich and C-rich sequences present in the 5′ end of retinoblastoma susceptibility genes (Rb) were employed in probing the competition between duplex, GQ and iM formation.[77] Both GQ and iM exhibited considerably higher fluorescence intensity as compared to the duplex form. Addition of complementary C-rich ON sequence to the pre-formed 2AP-labeled GQ resulted in quenching in fluorescence intensity at neutral pH as a consequence of slow conversion of GQ to duplex form. At acidic pH the formation of iM

structure facilitated the formation of GQ, thereby providing a competitor for the formation of duplex relative to the random coil of the C-rich ON sequence at neutral pH. In a recent study, Kankia and coworkers used 2AP to study the folding of thrombin and G3T aptamers into antiparallel and parallel GQ structure, respectively.[78] 2AP when was introduced in the loop regions of thrombin aptamer d(GGTTGGTGTGGTTGG) and G3T aptamer d(GGGTGGGTGGGTGGG) signaled the formation of respective GQ structures in the presence of K^+ ions, with enhancement in fluorescence intensity compared to unfolded state. However, the general drawback of 2AP is that it has excitation and emission maximum in the UV region and importantly, the quantum yield of 2AP incorporated into ONs and into duplexes is very low. For example, the quantum yield of free 2AP is 0.68, and when incorporated into GQ-forming sequences and corresponding duplexes the quantum yield decreases dramatically (up to 0.06 and 0.005, respectively), which compromises the sensitivity of the probe.[75a]

Another fluorescent purine anlogue 6-methyl-isoxanthopterin (6-MI), a fluorescent purine analog, significantly accelerated the formation of the GQ and showed drastically quenched fluorescence in the GQ state due to stacking interaction with a neighboring purine.[79] The responsiveness of 6-MI was used by Shamoo and coworkers in estimating the binding of unwinding protein (UP1) to the H-Telo repeat.[80] Upon addition of increasing concentration of UP1, the probe-labeled telomeric repeat DNA ON exhibited a dose-dependent increase in fluorescence intensity as a result of unstacking of 6-MI in the complexed state. The affinity of UP1 for the telomeric repeat was determined to be ~5 nm, and the results also indicated that the protein destabilized the GQ structure. In another report by Zhurkin and co-worker, 2AP and 6MI have reported the melting of parallel triplex DNA structures containing 2AP*T•A and 6MI*C•G triplets by exhibiting changes in the fluorescence intensity.[81]

1.3.5.2.2 Heterocycle-conjugated and heterocycle-fused pyrimidine nucleoside probes

Tor's group established a series of fluorescent nucleoside probe by either conjugating or fusing aromatic five-membered heterocycles for example of furan-, thiophene-, oxazole-, and thiazole- to pyrimidine bases.[69a] Furan-conjugated uridine (**18a**) and deoxyuridine (**18b**) analog showed emission maxima in the visible region and high sensitivity to changes in solvent polarity.[82] These fluorescent probes were successful in the detection of abasic sites[82a] and in reporting RNA-aminoglycosides[82b] and RNA-protein interaction.[82c] Tor and co-worker

also synthesized a thiophene-fused uridine analog which enabled the detection of mismatches in duplex and a depurination site in RNA, generated by ribosome inactivating protein (RIP) toxins.[83] Srivatsan and coworkers introduced a series of ribonucleoside analogs by conjugating indole, *N*-methyl indole, benzothiophene and benzofuran at the 5-position of uridine (**19−22**) (Figure 6).[84] The design of heterobicycle-conjugated nucleosides is based on the intrinsically fluorescent indole core of the amino acid tryptophan. Rewardingly, 5-benzothiophene- and 5-benzofuran-modified uridine analogs were found to be moderately emissive.[84a,84b] Importantly, both the nucleoside analogs displayed excellent fluorescence solvatochromism and viscochromism, meaning their fluorescence properties like emission maximum, quantum yield, lifetime and anisotropy were highly sensitive to the surrounding solvent polarity and viscosity. Subsequently, 5-benzothiophene- and 5-benzofuran-modified 2′-deoxyuridine and uridine analogs, when placed in different nucleobase environment were highly sensitive to flanking bases and base pair substitutions. Based on the fluorescence outcome, these analogs were appropriately utilized in designing assays to detect abasic sites (depurinated site) in DNA and RNA ONs,[84b,84c] and in studying the ON dynamics in cell-like confined environment.[84d]
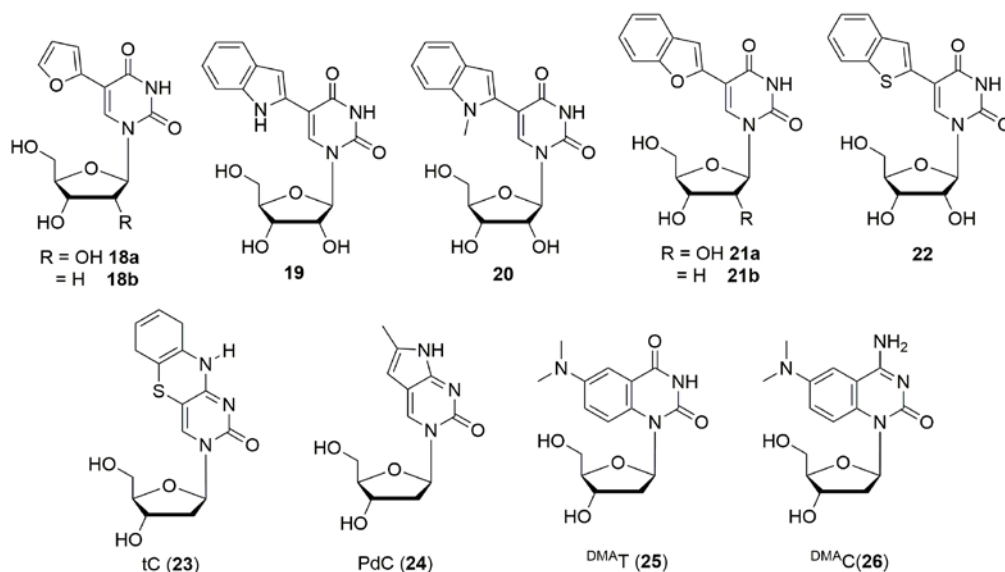
Figure 6. Chemical structure of furan (**18a** and **18b**), indole (**19**), *N*-methylindole (**20**), benzofuran (**21a** and **21b**) and benzothiophene (**22a** and **22b**) modified 2′-deoxyuridine and or uridine analogs. Chemical structure of tC (**23**), PdC (**24**), [DMA]T (**25**) and [DMA]C (**26**).

The non-perturbing nature and conformation-sensitivity of the heterobicycle-modified nucleoside analogs was put to use in addressing the challenges associated with probing GQ topologies and recognition. 5-benzofuran-modified 2′-deoxyuridine and uridine nucleosides were incorporated into loop residues of the H-Telo DNA and RNA repeat to study the

18

formation of respective GQ structures. (Figure 7).[85] The structurally non-invasive probe clearly distinguished hybrid-type GQ in KCl, antiparallel GQ in NaCl and duplex structures of H-Telo DNA by exhibiting changes in its quantum yield. On the other hand, in case of TERRA ON which adopts a parallel GQ topology irrespective of the ionic conditions the 5-benzofuran-modified uridine exhibited similar quatum yield and lifetime. Thus benzofuran-modified nucleosides photophysically reported the formation as well as distinguish different DNA and RNA GQ topologies. This feature was further succesfully utilized in determining the topology-specific and nucleic acid-specific binding of ligands like pyridostatin (PDS) and BRACO19 to telomeric DNA and RNA GQs. Recently in another study, the 5-benzofuan modified deoxyuridine analog when was incorporated into the iM forming c-rich sequence, it phophysically reported the p^H-dependent structural transition from random coil to iM by showing a significant reduction in fluorescence intensity and enabled the determination of transition pH for a specific sequence.[86] Along with CD and thermal melting studies, the distinct fluorescence properties of the probe in iM, GQ, duplex and random coil helped to understand the relative population of duplex and GQ/iM in a G-rich-C-rich double-stranded system under physiological condition. Juskowiak and co-worker developed pH sensors by incorporating a fluorescent cytidine analogue 1,3-diaza-2-oxophenothiazine (tC, **23**) into different cytidine positions of a proton-sensing aptamer, potential to form iM structure (Figure 6).[87] The pH sensing was based on gradual quenching of the fluorescence of tC due to protonation of the fluorophore (tC-H$^+$) in iM structure.
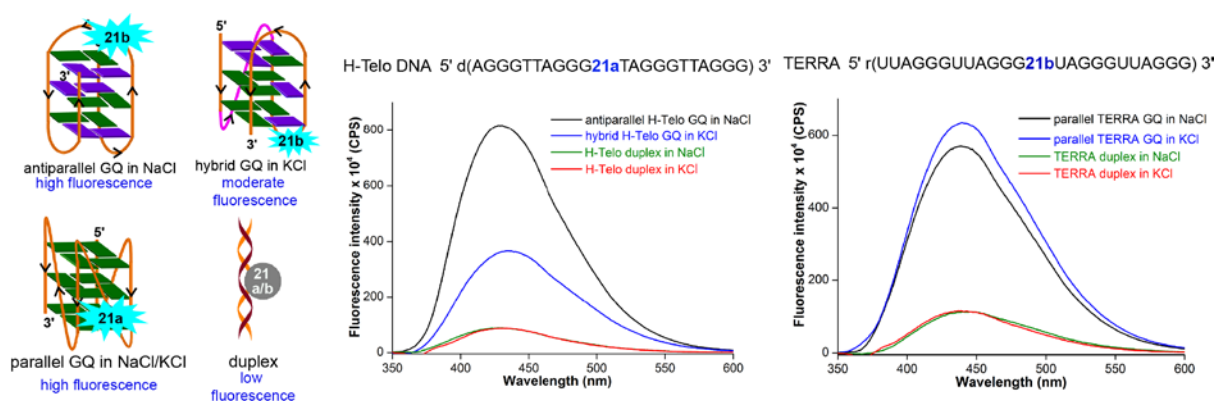
Figure 7 (**A**) 5-benzofuran-modified 2′-deoxyuridine (**21b**) incorporated into a H-Telo DNA ON fluorescently distinguishes different GQ structures. The antiparallel GQ in NaCl, mixed hybrid-type GQs in KCl and duplex show distinct fluorescence quantum yield and average lifetime. (**B**) Equivalent RNA repeat, TERRA, labeled with 5-benzofuran-modified uridine (**21a**) reports the formation of a parallel GQ structure, irrespective of the ionic conditions. The fluorescence intensity and average lifetime of TERRA GQ in NaCl and KCl are similar. The figure is adopted from reference 69c.

19

Recently fluorescent nucleosides have been attached site specifically in DNA oligonucleotides to photohysically report the non-canonical metallo base-pair formation in the respective DNA duplexes.[88,89] In a report from park group, a fluorescent cytidine analog pyrrolo-dC (PdC, **24**)-modified DNA duplex acted as Ag^I ion sensor by showing gradual quenching in fluorescence intensity with increasing the Ag^I concentration due to the formation of Pdc-Ag^I-C metallo basepair formation (Figure 6). This fluorescence reduction was attributed to the more efficient stacking and collisional quenching in the metallated state. Luetdke and co-worker have utilized thyamidine (^{DMA}T, **25**) and cytidine (^{DMA}C, **26**) fluorescent analogs to report the T-Hg-T and C-Hg-T formation.[89] They also studied the thermodynamic properties and kinetics of these base-pairs formation by monitoring the changes in the fluorescence intensity of the probes.

1.3.5.2.3 8-Substituted fluorescent purine nucleoside probes

8-substituted fluorescent guanosines incorporated into G-tetrads have provided efficient tools to study GQs.[90] Luedtke and coworkers developed a turn-on fluorescent GQ sensor, 8-(2-pyridyl)-2′-deoxyguanosine (2PyG, **27**), which after incorporation into the position where native guanine residue prefers to adopt *syn* glycosidic conformation did not affected the global GQ structure and stability.[91] 2PyG enabled the fluorescence detection of GQs and study the energy transfer in GQs. DNA ON labeled with 2PyG in different ionic conditions (Na^+, K^+, NH_4^+ and Rb^+) showed considerable enhancement in fluorescence intensity as well as higher energy transfer efficiency compared to random structure in Li^+ ionic conditions. (Figure 8A). Luedtke and Dumas introduced second generation GQ sensors 8-(2-phenylethenyl)-2′-deoxyguanosine (StG, **28**) and 8-[2-(pyrid-4-yl)-ethenyl]-2′-deoxyguanosine (4PVG, **29**) with extended conjugation (Figure 8B). High quantum yield and high molar extinction coefficient of ONs labeled with StG enabled the detection of GQ structures using conventional fluorescence spectrophotometer at concentrations as low as 0.25 nM.[92]
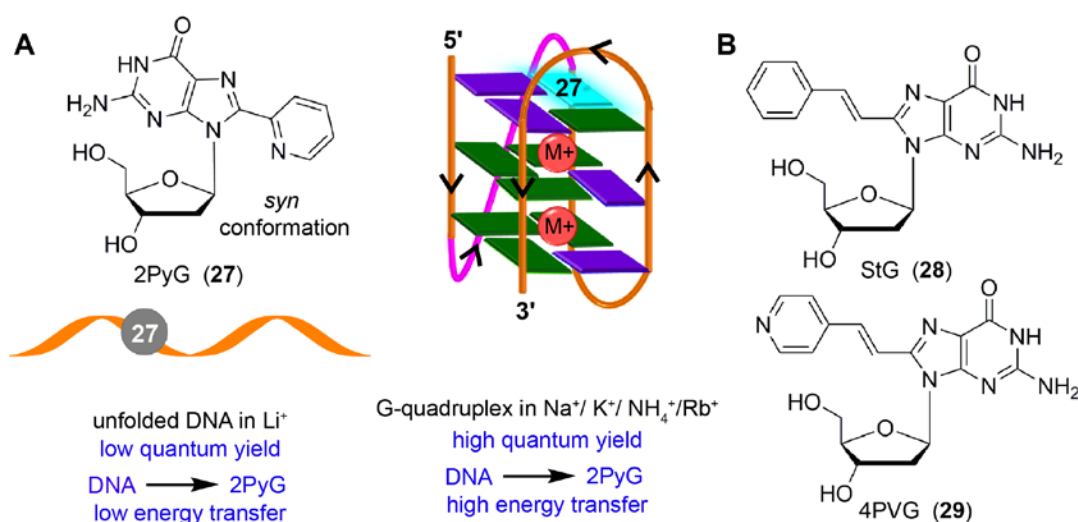
Figure. 8 (**A**) 2PyG in *syn* glycosidic conformation is shown. GQ-forming ONs containing 2PyG in the folded GQ state exhibits higher quantum yield and energy transfer efficiency compared to unfold state.[91] (**B**) Structure of 8-(2-phenylethenyl)-2′-deoxyguanosine (StG, **28**) and 8-[2-(pyrid-4-yl)-ethenyl]-2′-deoxyguanosine (4PVG, **29)** is shown.[92]

In another report by Diederichsen and coworker, a most conservatively modified fluorescent nucleoside analog, 8-vinyl-2′-deoxyguanosine (VdG, **30**) was incorporated into the human telomeric DNA ON repeat 5′-d[AGGG(TTAGGG)$_3$T] at different tetrad forming guanine positions.[93] Interestingly, telomeric ONs with VdG at guanine positions which participate in the formation of the middle tetrad were able to discriminate the GQ topology formed in NaCl (antiparallel) and in KCl (multiple conformations) with significant enhancement in fluorescence intensity (Figure. 9). However, these ONs show almost similar fluorescence intensity in duplex and quadruplex states in KCl with only a slight difference in the emission maximum.
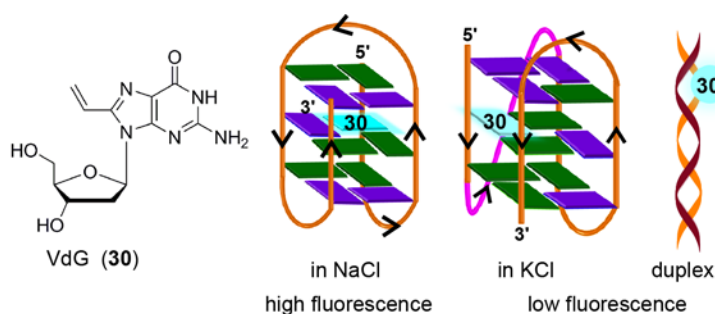
Figure 9. 8-vinyl-2′-deoxyguanosine VdG (**30**) adopts both *syn* and *anti* glycosidic conformations. When VdG is placed in the middle tetrad it shows higher fluorescence intensity for the antiparallel GQ in NaCl compared to the multiple GQs in KCl and duplex form.[93]

The group of Wetmore and Manderville utilized 8-furyl-dG (^{Fur}dG, **31**) and 8-(4′-cyanophenyl)-dG (^{CNPh}dG, **32**) as tools to detect the GQ folding of a 15-mer (5′-GGTTGGTGTGGTTGG) thrombin-binding aptamers.which forms a two tetrad chair-type antiparallel GQ structure (Figure 10A).[94] The guanosine analogs were placed at G_5 (*syn*), G_6 (*anti*) and G_8 (TGT) loop positions of the aptamer. The ^{Fur}dG placed at G_5 and ^{CNPh}dG placed at G_8 serve as a turn-on and turn-off fluorescence probe for duplex-quadruplex transition, respectively. Further, these probes were used in estimating the thrombin-mediated GQ folding of thrombin aptamer.[95] Upon increasing the thrombin concentration, ^{Fur}dG-modified aptamer displayed gradual increase in fluorescence intensity (Figure 10B) and the association constant (K_a) of thrombin binding to the aptamer was determined using fluorescence titration experiments. In another report, Manderville and coworker used the responsiveness of ^{Fur}dU (**18b**)[82a] to solvent polarity and viscosity changes to distinguishing thymine residues that are solvent-exposed in a GQ structure of a thrombin aptamer and identify the thymine residues that bind specifically to thrombin protein (Figure. 6).[96]

Kim group used pyrene-labeled adenosine nucleoside ^{Py}A to report iM structure formation and B DNA-Z DNA transition (Figure 10C).[97] 5' ^{Py}A labeled human telomeric C-rich sequence showed quenching in fluorescence intensity during the structural transition from random coil to iM.[97a] This fluorescence quenching in iM was attributed to the stacking interaction between planar C.CH$^+$ basepair and nonpolar pyrene ring. In an another report a C-rich sequence doubly labeled with ^{Py}A at two different loop positions reported the iM formation by showing quenched red-shifted emission due to exciplex formation by two pyrene residues. The ^{Py}A also has been used to monitor B-DNA to Z-DNA transition in high salt concentration.[97b] This study was based on enhanced stacking of fluorophore in B-DNA state, which led to fluorescence quenching and unstacking in Z-DNA resulting in fluorescence enhancement.
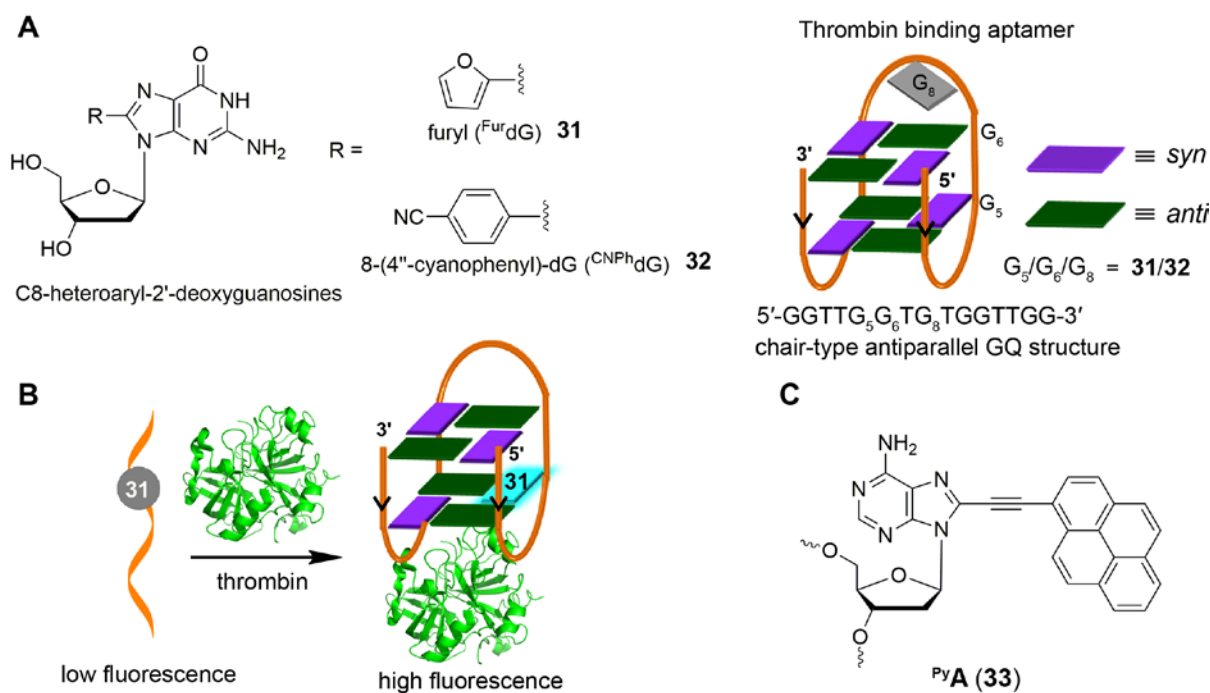
Figure 10. (**A**) C^8-heteroaryl-2′-deoxyguanosine fluorescent analogs ^{Fur}dG (**31**) and ^{CNPh}dG (**32**).[94] Site of incorporation of ^{Fur}dG (**31**)/^{CNPh}dG (**32**) into thrombin binding aptamer GQ at position G_5 (*syn*)/ G_6 (anti)/ G_8 (loop reside) is shown. (**B**) ^{Fur}dG- (**31**) modified thrombin aptamer shows higher fluorescence upon thrombin-mediated GQ formation compared to the aptamer in the unfolded state.[95] (**C**) Chemical structure of pyrene-modified adenosine nucleoside ^{Py}A (**33**).[97]

1.4 Structure-specific tools to probe nucleic acid structures in cells

1.4.1 Antibody-based tools

Although biochemical and biophysical *in vitro* analysis have long confirmed the formation of various canonical and noncanonical nucleic structures, probing the occurrence of those structures and finding their structure-function relationship *in vivo* has always remained a major challenge in nucleic acid field.[16] In this direction structure-specific antibody which enables the detection of a particular nucleic acid structure in cell has been observed to be a highly potential method. For example GQ structure-specific antibody has proven the existence of DNA and RNA GQs in cell. Plückthun and coworkers, for the first time selected high-affinity single-chain antibody fragments (scFvs) for the GQs formed by the *Stylonychia lemnae* telomeric repeat d(T$_4$G$_4$)$_2$ by ribosome display.[98] A selected fragment, Sty49 produced a clear signal in the macronuclei in immunofluorescence staining experiments, indicating the occurrence of GQ structure. Further, this antibody was used in studying the accumulation of GQs in the replication band and the unfolding of GQs by telomerase.[99]

23

Balasubramanian and coworkers developed an elegant method to pull-down GQ-forming sequences from mammalian cells by using a derivative of GQ-ligand PDS **34** having an affinity tag, biotin (Figure 12).[100] Pulling down telomeric repeats of HT1080 human cancer cells using **34** followed by high throughput sequencing (ChIP-Seq) provided the evidence for the formation of GQs in human cells. Based on this approach, a PDS derivative **35** containing a clickable alkyne group was developed. Cells treated with **35** for 12 h were click-stained with an Alexa-azide, which showed a fluorescent pattern resembling the staining of nucleoli containing GQ-forming sequences.[101]

Although these small-molecule GQ binders provide indirect evidence for the presence of GQs in cells, Balasubramanian's group developed a GQ-specific antibody to directly visualize DNA GQs in human cells.[7a] By employing phage display composed of 2.3×10^{10} different single-chain antibody clones, antibody called BG4 with nanomolar affinity for intramolecular and intermolecular GQ structures was indentified. Cells treated with BG4 were visualized by using a secondary antibody and then a tertiary fluorochrome-labeled antibody. Punctate nuclear staining in all cell lines tested along with appropriate control experiments confirmed the presence of DNA GQs in human cells. Further, it was identified that BG4 also binds to RNA GQs of TERRA, *BCL2* and *NRAS* with affinities in the nanomolar range. Encouraged by this observation, the presence of endogenous RNA GQs was confirmed in the cytoplasm of human normal endothelial, immortalized and cancer cell lines by treating the cells with BG4 followed by immunostaining with secondary and tertiary antibody.[7b] Interestingly, a derivative of PDS, carboxyPDS **36**, which binds strongly to RNA GQs as compared to DNA GQs, increased the number of GQs in the cytoplasm. These results suggest that structure-specific small molecule ligands can be used to target as well as trap GQs in cells.
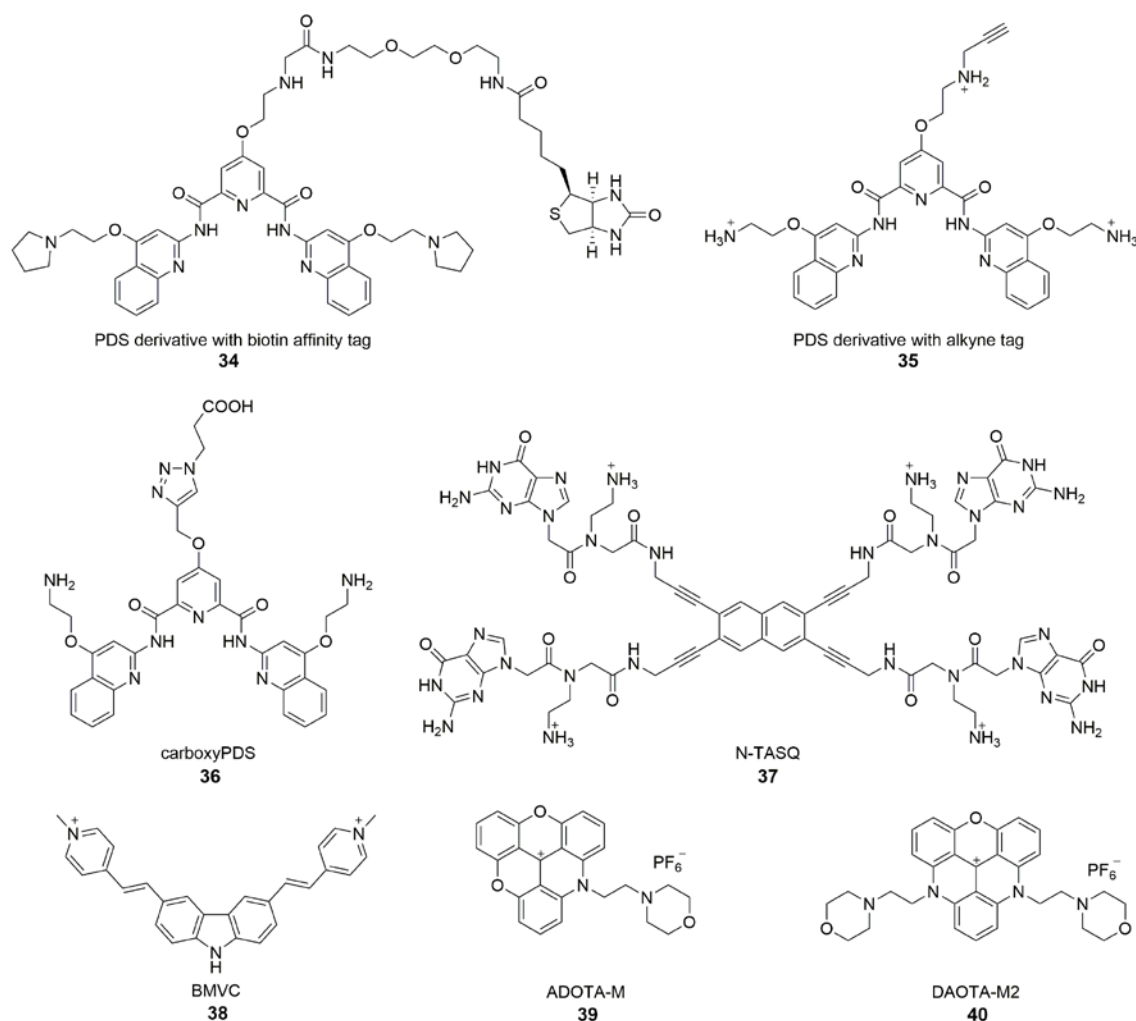
Figure 11. GQ-structure-specific small molecule ligands used in the visualization of DNA and RNA GQs in cells.

Lansdorp and coworkers identified a new murine monoclonal antibody specific for GQ structures by immunizing spleen cells from mice with stable DNA GQs.[7c] The antibody designated as 1H6 displayed a strong nuclear signal from human and murine cells in immunofluorescence microscopy studies. Addition of a GQ stabilizing ligand (TMPyP4) increased the number of nuclear foci. While BG4 and 1H6 are useful in GQ visualization experiments, they fail to provide information on the topology as these antibodies don't have preference for a certain GQ topology. In this context, Huang and coworkers recently developed a new scFv antibody, D1, which exhibits very high binding affinity and selectivity for the parallel GQ structure as opposed to other GQ forms.[102] Genome-wide identification of consensus sequences for parallel GQ structure was performed by chromatin immunoprecipitation using D1, followed by deep sequencing. The antibody enabled the visualization of parallel GQ structure at the human telomere in fixed cells. Importantly, using

ligands like NMM and QPB-15e, which are known to induce parallel GQ conformation, the formation of parallel GQ structure in cells could be modulated. Further, immunofluorescence experiments with fixed cells incubated with both BG4 and D1 antibodies indicated that the human telomeric GQs predominantly (77%) exist in the parallel form.

In an important contribution, Christ, Dinger and Bryan groups reported the direct visualization of iM structures in the nuclei of human cells by using an iM structure-specific antibody.[7d] This finding is important because the existence of iM structure *in vivo* was highly debated till date. Utilizing Garvan-2 human single-chain variable fragment (scFv) library and performing three round of phase selection they isolated the antibody iMab, which specifically binds to iM structures with picomolar to nanomolar affinity. Fluorescence immunostaining using iMab showed punctuate dots in human cells-nuclei, which confirmed the formation of iM in this region. Their studies also revealed that iM structures are formed in regulatory regions like in telomere and promoters of human genome and formation of these structures are pH and cell-cycle dependent. However, immunofluorescence detection of GQs and iMs using currently available antibodies has its own limitations, which include elaborate assay setup and restricted application to fixed and permeabilized cells.[103]

1.4.2 Small molecule ligand-based tools

Alternatively, structure-specific light-up and fluorogenic small molecule probes have been developed to visualize non-canonical nucleic structures particularly GQ in live cells. In one of the first efforts, Komiyama and coworkers used a light-switching pyrene probe, which forms excimers, to study the structure of TERRA RNA in living cells.[104] ONs containing TERRA RNA sequence were labeled with a pyrene moiety at both 5′ and 3′ ends showed fluorescence emission at ~400 nm from pyrene monomer in the unfolded state and showed a red-shifted emission at ~480 nm from an excimer formed by two pyrenes of close proximity in GQ state. This change in emission color was used to study the formation of GQ in cells. Treatment of HeLa cells with a dual-pyrene labeled ON containing TERRA RNA repeat in the presence of a transfecting agent and further imaging using fluorescence microscopy clearly revealed the excimer emission from pyrene due to the GQ folded state of the ON. Further, colocalization experiments indicated that the TERRA RNA GQ is restricted to nuclei. These observations indicate that RNA GQ can be potentially stabilized *in vivo*.

Monchaud and coworkers developed a GQ ligand called NaphthoTASQ (N-TASQ), which serves as a multi-photon fluorescence probe to directly visualize RNA GQs in live

cells (Figure 11).[103] N-TASQ binding to TERRA GQ resulted in 22-fold improvement in fluorescence of the probe. The probe also showed preference for RNA GQs as compared to DNA GQs. In cell based experiments, N-TASQ stained cytoplasmic and nucleolar compartments. Further, experiments with nucleases and BG4 antibody confirmed the GQ staining ability of the multi-photon probe. In another report, the GQ specific fluorogenic N-TASQ probe was observed to exhibit red-edge effect, which was utilized in visualizing DNA and RNA in human cell.[105] The unique red-edge fluorescence properties of N-TASQ-GQ complex enabled it to absorb light upto near infrared region and the emission maximum was observed to be linearly dependent on the excitation wavelength. These properties made N-TASQ-GQ complex compatible to be imaged by both confocal and two-photon microscopy. Further, the quadruplex specific red edge probe (G4-REP) N-TASQ was utilized to quantify BRACO-19 and TmPyP4 induced GQ in HeLa cells.[106] Additionally, the confocal imaging of N-TASQ treated live cancerous and non-cancerous cells displayed different staining pattern in cytoplasm, which proved that N-TASQ can enter both type of cells but the cellular response to N-TASQ is different in cancerous cell than non-cancerous cell. Most interestingly, a microplate-compatible LED based microscope housed inside a cell incubator was utilized in real-time analysis of N-TASQ treated live MDA-MB 231 breast cancer cells for longer periods (~120 h).

Chang and coworkers employed a fluorogenic GQ-binding ligand BMVC (3,6-bis(1-methyl-4-vinylpyridinium) carbazole diiodide) to investigate the cellular uptake and localization of various exogenously added GQ-forming DNA sequences (Figure 11).[107] The G-rich sequences, which form a parallel conformation, were found to accumulate in the lysosome, and the G-rich sequences, which form non-parallel GQ structures, were found to accumulate in the mitochondria of CL1-0 lung cancer cells. Further, o-BMVC and derivatives of BMVC, which have longer fluorescence decay time when bound to GQs than when bound to duplex structure, enabled the development of a method to visualize GQs in live cells by using fluorescence lifetime imaging microscopy (FLIM).[108] In a similar approach, Vilar and coworkers developed small molecule optical probes, which exhibit fluorescence enhancement and significant difference in lifetime depending on the nucleic acid topology.[109] They developed new triangulenium derivatives, ADOTA-M and DAOTA-M2 (Figure 11). In particular, DAOTA-M2 upon binding to GQs served as a turn-on fluorescence probe, and importantly, exhibited distinctly different lifetimes for GQ, duplex and single-stranded

nucleic acids. This feature was further utilized in visualizing GQs and also in studying the interaction of small molecule ligands with GQs in cells by using FLIM.

Richter and Freccero group developed a red-NIR core-extended NDI probe, which showed quenching in fluorescence intensity in water due to aggregation compared to monomer emission in organic solvent.[110] Interestingly, the probe was observed to bind GQ structure selectively compared to duplex and single stranded DNA. The presence of H-Telo DNA made the probe monomeric, and enhanced the fluorescent intensity significantly. Further, the location of NDI probe in the nucleus was visualized by confocal microscopy, and good colocalization of the NDI foci and GQ-specific antibody 1H6 proved that the probe indeed targets GQs in cells. Chow and coworker examined the efficiency of different pyridinium-based fluorescent dyes with varying symmetry and different styrene like side groups in detecting GQs.[111] Pyridinium molecule with C2 symmetry and indolyl side group, which was the best GQ-specific fluorescent probe was used in live cell imaging. The highest enhancement of fluorescence intensity in S phase of the cell cycle supported the replication-dependent GQ formation in cell.

In another study, Tan and coworkers developed hybridization ON probe, composed of a fluorogenic reporter ISCH-1, to identity GQs in cells.[112] ISCH-1, which shows remarkable enhancement in fluorescence efficiency, was click-tagged to an ON sequence complementary to the sequence adjacent to the GQ-forming motif of NRAS UTR. Since the target mRNA concentration in a single cell is less, the NRAS RNA ON was transfected into the cells followed by annealing with the hybridization probe. Under these conditions the cells displayed intense fluorescence spots in the cytoplasm. However, mutated version of the NRAS did not produce any signal. This method confirms that an exogenous RNA ON of NRAS can adopt GQ structure in cellular environment.

1.4.3 In-cell NMR and in-cell EPR

In-cell NMR and EPR spectroscopy have emerged as potential techniques to study nucleic acids in their native environment. Owing to the very low in-cell concentration of naturally occurring nucleic acids and very large size of genomic DNA, these methods mainly depends on introduction of unlabeled or labeled nucleic acids in cell followed by spectroscopic measurements.[16] Trantírek and Dötsch, in collaboration, established an in-cell NMR spectroscopy method to study different nucleic acid structural motifs like GQs, DNA and RNA hairpin in *Xenopus laevis* (frog) oocytes.[113] They microinjected ^{13}C and ^{15}N labeled

DNA and RNA hairpin into oocyte and performed 1D and 2D in-cell NMR analysis. It was observed that there was appearance of some new peaks over time which was because of degradation of the ONs or changes in their conformations by interactions with endogenous macromolecules. However chemical modifications like substitution of the phosphate group in the backbone by a thiophosphate moiety or by methylation of the O2'-hydroxy group enhanced the stability of ON in cell. In-cell NMR analysis was also performed with unlabeled GQ forming H-Telo DNA repeat $d(AG_3(TTAGGG)_3$ by monitoring the imino pron region of 1D ^1H NMR spectra. While degradation of the nucleic acid inside living cell is a concern for in-cell nucleic acid study, GQ structures were found to be quite stable in cell for the longer period. The imino proton signatures from a H-Telo DNA ON sequence $d(AG_3(TTAGGG)_3$ in egg extract and live oocytes suggested that the parallel conformation adopted in the presence of synthetic crowding agents (e.g., PEG) is not the preferred GQ conformation in cellular environment.[114] Mergny and coworkers, reported a high resolution tetramolecular GQ structure of a short isotope-labeled ON $d(TG_4T)_4$ in live oocytes by using $^1H–^{15}N$ SOFAST-HMQC analysis.[115] Further, NMR analysis of exogenous iM forming sequences from the promoter region of the human genome in live HeLa cells revealed that certain sequences support iM formation under physiological conditions.[116] However, it is worth mentioning here that the resolution of imino proton signals is poor in live cells and egg extract as compared to in cell lystae due to high viscosity and inherent inhomogeneity of cellular system.[16] Recently, Xu and co-worker used simple ^{19}F NMR method to study TERRA RNA GQ structure *in vitro* and in live cell.[117] They used 3,5-bis(trifluoromethyl)benzene labeled TERRA RNA which displayed distinct ^{19}F signal for dimeric GQ and two-subunits stacked higher-order RNA GQ structure *in vitro*. Interestingly, further in-cell ^{19}F NMR analysis of this labeled TERRA RNA demonstrated that TERRA RNA preferably adopts the higher-order structure in the crowded environment of the live cell. They have also utilized 3,5-bis(trifluoromethyl)benzene modified guanosine analog as ^{19}F NMR probe which after incorporation into human telomeric DNA, reported intermolecular GQ formation *in vitro* and in live cell.[61b]

In-cell EPR analysis is considered a potential technique to study the effect of cellular environment on nucleic acid structure because of some advantages like high sensitivity which leads to the requirement of lower intracellular nucleic acid concentration, absence of any background signal from cellular compartment and ability to detect long-range dipolar coupling between the spin-labels. Samples for in-cell EPR analysis have been prepared

similarly like in-cell NMR by microinjecting the spin-labeled ON into *Xenopus laevis* (frog) oocytes. *In vito* and in-cell PELDOR study were performed with duplex DNA, RNA hairpin and neomycin-sensing riboswitch to examine whether these structures are maintained or altered in complex cellular environment.[17] Recently *in vivo* EPR studies with GQ forming H-Telo DNA have suggested that this sequence exists as a mixture of parallel and 3-tetrad antiparallel basket conformations in cellular environment.[118] However, the major disadvantage of EPR technique is the fast reduction of nitroxide radical in cellular environment at physiological temperature.[17] Because of this, in-cell EPR experiments are performed with shock-frozen cell at cryogenic temperature where the morphological integrity of a native cell is compromised. Another major challenge in both of these techniques is the high cellular concentration of NAs of interest (20-250 µM).[16,17] This concentration range is quite higher than the natural concentration of NAs in cell (~ 1 µM). Hence major efforts are going on to enhance the sensitivity of these techniques so that nucleic acids can be investigated at physiological concentrations and conditions.

1.5 Challenges in present tools and scope of the thesis

It is quite clear from the above studies that appropriately labeled nucleoside probes are quite useful in providing information of various nucleic acid structures, dynamics and interactions with small molecules, metal ions and proteins. However, our understanding of what structure a nucleic acid sequence adopts inside cells, which structure is responsible for its function and how to target a specific nucleic acid topology using small molecule ligands is limited. This is because of the paucity of chemical probes that can efficiently differentiate different nucleic acid topologies, and majority of biophysical tools cannot be implemented in both cell-free and cellular systems. Therefore, development of biophysical tools, which will enable the probing of nucleic acid structure and recognition properties in both cell-free and cellular environments, will be highly beneficial in analyzing nucleic acids and also in setting up discovery platforms for identifying structure-specific binders.

My thesis is mainly focused on the development of biophysical platforms to study nucleic acid structure and recognition in cell-free and cellular environments by using conformation-sensitive nucleoside probes. The nucleoside probes have been developed by attaching heterobicycles like benzofuran, 5-fluoro-benzofuran and 5-methoxybenzofuran at the 5 position of 2'-deoxyurindine and uridine. The responsiveness of the nucleoside probes to their local environment has been appropriately implemented in assays to investigate

different nucleic acid secondary structures mainly non-canonical structures and their recognition properties (Figure 12).

(i) The microenvironment sensitivity of fluorescent benzofuran-modified nucleoside analogs was employed in investigating the GQ structure and ligand binding of the human telomeric DNA and RNA repeats in aqueous buffer and cell-like confined environment of reverse micelles (RM).

(ii) 5-fluorobenzofuran-modified nucleoside analogs composed of a microenvironment-sensitive fluorophore and a ^{19}F NMR label, serve as dual-app probes. In particular, distinct signatures displayed by ^{19}F-labeled nucleoside for different GQs enabled a systematic study in *Xenopus laevis* oocytes to provide new structural insights into the GQ topologies adopted by the human telomeric overhang in cells, which so far has remained unclear.

(iii) The scope of the dual-app probe was further expanded in studying the conformational change occurring at a functionally important domain in the internal ribosomal entry site of hepatitis C-virus RNA genome.

(iv) The fluorescence properties of 5-methoxybenzofuran-modified uridine analog were found to be highly sensitive to subtle changes in the polarity of the surrounding environment. This probe photophysically reported pyrimidine-pyrimidine mismatch and metal-mediated base pairing in RNA-DNA and RNA-RNA duplexes.

Taken together, design, synthesis and usefulness of our environment-sensitive nucleoside probes will provide new means to effectively study nucleic acids and also help in devising new screening platforms, which could have profound impact on nucleic acid-targeted drug discovery.
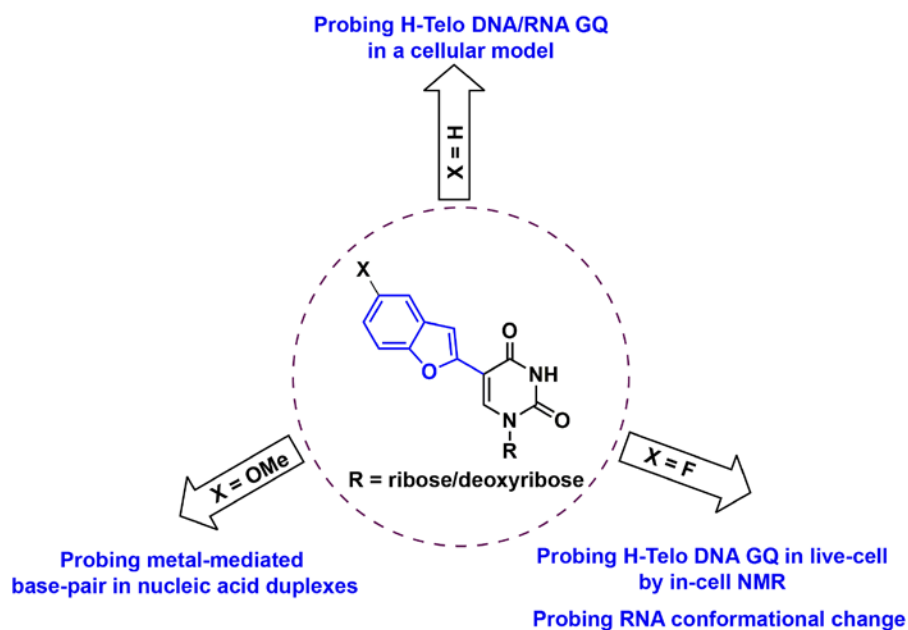
Figure 12. Schematic representation of the development of responsive nucleoside probes and their application in studying nucleic acid structures and recognition

1.6 References

1. (a) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. *Nat. Rev. Mol. Cell Biol*. **2004**, *5*, 1013–1023. (b) Saini, N.; Zhang, Y.; Usdin, K.; Lobachev, K. S. *Biochimie* **2013**, *95*, 117–123. (c) Brooks, T. A.; Kendrick, S.; Hurley, L. *FEBS J*. **2010**, *277*, 3459–3469.

2. Belmont, P.; Constant, J.-F.; Demeunynck, M. *Chem. Soc. Rev.* **2001**, *30*, 70–81.

3. (a) Hermann, T.; Patel, D. *Structure* **2000**, *8*, R47–R54. (b) Hall, K. B. *Curr. Opin. Chem. Biol.* **2008**, *12*, 612–618. (c) Al-Hashimi, H. M.; Walter, N. G. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321–329. (d) Choi, J.; Majima, T. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909.

4. (a) Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867–5892. (b) Cammas, A.; Millevoi, S. *Nucleic Acids Res.* **2017**, *45*, 1653–1668. (c) Theisen, A.; Shaffer, L. G. *Appl Clin Genet.* **2010**, *3*, 159–174.

5. (a) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916. (b) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901–2907. (c) Mir, B.; Serrano, I.; Buitrago, D.; Orozco, M.; Escaja, N.; Gonzalez, C. *J. Am. Chem. Soc.* **2017**, *139*, 13985–13988.

6. (a) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627–8637. (b) Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 279–284. (c) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res*. **2018**, *46*, 8038–8056.

7. (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186. (b) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. *Nat. Chem*. **2014**, *6*, 75–80. (c) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh Jr., R. M.; Sen, D.; Lansdorp, P. M. *Nucleic Acids Res.* **2013**, *42*, 860–869. (d) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. *Nat. Chem.* **2018**, *10*, 631–637.

8.   (a) Fletcher, T. M. *IUBMB Life* **2003**, *55*, 443–449. (b) Cuesta, J.; Read, M.; Neidle, S. *Mini-Rev. Med. Chem.* **2003**, *3*, 11–21. (c) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat.Rev. Drug Discov.* **2011**, *10*, 261–275. (d) Niu, K.; Zhang, X.; Deng, H.; Wu, F.; Ren, Y.; Xiang, H.; Zheng, S.; Liu, L.; Huang, L.; Zeng, B.; Li, S.; Xia, Q.; Song, Q.; Palli, S. R.; Feng, Q. *Nucleic Acids Res.* **2018**, *46*, 1711–1723. (e) Miglietta, G.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. *Sci Rep.* **2015**, *5*, 18097.

9.   (a) Rhee, S.; Han, Z.-J.; Liu, K.; Miles, H. T.; Davies, D. R. **1999**, *38*, 16810–16815. (b) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680–686.

10.   (a) Bissler, J. J. *Front Biosci.* **2007**, *12*, 4536–4546. (b) Vongsutilersa, V.; Gannett, P. M. *Org. Biomol. Chem.* **2018**, *16*, 2198–2209.

11.   Takezawa, Y.; Shionoya, M. *Acc. Chem. Res.* **2012**, *45*, 2066–2076.

12.   Scharf, P; Müller, J. *ChemPlusChem* **2013**, *78*, 20–34.

13.   (a) Chen, Y.; Yang, D. *Curr. Protoc. Nucleic Acid Chem.*, **2012**, *50*, 17.5.1−17.5.17. (b) Zhang, S.; Wu, Y.; Zhang, W. *ChemMedChem*, **2014**, *9*, 899–911. (c) Miyoshi, D.; Fujimoto, T.; Sugimoto, N. *Top. Curr. Chem.* **2013**, *330*, 87–110. (d) Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833. (e) Rajendran, A.; Nakano, S.; Sugimoto, N. *Chem.Commun.* **2010**, *46*, 1299–1301.

14.   (a) Shi, X.; Herschlag, D.; *Methods Enzymol.* **2009**, *469*, 287–302. (b) Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. *ChemBioChem* **2013**, *14*, 540–558. (c) Juskowiak, B. *Anal. Bioanal. Chem.* **2011**, *399*, 3157–3176. (d) Zhang, X.; Cekan, P.; Sigurdsson, S. T.; Qin, P. *Methods Enzymol.* **2009**, *469*, 303–328. (e) Piton, N.; Mu, Y.; Stock, G.; Prisner, T. F.; Schiemann, O.; Engels, J. W. *Nucleic Acids Res.* **2007**, *35*, 3128–3143. (f) Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *Trends Biochem. Sci.* **2003**, *28*, 259–266. (g) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struc. Biol.* **2004**, *14*, 350–359.

15.   Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.

16.   Giassa, I.-C.; Rynes, J.; Fessl, T.; Foldynova-Trantirkova, S.; Trantirek, L. *FEBS Letters* **2018**, *592*, 1997–2011.

17.   Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantirek, L.; Dötsch, V. *Angew. Chem. Int. Ed.* **2014**, *53*, 10300–10314.

18.   Burge, S.; G. Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.

19.   (a) Wang, Y.; Patel, D. J. *Structure*, **1993**, *1*, 263–282. (b) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.

20.   (a) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. (b) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927–4940.

21.   Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature*, **2002**, *417*, 876–880.

22.   (a) Chen, Z.; Zheng, K.-W.; Hao, Y.-H.; Tan, Z. *J. Am. Chem. Soc.* **2009**, *131*, 10430–10438. (b) Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. *J. Am. Chem. Soc.* **2010**, *132*, 17105–17107. (c) Lannan, F. M.; Mamajanov, I.; Hud, N. V. *J. Am. Chem. Soc.* **2012**, *134*, 15324–15330.

23.   Shrestha, P.; Jonchhe, S.; Emura, T.; Hidaka, K.; Endo, M.; Sugiyama, H.; Mao, H. *Nat. Nanotechnol.* **2017**, *12*, 582–588.

24.   (a) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry*, **2005**, *44*, 2048–2058. (b) Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. *Nucleic Acids Res.* **2011**, *39*, 9023–9033.

25.   (a) Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D. *Nucleic Acids Res.* **2008**, *36*, 4598–4608. (b) De Armond, R.; Wood, S.; Sun, D. Y.; Hurley, L. H.; Ebbinghaus, S. W.

Biochemistry, **2005**, *44*, 16341–16350. (c) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry*, **2006**, *45*,7854–7860. (d) Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **200**7, *129*, 10220–10228. (e) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. *Nucleic Acids Res.* **2007**, *35*, 7698–7713.

26. Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 5133–5144.

27. Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S. S.; Korsmeyer, J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. *Nature* **2005**, *435*, 677–681.

28. Young, R. L.; Korsmeyer, S. J. *Mol. Cell. Biol.* **1993**, *13*, 3686–3697.

29. Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. *J Am. Chem. Soc.* **2014**, *136*, 1750–1753.

30. Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palù, G.; Richter, S. N. *J. Med. Chem.* **2013**, *56*, 6521–6530.

31. Nicola, B. D.; Lech, C. J.; Heddi, B.; Regmi, S.; Frasson, I.; Perrone, R.; Richter, S. N.; Phan, A. T. *Nucleic Acids Res.* **2016**, *44*, 6442–6451.

32. Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M.-L.; Mergny, J.-L. *J. Am. Chem. Soc.* **2014**, *136*, 5249–5252.

33. (a) Agarwal, P.; Pandey, S.; Maiti, S. *Org. Biomol. Chem.* **2015**, *13*, 5570–5585. (b) Arora, A.; Maiti, S. *J. Phys. Chem. B* **2009**, *113*, 10515–10520.

34. (a) Xu, Y.; Kaminaga, K.; Komiyama, M. *J. Am. Chem. Soc.* **2008**, *130*, 11179–1184. (b) Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570–2578.

35.   Bugaut, A.; Balasubramanian, S. *Nucleic Acids Res.* **2012**, *40*, 4727–4741.

36. Sacca, B.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res.* **2005**, *33*, 1182–1192.

37. (a) Guéron, M.; Leroy, J.-L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326–331. (b) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. *RSC Adv.* **2014**, *4*, 26956–26980. (c) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res.* **2017**, *45*, 2951–2959. (d) Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. *J. Am. Chem. Soc.* **2017**, *139*, 4682–4689.

38. (a) Lacroix, L.; Liénard, H.; Labourier, E.; Djavaheri-Mergny, M.; Lacoste, J.; Leffers, H.; Tazi, J.; Hélène, C.; Mergny, J.-L. *Nucleic Acids Res.* **2000**, *28*, 1564–1575. (b) Cornuel, J.-F.; Moraillon, A.; Guéron, M. *Biochimie* **2002**, *84*, 279–289. (c) Uribe, D. J.; Guo, K.; Shin, Y.-J.; Sun, D. *Biochemistry* **2011**, *50*, 3796–3806. (d) Yoga, Y. M. K.; Traore, D. A. K.; Sidiqi, M.; Szeto, C.; Pendini, N. R.; Barker, A.; Leedman, P. J.; Wilce, J. A.; Wilce, M. C. J. *Nucleic Acids Res.* **2012**, *40*, 5101–5114.

39. Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4172–4185.

40. Chen, Y.; Qu, K.; Zhao, C.; Wu, L.; Ren, J.; Wang, J.; Qu, X. *Nat. Commun.* **2012**, *3*, 1074.

41. Cui, Y.; Koirala, D.; Kang, H.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H.; *Nucleic Acids Res.* **2014**, *42*, 5755–5764.

42. Day, H. A.; Pavlou, P; Waller, Z. A. E. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.

43. Jain, A.; Wang, G.; Vasquez, K. M. *Biochimie* **2008**, *90*, 1117–1130.

44. Rich, A.; Zhang, S. *Nat. ReV. Genet.* **2003**, *4*, 566–572.
45. (a) Phan, A. T.; Maurice Guéron, M.; Leroy, J.-L. *J. Mol. Biol.* **2000**, *299*, 123–144. (b) Tereshko, V. *et. al. Nucleic Acids Res.* **2001**, *29*, 1208–1215. (c) Tarköy, M.; Kathryn Phipps, A.; Schultze, P.; Feigon, J. *Biochemistry* **1998**, *37*, 5810–5819. (d) Kondo, J.; Yamada, T.; Hirose, C.; Okamoto, I.; Tanaka, Y.; Ono, A. *Angew. Chem. Int. Ed.* **2014**, *53*, 2385–2388.
46. Ono, A.; Torigoe, H.; Tanaka, Y.; Okamoto, I. *Chem. Soc. Rev.* **2011**, *40*, 5855–5866.
47. (a) Urata, H.; Yamaguchi, E.; Funai, T.; Matsumura, Y.; Wada, S.-i. *Angew. Chem. Int. Ed.* **2010**, *49*, 6516–6519. (b) Funai, T.; Nakamura, J.; Miyazaki, Y.; Kiriu, R.; Nakagawa, O.; Wada, S.-i., Ono, A.; Urata, H. *Angew. Chem. Int. Ed.* **2014**, *53*, 6624–6627. (c) Park, K. S.; Jung, C.; Park, H. G. *Angew. Chem. Int. Ed.* **2010**, *49*, 9757–9760.
48. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
49. Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. *Methods* **2012**, *57*, 64–75.
50. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. *FEBS Letters* **2002**, *526*, 77–81.
51. Yu, H.-Q. Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2006**, *128*, 15461–15468.
52. Doktycz, M. J. *Encyclopedia of Life Science. John Wiley & Sons, Chichester* **1997**, 3123–3140.
53. Brazier, J. A.; Shaha, A.; Brown, G. D. *Chem. Commun.* **2012**, *48*, 10739–10741.
54. (a) Rachwal, P. A.; Fox, K. R. *Methods* **2007**, *43*, 291–301. (b) Mergny, J.-L.; Lacroix, L. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37*, 17.1.1–17.1.15.
55. (a) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936 – 962. (b) Adrian, M.; Heddi, B.; Phan, A. T. *Methods*, **2012**, *57*, 11–24.
56. (a) Feigon, J.; Koshlap, K.; Smith, M. F. W. *Methods Enzymol.* **1995**, *261*, 225–255. (b) Phan, A. T.; Mergny, J.-L. *Nucleic Acids Res.* **2002**, *30*, 4618–4625.
57. (a) Esmaili, N.; Leroy, J. L. *Nucleic Acids Res.* **2005**, *33*, 213–224. (b) Jeong, M.; Lee, A.-R.; Kim, H.-E.; Choi, Y.-G.; Choi, B.-S.; Lee, J.-H. *Arch Biochem Biophys* **2014**, *558*, 95–103.
58. (a) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167–173. (b) Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680. (c) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. *Angew.Chem. Int. Ed.* **2014**, *53*, 999–1002.
59. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.
60. (a) Fauster, K.; Kreutz, C.; Micura, R. *Angew Chem Int Ed.* **2012**, *124*, 13257–13261.(b) Puffer, B.; Kreutz, C.; Rieder, U; Ebert, M.-O.; Konrat, R.; Micura, R. *Nucleic Acids Res.* **2009**, *37*, 7728–7740. (c) Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. *J. Org. Chem.* **2012**, *77*, 1026−1044. (d) Granqvist, L; Virta, P. *Chem. Eur. J.* **2016**, *22*, 15360–15372. (e) Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. A. *Angew. Chem. Int. Ed.* **2006**, *45*, 3450–3453. (f) Olszewska, A.; Pohl, R.; Hocek, M. *J. Org. Chem.* **2017**, *82*, 11431–11439.
61. (a) Ishizuka, T.; Yamashita, A.; Asada, Y.; Xu, Y. *ACS Omega* **2017**, *2*, 8843–8848. (b) Ishizuka, T.; Zhao, P.-Y.; Bao, H.-L.; Xu, Y. **2017** *Analyst 142*, 4083–4088.
62. Nakamura, S.; Yang, H.; Hirata, C.; Kersaudy, F.; Fujimoto, K. *Org. Biomol. Chem.* **2017**, *15*, 5109–5111.
63. (a) Grytz, C. M.; Marko, A.; Cekan, P.; Sigurdsson, S. T.; Prisner, T. F. *Phys. Chem. Chem. Phys.* **2016**, *18*, 2993–3002 (b) Reginsson, G. W.; Shelke, S. A.; Rouillon, C.; White, M. F.; Sigurdsson, S. T.; Schiemann, O. *Nucleic Acids Res.* **2013**, *41*, e11. (c)

Nguyen, P.; Qin, P. Z. *Wiley Interdisc. Rev. RNA* **2012**, *3*, 62–72. (d) Wunnicke, D.; Strohbach, D.; Weigand, J. E.; Appel, B.; Feresin, E.; Suess, B.; Müller, S.; Steinhoff, H.-J. *RNA* **2011**, *17*, 182–188.

64. Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M. *Angew. Chem. Int. Ed.* **2009**, *48*, 9728–9730.

65. Zhang, X.; Xu, C.-X.; Felice, R. D.; Sponer, J.; Islam, B.; Stadlbauer, P.; Ding, Y.; Mao, L.; Mao, Z.-W.; Qin, P. Z. *Biochemistry* **2016**, *55*, 360–372.

66. (a) Mooers, B. H. *Methods* **2009**, *47*, 168. (b) Neidle, S. *J. Med. Chem.* **2016**, *59*, 5987–6011.

67. (a) Salon, J.; Gan, J.; Abdur, R.; Liu, H.; Huang, Z. *Org. Lett.* **2013**, *15*, 3934–3937. (b) Pallan, P. S.; Egli, M. *Nature Protocols* **2007**, *2*, 647–651. (c) Nuthanakanti, A.; Boerneke, M. A.; Hermann, T.; Srivatsan, S. G. *Angew. Chem. Int. Ed.* **201**7, *56*, 2640–2644.

68. Neidle, S. *Nat. Rev. Chem.* **2017**, *1*, 0041.

69. (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579–2619. (b) Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem.* **2013**, *53*, 366–378. (c) Manna, S.; Srivatsan, S. G. *RSC Adv.* **2018**, *8*, 25673–25694.

70. (a) Mergny, J.-L.; Maurizot, J.-C. *ChemBioChem* **2001**, *2*, 124–132. (b) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. *Nucleic Acids Res.* **2010**, *38*, 7858–7868. (c) Ying, L.; Green, J. J.; Li, H.; Kleneram, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 14629–14634. (d) Choi, J.; Kim, S.; Tachikawa, T.; Fujitsuka, M.; Majima, T. *J. Am. Chem. Soc.* **2011**, *133*, 16146–16153. (e) Schneider, U. V.; Severinsen, J. K.; Geci, I.; Okkels, L. M.; Jøhnk, N.; Mikkelsen, N. D; Klinge, T.; Pedersen, E. B.; Westh, H.; Lisby, G. *BMC Biotechnol* **2010**, *10*, 4.

71. (a) He, F.; Tang, Y.; Yu, M.; Feng, F.; An, L.; Sun, H.; Wang, S.; Li, Y.; Zhu, D.; Bazan, G. C. *J. Am. Chem. Soc.* **2006**, *128*, 6764-6765. (b) Kumar, N.; Maiti, S. *Biochem. Biophys. Res. Commun.* **2004**, *319*, 759. (c) Lee, I.-B.; Hong, S.-C.; Lee, N.-K.; Johner, A. *Biophys. J.* **2012**, *103*, 2492−2501. (d) Dumat, B.; Larsen, F. A.; Wilhelmsson, L. M. *Nucleic Acids Res.* **2016**, *44*, e101.

72. (a) Rache, A. D.; Mergny, J.-L. *Biochimie* **2015**, *115*, 194–202. (b) Sheng, Q.; Neaverson, J. C.; Mahmoud, T.; Stevenson, C. E. M.; Matthews, S. E.; Waller, Z. E. *Org. Biomol. Chem.* **2017**, *15*, 5669–5673. (c) Maleki, P.; Ma, Y.; Iida, K.; Nagasawa, K.; Balci, H. *Nucleic Acids Res.* **2017**, *45*, 288–295. (d) Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4161–4171. (e) Wright, E. P.; Day, H. A.; Ibrahim, A. M.; Kumar, J.; Boswell, L. J.; Huguin, C.; Stevenson, E.; Pors, K.; Waller, Z. A. E. *Sci. Rep.* **2016**, *6*, 39456.

73. (a) Ohtsuka, K.; Sato, S.; Sato, Y.; Sota, K.; Ohzawa, S.; Matsuda, T.; Takemoto, K.; Takamune, N.; Juskowiak, B.; Nagai, T.; Takenaka, S. *Chem. Commun.* **2012**, *48*, 4740–4742. (b) Li, J. J.; Fang, X.; Tan, W. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 31–40. (c) De Tito, S.; Morvan, F.; Meyer, A.; Vasseur, J.-J.; Cummaro, A.; Petraccone, L.; Pagano, B.; Novellino, E.; Randazzo, A.; Giancola, C.; Montesarchio, D. *Bioconjug. Chem.* **2013**, *24*, 1917–1927. (d) Ono, A.; Togashi, H. *Angew. Chem. Int. Ed*. **2004**, *43*, 4300–4302. (e) Teh, H. B.; Wu, H.; Zuo, X. *Sensors and Actuators B* **2014**, *195*, 623–629.

74. (a) Largy, E.; Granzhan, A.; Hamon, F.; Verga, D.; Teulade-Fichou, M. P. *Top. Curr. Chem*. **2013**, *330*, 111–177. (b) Sabharwal, N. C.; Savikhin, V.; Turek-Herman, J. R.; Nicoludis, J. M.; Szalai, V. A.; Yatsunyk, L. A. *FEBS J.* **2014**, *281*, 1726–1737. (c) Lee, I. J.; Patil, S. P.; Fhayli, K.; Alsaiari, S.; Khashab, N. M. *Chem. Commun.* **2015**,

51, 3747–3749. (d) Xu, L.; Hong, S.; Sun, N.; Wang, K.; Zhou, L.; Ji, L.; Pei, R. *Chem. Commun.* **2016**, *52*, 179–182. (e) Chiang, C.-K.; Huang, C.-C.; Liu, C.-W.; Chang, H.-T. *Anal. Chem.* **2008**, *80*, 3716–3721. (f) Chan, D. S.-H.; Lee, H.-M.; Che, C.-M.; Leung, C.-H.; Ma, D.-L. *Chem. Commun.* **2009**, 7479–7481.

75. (a) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem. Commun.* **2004**, 1438–1439.(b) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem. Commun.* **2006**, 401–402.

76. Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649–4659.

77. Xu, Y.; Sugiyama, H.; *Nucleic Acids Res.* **2006**, *34*, 949–954.

78. Johnson, J.; Okyere, R.; Joseph, A.; Musier-Forsyth, K.; Kankia, B. *Nucleic Acids Res.* **2013**, *41*, 220–228.

79. Gros, J.; Rosu, F.; Amrane, S.; Cian, A. D.; Gabelica V.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res.* **2007**, *35*, 3064–3075.

80. Myers, J. C.; Moore, S. A.; Shamoo, Y. *J. Biol. Chem.* **2003**, *278*, 42300–42306.

81. Shchyolkina, A. K.; Kaluzhny, D. N.; Borisova, O. F.; Hawkins, M. E.; Jernigan, R. L.; Jovin, T. M.; Arndt-Jovin, D. J.; Zhurkin, V. B. *Nucleic Acids Res.* **2004**, *32*, 432–440.

82. (a) Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784–10785. (b) Srivatsan, S. G.; Tor, Y. *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053. (c) Srivatsan, S. G.; Tor, Y. *Tetrahedron* **2007**, *63*, 3601–3607.

83. Srivatsan, S. G.; Greco, N. J.; Tor, Y. *Angew. Chem. Int. Ed.* **2008**, *47*, 6661–6665.

84. (a) Pawar, M. G.; Srivatsan, S. G. *Org. Lett.* **2011**, *13*, 1114–1117. (b) Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J.* **2011**, *17*, 12820–12827. (c) Tanpure, A. A.; Srivatsan, S. G. *Chembiochem.* **2012**, *13*, 2392–9156. (d) Pawar, M. G.; Srivatsan S. G. *J. Phys. Chem. B* **2013**, *117*, 14273–14282.

85. Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149.

86. Sabale, P. M.; Tanpure, A. A.; Srivatsan, S. G. *Org. Biomol. Chem.* **2018**, *16*, 4141–4150.

87. Bielecka, P.; Juskowiak, B. *Molecules* **2015**, *20*, 18511–18525.

88. Park, K. S.; Lee, J. Y.; Park, H. G. *Chem. Commun.* **2012**, *48*, 4549–4551.

89. (a) Schmidt, O. P.; Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2016**, *138*, 14733–14739. (b) Schmidt, O. P.; Benz, A. S.; Mata, G; Luedtke, N. W. *Nucleic Acids Res.* **2018**, *46*, 6470–6479.

90. Manderville, R. A.; Wetmore, S. D. *Chem. Sci.* **2016**, *7*, 3482–3493.

91. Dumas, A.; Luedtke, N. W. *J. Am. Chem. Soc.* **2010**, *132*, 18004–18007.

92. Dumas, A.; Luedtke, N. W. *Nucleic Acid Res.* **2011**, *39*, 6825–6834.

93. Nadler, A.; Strohmeier, J.; Diederichsen, U. *Angew. Chem. Int. Ed.* **2011**, *50*, 5392–5396.

94. Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A.; Sharma, P.; Wetmore, S. D. *Chem. Sci.* **2014**, *5*, 788–796.

95. Sproviero, M.; Manderville, R. A. *Chem. Commun.* **2014**, *50*, 3097–3099.

96. Cservenyi, T. Z; Riesen, A. J. V.; Berger, F. D.; Desoky, A.; Manderville, R. A. *ACS Chem. Biol.* **2016**, *11*, 2576–2582.

97. (a) Lee, J.; Yi, J. W.; Kim, B. H. *Chem. Commun.* **2009**, 5383–5385. (b) Seo, Y. J.; Kim, B. H. *Chem. Commun.* **2006**, 150–152.

98. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 8572–8577.

99. Paeschke, K.; Juranek, S.; Simonsson, T.; Hempel, A.; Rhodes, D.; Lipps, H. J. *Nat Struct Mol Biol* **2008**, *15*, 598–604.

100. Müller, S.; Kumari, S.; Rodriguez, R.; Balasubramanian, S. *Nat. Chem.* **2010**, *2*, 1095–1098.

101. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* **2012**, *8*, 301–310.

102. Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ren, J.; Zhao, Y.; Huang, Z.-S. *Cell Chem. Biol.* **2016**, *23*, 1261–1270.

103. Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M. Y.; Monchaud, D. *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525.

104. Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 14579–14584.

105. Laguerre, A.; Wong, J. M. Y.; Monchaud, D. *Sci. Rep.* **2016**, *6*, 32141.

106. Yang, S. Y.; Amor, S.; Laguerre, A.; Wong, J. M. Y.; Monchaud, D. *Biochim. Biophys. Acta.* **2017**, *1861*, 1312–1320.

107. Tseng, T.-Y.; Wang, Z.-F.; Chien, C.-H.; Chang, T.-C. *Nucleic Acids Res.* **2013**, *41*, 10605–10618.

108. (a) Tseng, T.-Y.; Chien, C.-H.; Chu, J.-F.; Huang, W.-C.; Lin, M.-Y.; Chang, C.-C.; Chang, T.-C. *J. Biomed. Opt.* **2013**, *18*, 101309-1–101309-6. (b) Huang, W.-C.; Tseng, T.-Y.; Chen, Y.-T.; Chang, C.-C.; Wang, Z.-F.; Wang, C.-L.; Hsu, T.-N.; Li, P.-T.; Chen, C.-T.; Lin, J.-J.; Lou, P.-J.; Chang, T.-C. *Nucleic Acids Res.* **2015**, *43*, 10102–10113.

109. Shivalingam, A.; Izquierdo, M. A.; Marois, A. L.; Vyšniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. *Nat. Commun.* **2015**, *6*, 8178.

110. Doria, F.; Nadai, M.; Zuffo, M.; Perrone, R.; Freccero, M.; Richter, S. N. *Chem. Commun.* **2017**, *53*, 2268–2271.

111. Lu, Y.-J.; Hu, D.-P.; Zhang, K.; Wong, W.-L.; Chow, C.-F. *Biosens. Bioelectron.* **2016**, *81*, 373–381.

112. Chen, S.-B.; Hu, M.-H.; Liu, G.-C.; Wang, J.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Tan, J.-H. *J. Am. Chem. Soc.* **2016**, *138*, 10382–10385.

113. Hänsel, S. Foldynová-Trantírková, F. Löhr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Dötsch and L. Trantírek, *J. Am. Chem. Soc.* **2009**, *131*, 15761–15768.

114. (a) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. *Nucleic Acids Res.* **2011**, *39*, 5768–5775. (b) Hänsel, R.; Löhr, F.; Trantírek, L.; Dötsch, V. *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.

115. Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J,-L. *Chem. Sci.* **2015**, *6*, 3314–3320.

116. Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldynova-Trantirkova, S.; Trantírek, L. *Angew. Chem. Int. Ed.* **2018**, *57*, 2165–2169.

117. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. *Nucleic Acids Res.* **2017**, *45*, 5501–5511.

118. Azarkh, M.; Singh, V.; Okle, O.; Dietrich, D. R.; Hartig, J. S.; Drescher, M. *ChemPhysChem* **2012**, *13*, 1444–1447.

Chapter 2

Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model
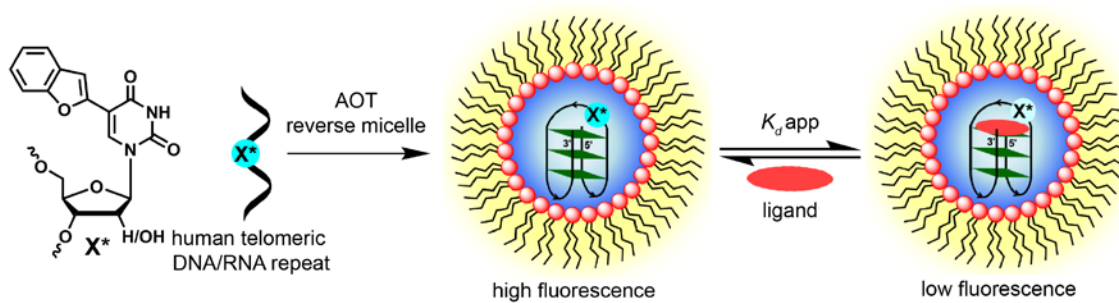
2.1 Introduction

Guanine-rich nucleic acid sequences that have the potential to form four-stranded noncanonical secondary structures, called G-quadruplexes (GQs), are frequently found in the genome of mammals, bacteria and viruses.[1] The position and conservation of putative GQ forming motifs, particularly among mammalian species, as evident from bioinformatics and sequencing studies suggest that GQ could be an important class of structural element for gene regulation.[2] In the human genome, GQ forming sequences are mostly found in the telomeric region (e.g., telomeric DNA and RNA repeats) and in several DNA promoters (e.g., *C-myc*, *C-kit*) and untranslated regions of mRNA (e.g., *NRAS, BCL-2*).[3,4] Recent biochemical investigations point out that these G-rich sequences play crucial roles in cellular processes like chromosome maintenance, transcriptional and translational regulation of several proto-oncogenes, which have been found to be in consensus with the ability of these sequences to form stable GQ structures *in vitro*.[5,6] Consequently, stabilization of GQ structures by small molecule ligands has emerged as a novel approach to cancer therapeutics.[7] In this context, several small molecule ligands that bind and stabilize GQs are being rigorously evaluated as chemotherapeutic candidates and also have been used as tools to understand the biological role of GQ forming sequences.[8,9]

In terms of structure, G-rich sequences form wide varieties of geometry *in vitro* depending on sequence and ionic environment.[3,9] These sequences typically form either one or a combination of antiparallel, parallel and mixed parallel-antiparallel stranded GQ structures. Several methods based on circular dichroism (CD), fluorescence, NMR and X-ray crystallography techniques have provided valuable information on the structure and ligand-binding affinities of GQs in cell-free systems.[10] Notably, FRET pair-containing ONs,[11] ligands,[12] metal complexes[10c] and fluorescent nucleobase analogs,[13] which exhibit changes in their fluorescence properties upon folding have been widely used to probe the formation and recognition property of GQs. More recently, GQ-specific ligands and antibodies have been used to directly visualize DNA and RNA GQs in mammalian cells.[14] Model systems, which closely mimic the physical properties (e.g., polarity and viscosity) and crowded environment of the cell have also been used to study the structure, stability and recognition properties of human telomeric (H-Telo) DNA and RNA repeats.[15] Typically in aqueous buffer, the H-Telo

DNA repeat (TTAGGG)$_n$ forms antiparallel GQ in Na$^+$ ionic conditions and a combination of GQ topologies in the presence of K$^+$ ions in which hybrid-type mixed parallel-antiparallel stranded GQ structures predominate.[16] However, crowding agents and high viscous cosolutes like polyethylene glycol (PEG), polysaccharides and deep eutectic solvents in the presence of K$^+$ ions have been shown to stabilize parallel GQ structure which exhibits slower folding dynamics, reduced ligand-binding affinity and decreased stabilization by ligands compare to those in dilute solution.[17-20] It has been suggested that the dehydrating nature of crowding agents (e.g., PEG) drives the formation of parallel GQ in the presence of K$^+$ ions.[21] Hence, the use of PEG as a molecular crowding agent in the study of GQs may not be appropriate.[22] Recent NMR analyses of H-Telo DNA repeat in live *Xenopus laevis* oocytes and oocytes egg extract support the notion that cellular environment favors conformations that closely resemble the ones observed *in vitro* under K$^+$ ionic conditions as opposed to the parallel topology predicted in the presence of synthetic crowding agents.[23] However, equivalent telomeric repeat containing RNA, TERRA (UUAGGG)$_n$, folds into parallel GQ structure irrespective of ionic conditions,[24] and the topology does not change in solution containing 40% PEG.[15b]

Despite extensive studies, probing different GQ topologies and their ligand-binding abilities in a cellular environment has remained a major challenge.[25-27] For example, structural analysis of GQs in cell requires elaborate assay setups and expensive isotope-labeled ONs in non-apoptotic concentrations, which often leads to obscure signal due to inhomogeneity of cellular samples.[25] Further, paucity of efficient biophysical probes that can differentiate and quantitatively report ligand-binding to different GQ structures and nucleic acid type has hampered the discovery of clinically viable GQ binders. Therefore, development of screening-compatible biophysical platforms that would enable the easy detection and estimation of ligands binding to different GQ structures in a cellular model will be highly beneficial in not only advancing our understanding of GQ structures and recognition in cellular milieu, but also could support approaches to discover efficient GQ-binders.

Herein, we describe the development of a platform to investigate the structure and ligand-binding ability of H-Telo DNA and TERRA RNA in a cell-like confined environment by using microenvironment-sensitive fluorescent nucleoside probes and reverse micelles (RM, Figure 1). Nano-sized water pool encapsulated in RM is an established membrane model, which is well known to mimic the physical characteristics and crowded environment

of cells.[28] The emissive nucleoside probe is based on a 5-benzofuran uracil core,[29,30] and it faithfully reports the environment of water encapsulated in RM via changes in its fluorescence properties. The useful features of RM as well as the conformation-sensitivity of the nucleoside probe facilitated the comparison of GQ topologies adopted by H-Telo DNA and RNA in aqueous buffer and a confined environment. Furthermore, this GQ sensor enabled the development of a simple fluorescence assay to quantify the ligand-binding ability of H-Telo DNA GQ structure in a confined environment. Our results indicate that such an emissive GQ sensor could provide new opportunities to study and step up discovery assays in cell-like environment to identify efficient GQ binders of therapeutic potential.
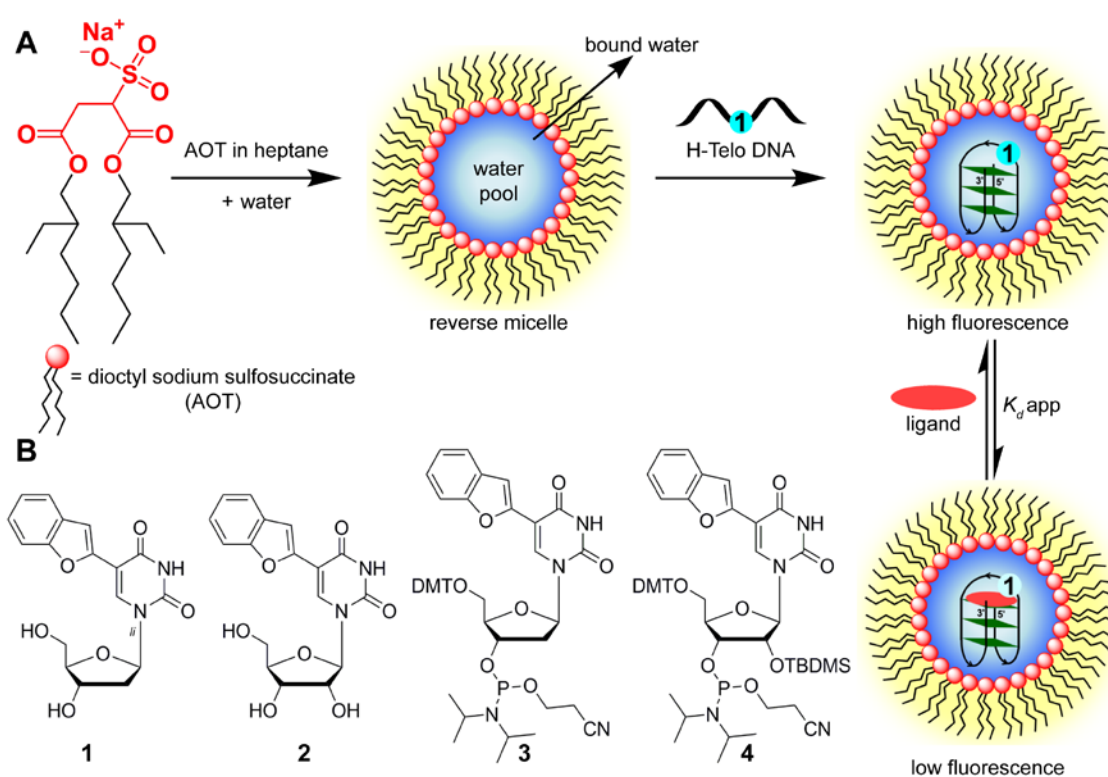
Figure 1. (**A**) A schematic illustration of the platform to study the structure and ligand-binding ability of GQ forming ONs (e.g., H-Telo DNA repeat) in a cell-like confined environment using 5-benzofuran-modified 2′-deoxyuridine (**1**) and AOT RM. The water pool solubilizes the labeled H-Telo DNA and RNA ONs and supports the formation of GQ structure. The emissive nucleoside probe detects the formation of GQ structure and also helps in determining the binding affinity of a ligand to H-Telo DNA and RNA GQ structure in confined environment. (**B**) Chemical structure of emissive nucleosides **1** and **2** and phosphoramidite substrate **3** and **4** used in the synthesis of labeled H-Telo DNA and RNA ONs, Respectively.

2.2 Results and Discussion

2.2.1 Platform design

To establish a fluorescence-based platform to study GQ structures in cell-like confined environment, we chose a combination of a widely used membrane model, AOT RM and fluorescent GQ sensors, 5-benzofuran-2′-deoxyuridine (**1**) and uridine (**2**) analogues (Figure 1). A ternary mixture of AOT, apolar solvent and water forms stable micellar aggregates in which the polar head groups of AOT face the water and hydrophobic chains extend towards the apolar solvent.[31] The size of the encapsulated water droplet increases linearly with increasing w_0 value (w_0 = [water]/[AOT]). Notably, the dynamics, polarity, viscosity and proton-transfer efficiency of the aqueous micellar core are quite different compared to that of bulk water or aqueous buffer in which most biophysical analyses are performed.[32] In addition, RM are transparent and can solubilize nucleic acids uniformly, which qualify them as a convenient cellular model for studying nucleic acid structure by various spectroscopy techniques.[33] For example, the effect of confinement on the ON dynamics and conformational flexibility of therapeutically important hairpin RNA motifs like HIV TAR and U4 snRNA has been studied by fluorescence and NMR techniques.[34,35] In another report, slower hybridization rate of DNA ONs in RM has been aptly utilized in detecting the mismatches in DNA duplexes by circular dichroism (CD).[36] More recently, absorption and CD techniques have been used to evaluate the stability and conformation of GQ in RM formed by cationic and anionic surfactants.[37-39]

We have recently introduced fluorescent nucleoside analogs derived by attaching a benzofuran moiety at the 5-position of 2′-deoxyuridine (**1**) and uridine (**2**).[29,30] The nucleoside analogs are reasonably emissive and their fluorescence properties are highly sensitive to changes in solvent polarity and viscosity. Interestingly, upon incorporation into one of the loop residues (TTA) of the human telomeric DNA and equivalent telomeric repeat containing RNA (TERRA) ONs, the probes photophysically distinguished the formation of different DNA and RNA GQ structures in aqueous buffer.[40] Together, the environment-sensitivity of the nucleoside probes and the ability of the aqueous micellar core to mimic the intracellular environment, as described above, provided the impetus to set up an efficient platform to probe the topology and binding affinity of H-Telo DNA and TERRA RNA repeats in cell-like confined environment.

2.2.2 Benzofuran-modified nucleoside probe senses the microenvironment of AOT RM

The water encapsulated in RM is configured into distinct domains namely "bound" water, which hydrates the head groups and counter ions, and "free" water in the inner core (Figure 1).[41] Typically at $w_0 < 8$, the majority of water molecules interact with head groups and counter ions forming a structured domain, which is more viscous and less polar. However, upon increasing w_0 values a well defined water pool emerges, which has higher polarity and lower viscosity as compared to the water at the interface. In order to evaluate the efficacy of the nucleoside probe to report the microenvironment of AOT RM, we measured the fluorescence properties of benzofuran-modified uridine analog as a function of increasing w_0 values. Though emissive 2′-deoxynucleoside **1** and ribonucleoside **2** exhibit similar photophysical properties in solvents of different polarity and viscosity (Table 1),[29,30] the ribonucleoside analog **2** was preferred in this study because of its higher solubility in AOT RM.

Nucleoside **2** exhibits a significantly higher quantum yield (0.21) and lifetime (2.55 ns) in water as compared to in a nonpolar solvent like dioxane (0.10 and 0.43 ns, respectively, Table 1). In water the emission maximum is centered at ~447 nm, which considerably blue-shifted to 404 nm in dioxane. The presence of a molecular rotor element (rotatable bond between benzofuran and uracil rings) also affected the fluorescence properties of the nucleoside in solvents of different viscosity (Table 1). Increasing the viscosity of the medium from ethylene glycol to glycerol resulted in a discernible increase in fluorescence quantum yield and anisotropy with no apparent change in emission maximum. Collectively, these results indicate that the nucleoside is environment- and conformation-sensitive.[42] The fluorescence signatures obtained from bulk solvents were further used to determine the responsiveness of the nucleoside analog in the confined water of AOT RM.

Table 1. Fluorescence properties of nucleoside analogs **1** and **2** in solvents of different polarity and viscosity.[29,30]

Nucleoside Analog	Solvent	$\lambda_{max}{}^a$	λ_{em} (nm)	$I_{rel}{}^b$	Φ	$\tau_{av}{}^c$ (ns)	$r^{\,c}$
1	water	322	446	1.0	0.190	2.38	0.018
	methanol	322	423	0.8	0.120	0.78	nd
	acetonitrile	322	411	0.3	0.040	0.27	nd
	dioxane	322	406	0.4	0.070	0.33	nd
	ethylene glycol	322	429	2.0	0.399	2.50	0.140
	glycerol	322	427	2.7	0.537	3.56	0.323
2	water	322	447	1.0	0.212	2.55	0.016
	methanol	322	423	0.8	0.145	0.94	nd
	acetonitrile	322	410	0.4	0.063	0.33	nd
	dioxane	322	404	0.6	0.099	0.43	nd
	ethylene glycol	322	428	1.9	0.377	3.13	0.125
	glycerol	322	427	2.4	0.528	3.35	0.315

[a]lowest absorption energy maximum is given. [b]Relative emission intensity is given with respect to intensity in water. [c]Standard deviations for τ_{av} (average lifetime) and r (anisotropy) in different solvents are ≤ 0.02 ns and ≤ 0.003 respectively. nd = not determined

RM samples of increasing w_0 value were prepared by adding appropriate amounts of the nucleoside **2** dissolved in water to AOT (sodium salt) in heptane such that the concentration of nucleoside and AOT was maintained at 1 µM and 200 mM, respectively. At low water content ($w_0 = 0.6$–2.2), the nucleoside exhibited an emission peak around 403 nm, which is similar to the λ_{em} in dioxane but considerably blue-shifted compared to in bulk water ($\lambda_{em} = 447$ nm, Figure 2, compare Table 2 and Table 1). As the water content was increased, discernible quenching in fluorescence intensity accompanied by a red-shifted emission was observed. The intensity and emission maximum saturated out at $w_0 = 11$, which apparently is similar to the emission maximum of the nucleoside **2** in methanol (~423 nm, Figure 2, Table 1, and Table 2). This result is in good agreement with literature reports, which have also predicted a polarity equivalent to methanol for the water encapsulated in the AOT RM.[32b,43] Consistent with the steady-state fluorescence data, the excited-state lifetime of the nucleoside

decreased with increase in the size of the water pool. Similarly, anisotropy measurements revealed a gradual reduction in anisotropy of the nucleoside with increasing w_0 (Table 2). This observation is consistent with the lower viscosity of the aqueous micellar core at higher w_0 as compared to the high viscous water domain at lower w_0.[32c]
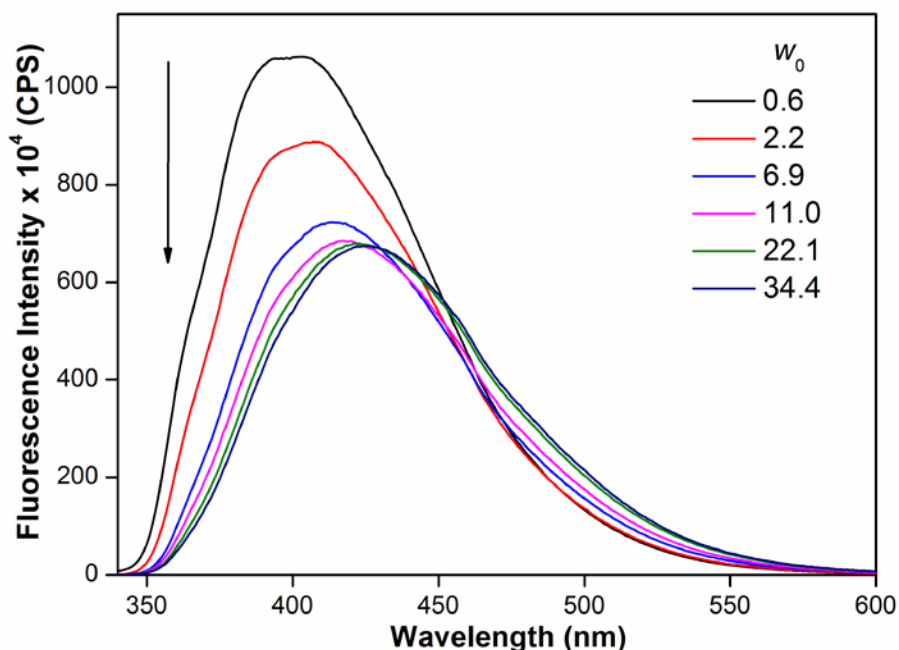
Figure 2. Emission spectra of nucleoside **2** (1 µM) in AOT RM (200 mM, *n*-heptane) at different w_0 values. Samples were excited at 322 nm with an excitation and emission slit width of 3 nm and 4 nm, respectively. The fluorescence studies of nucleoside **2** in AOT RM with different w_0 vaue was carried out by Dr. Cornelia Panse.
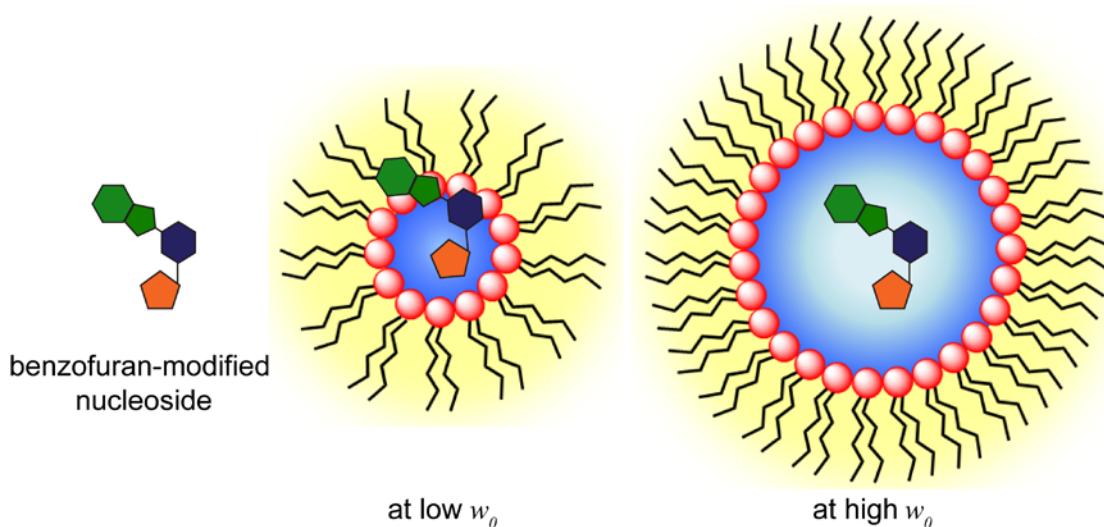
Table 2. Fluorescence properties of nucleoside 2 in AOT RM (200 mM in *n*-heptane) as a function of increasing w_0.

w_0	λ_{em} (nm)	τ_{av}^{a} (ns)	r^{a}
0.6	403	1.80	nd
2.2	407	1.70	0.192
6.9	414	1.46	0.171
11.0	420	1.48	0.132
22.1	423	1.44	0.108
34.4	427	1.57	nd

[a] Standard deviation for average lifetime (τ_{av}) and anisotropy (r) at different w_0 values are ≤ 0.06 ns and ≤ 0.006 respectively. nd = not determined.

The ability of the nucleoside probe to photophysically report the microenvironment of the water pool in RM at different w_0 values will depend on the location of the probe within RM at a given w_0 value. This can be deduced by comparing the fluorescence properties of the

nucleoside in bulk solvents and RM. Rigidification of the fluorophore (benzofuran-modified uracil) in viscous medium will result in an increase in fluorescence intensity and anisotropy, whereas, change in polarity of the medium from nonpolar to polar will result in an increase in fluorescence intensity and a red shift in emission maximum. In the absence of a well-defined water pool at low w_0 values, it is likely that the nucleoside is solubilized by "bound" water at the AOT-water interface, which is less polar and more viscous (Figure 3).[41] Hence, we observed an intense blue-shifted emission band and higher anisotropy due to a nonpolar environment around the rigidified fluorophore (Figure 2 and Table 2). However, when the water/AOT ratio is increased, the probe permeates from a more viscous and nonpolar domain to a less viscous and more polar water pool (Figure 3). Hence, the nucleoside probe located in the aqueous micellar core is derigidified and surrounded by more polar water resulting in fluorescence quenching, red-shifted emission band and reduced anisotropy. From these observations, it can be concluded that the emissive nucleoside analog faithfully reports the microenvironment of the confined water. Hence, we decided to deploy the nucleoside analog in a fluorescence assay to probe the structure and recognition properties of H-Telo DNA and RNA repeats in RM.

benzofuran-modified
nucleoside

at low w_0

at high w_0

Figure 3. Based on the fluorescence properties of nucleoside probe **2** in AOT RM at different w_0 values a working model depicting the location of the probe in RM at low and high w_0 values is given. At low w_0 values (0.6–2.2), the nucleoside is solubilized in a less polar and viscous domain at the interface of polar head groups and structured water associate with the head groups. At higher w_0 values (above 11), well defined water pool appears, and the nucleoside permeates to more polar and less viscous water pool.[32b,44]

### 2.2.3 Probing H-Telo DNA and RNA GQ structure in buffer and AOT RM

Among the various GQ forming sequences, the H-Telo DNA repeat is the most studied because of its structural polymorphism and biological roles.[2,5] Telomeric DNA, which end-caps and protects the chromosomes from degradation and fusion has been implicated in carcinogenesis and ageing related diseases.[6] Biochemical investigations using GQ stabilizing ligands point out that telomeric DNA uses this structural element to carry out its function. However, it is still not clear which topology it adopts and how different topologies interact with ligands in cellular environment.

Telomeric DNA forms various GQ structures *in vitro* depending on the ionic conditions.[16] In general, H-Telo GQ structures are made of three loops and three G-tetrads, which are stacked above each other. Notably, the conformation of the loop residues (TTA) in each topology is distinctly different. Hence, we replaced one of the dT residues in the first, second and third loop of H-Telo DNA repeat $AGGG(TTAGGG)_3$ with benzofuran-modified 2′-deoxyuridine analog **1** and utilized its fluorescence readout to detect and discriminate different GQ structures (Figure 4). The modified H-Telo DNA ONs **5**–**7** were synthesized by site-specifically incorporating the phosphoramidite **3** using solid-phase ON synthesis method.

Although TERRA is an integral part of the telomere, the majority of studies have focused on understanding the structural features and function of the telomeric DNA repeat. Biochemical investigations reveal that TERRA plays an important role in the maintenance of the telomere structure, heterochromatinization, and replication.[45] In this regard, we have synthesized fluorescently modified TERRA ON **8** $(U_2AG_3)_4$, in which the uridine residue in the middle loop has been replaced with **2** by solid-phase ON synthesis using corresponding phosphoramidite **4**. The phosphoramidites **3** and **4** required for solid-phase DNA and RNA synthesis were synthesized by following reported procedure.[29,40] Modified ONs were purified by denaturing polyacrylamide gel electrophoresis. Purity and integrity of DNA ONs **5**–**7** and RNA ON **8** were confirmed by HPLC and MALDI-TOF MS analyses, respectively (Figures 5 and 6, Table 3).

5 5' d(AGGGTTAGGG**1**TAGGGTTAGGG) 3'

6 5' d(AGGG**1**TAGGGTTAGGGTTAGGG) 3'

7 5' d(AGGGTTAGGGTTAGGG**1**TAGGG) 3'

8 5' r(UUAGGGUUAGGG**2**UAGGGUUAGGG) 3'

9 5' d(AGGGTTAGGGTTAGGGTTAGGG) 3'

10 5' r(UUAGGGUUAGGGUUAGGGUUAGGG) 3'

11 5' d(CCCTAACCCTAACCCTAACCCT) 3'

Figure 4. Benzofuran-modified H-Telo DNA ONs **5**–**7**. dT residue in the first (**6**), second (**5**) and third (**7**) loop of H-Telo DNA ONs was replaced with emissive nucleoside **1**. Benzofuran modified TERRA RNA **8**. rU residue in the second loop was replaced with emissive nucleoside **2**. ON **9** and **10** are the control unmodified H-Telo DNA and TERRA RNA respectively. ON **11** is the complementary to DNA ONs **5**–**7** and RNA ON **8**.
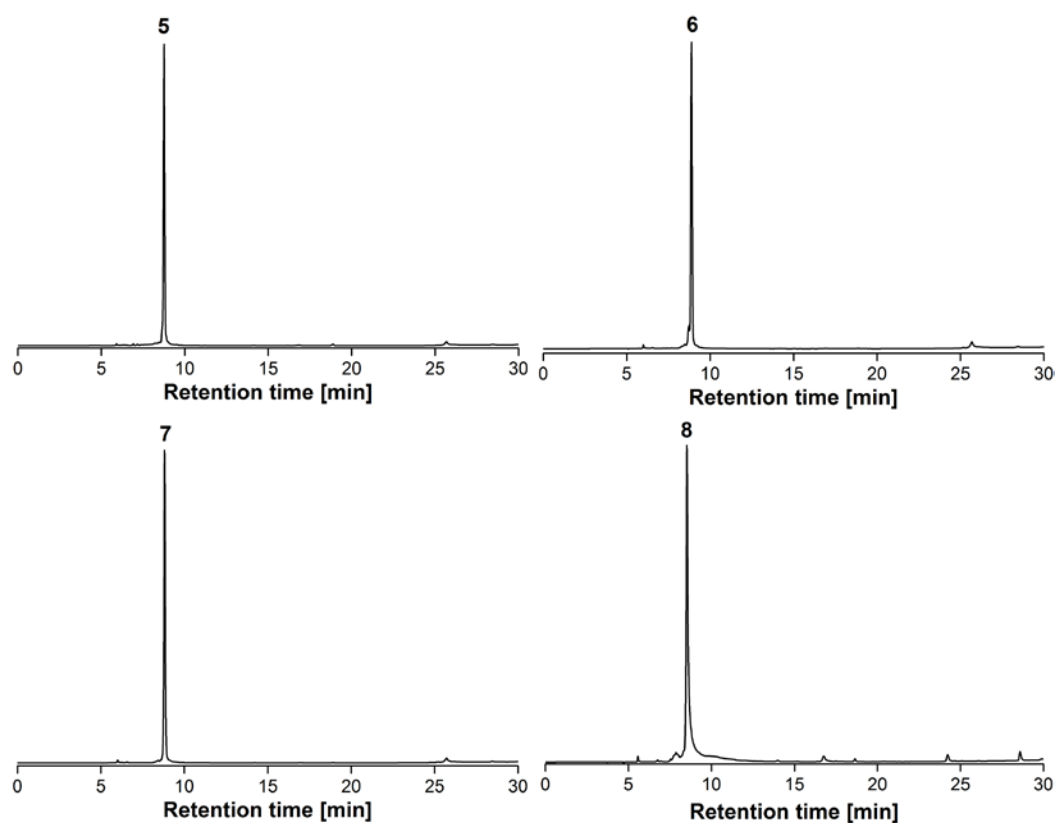
Figure 5. HPLC chromatograms of PAGE purified fluorescent ONs **5**–**8** at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min.
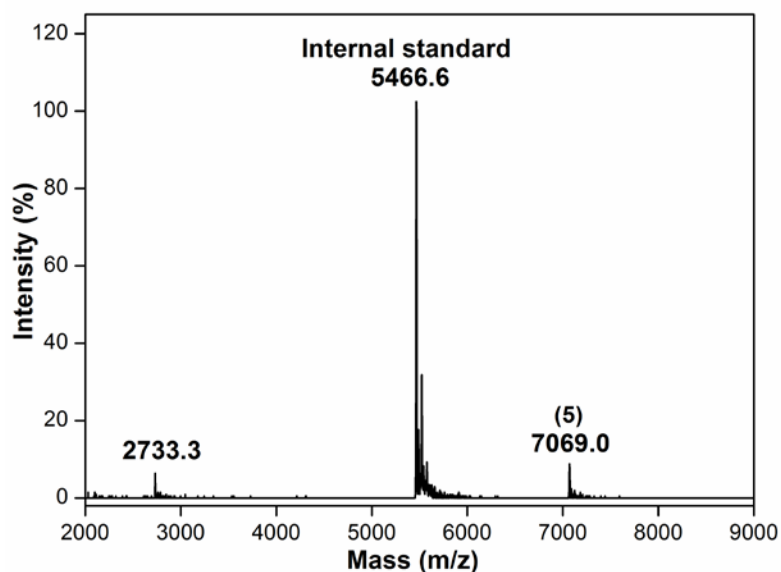
Figure 6. MALDI-TOF MS spectrum of modified H-Telo DNA ON **5** calibrated relative to the +1 and +2 ion of an internal 18-mer DNA ON standard (m/z for +1 and +2 ions are 5466.6 and 2733.3, respectively). Calcd. for **5**: 7068.6 [M]$^+$; found: 7069.0.

Table 3. Extinction coefficient and mass of modified H-Telo DNA and RNA ONs **5–8**.

| H-Telo DNA ON	$\varepsilon_{260}{}^a$ (M^{-1}cm^{-1})	Calculated mass	Observed mass
5	232713	7068.6	7069.0
6	232713	7068.6	7068.6
7	232713	7068.6	7067.9
8	263253	7962.8	7962.9

^{a} Molar absorption coefficient ε of the modified ONs was determined by using OligoAnalyzer 3.1, which was used for the determination of concentration of modified ONs. The extinction coefficient of nucleoside **1** (ε_{260} = 12613 M^{-1}cm^{-1}) and **2** (ε_{260} = 17253 M^{-1}cm^{-1}) was used in the place of thymidine and uridine, respectively.

CD and thermal melting (T_m) studies were carried out to investigate the effect of modification on the structure and stability of GQs of ONs **5–10**. ONs **5–10** were annealed to form GQ structures in Tris-HCl buffer containing NaCl or KCl. In K$^+$ solution, the DNA ONs **5–7** and **9** displayed a positive peak at ~290 nm and a shoulder at ~270 nm characteristic of hybrid-type structures (Figure 7A, **5–7** and **9** in KCl).[16b] In the presence of Na$^+$ ions, the ONs **5–7** and **9** gave a positive peak at ~294 nm and a strong negative peak at ~263 nm reminiscent of an antiparallel GQ structure (Figure 7B, **5–7** and **9** in NaCl).[16a] On the other hand, irrespective of metal ion Na$^+$ or K$^+$, RNA ONs **8** and **10** showed a positive peak at

~265 nm and a negative peak at ~240 nm characteristic of parallel GQ structure (Figure 8).[24a,46] T_m values of unmodified and modified DNA and RNA GQs in KCl/NaCl were found to be similar and consistent with literature reports (Figure 9, Figure 10 and Table 4).[47] These results indicate that the benzofuran modification is structurally minimally invasive and does not hamper the formation of respective GQs in the presence of K^+ and Na^+ ions.
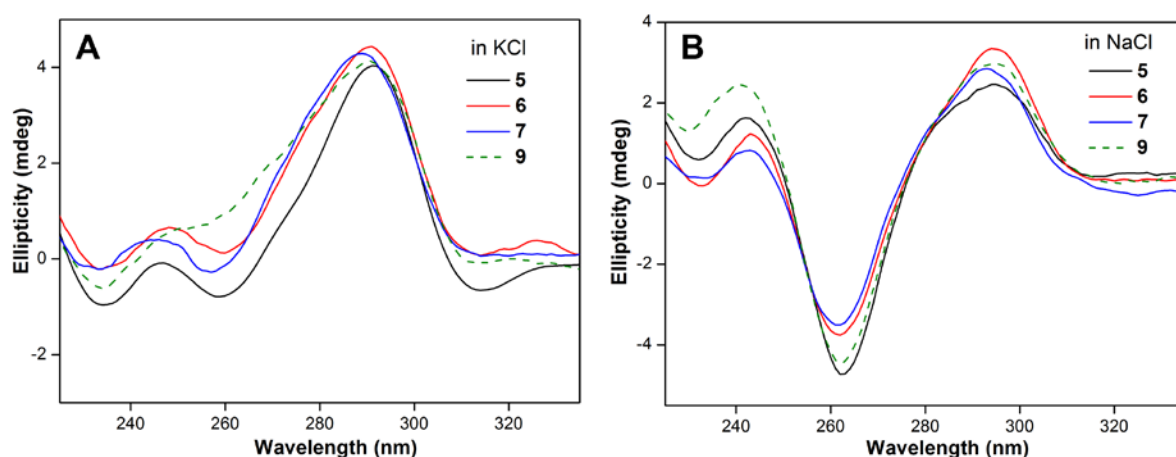

Figure 7. CD spectra of fluorescently modified H-Telo DNA ONs **5**–**7** (8 µM, solid lines) and control unmodified H-Telo DNA ON **9** (8 µM, dashed green lines) in aqueous buffer. (**A**) In Tris-HCl buffer (pH 7.5) containing 50 mM KCl and (**B**) in Tris-HCl buffer (pH 7.5) containing 50 mM NaCl.
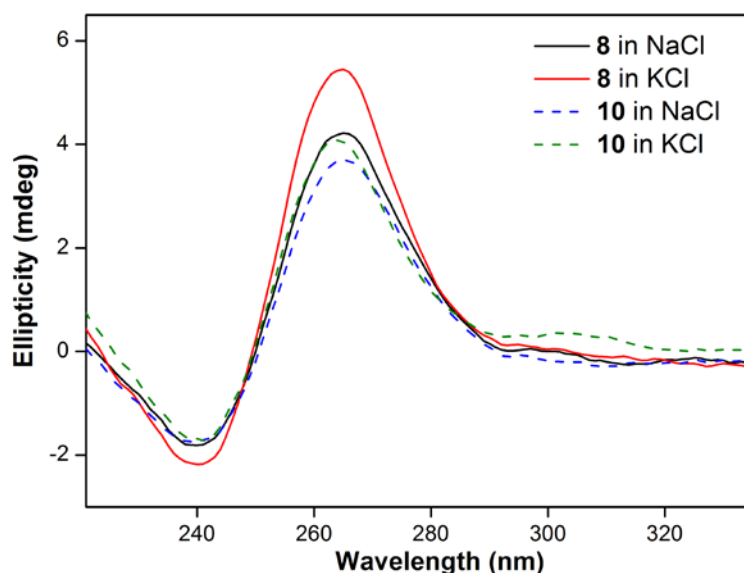
Figure 8. CD spectra of fluorescently modified TERRA RNA ONs **8** (8 µM, solid lines) and control unmodified TERRA RNA ON **10** (8 µM, dashed lines) in Tris-HCl buffer (pH 7.5) containing either 50 mM KCl or 50 mM NaCl.
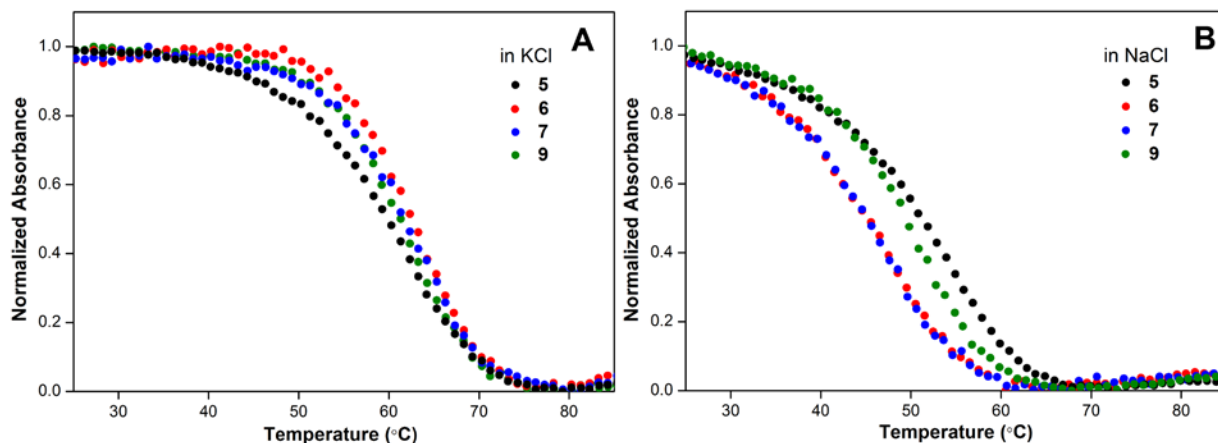
Figure 9. UV-thermal melting profile of fluorescent H-Telo DNA ONs (**5**–**7**) and control unmodified DNA ON **9** in Tris-HCl buffer (pH 7.5) containing 50 mM KCl (**A**) and 50 mM NaCl (**B**) at 295 nm. For experimental details see experimental section and for T_m values see Table 4.
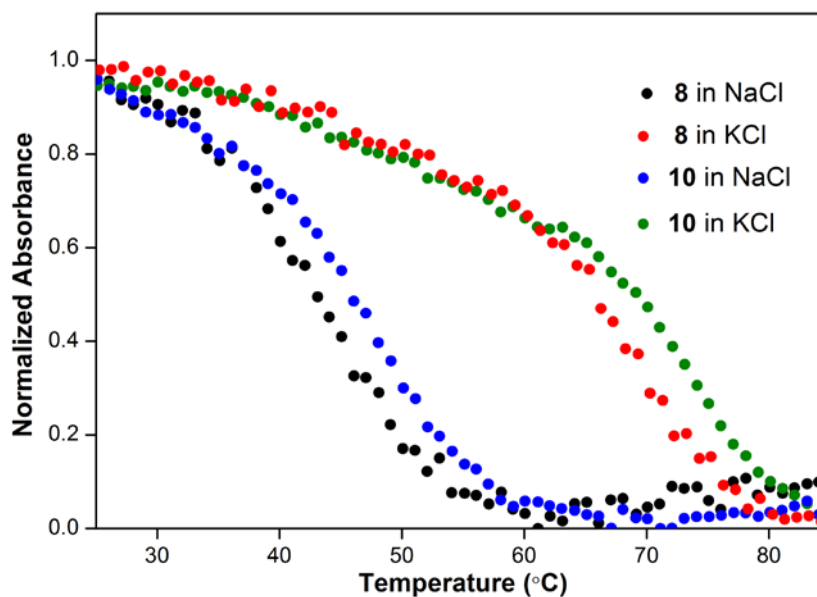
Figure 10. UV-thermal melting profile of fluorescent TERRA RNA ON **8** and control unmodified TERRA ON **10** in Tris-HCl buffer (pH 7.5) containing either 50 mM KCl or 50 mM NaCl at 295 nm. For experimental details see experimental section and for T_m values see Table 4.

Table 4. T_m values of modified ONs (**5**–**8**) and control unmodified ONs **9** and **10**

Modified and control unmodified H-Telo DNA and RNa ONs	T_m (°C) in NaCl	T_m (°C) in KCl
H-Telo DNA **5**	53.9 ± 1.1	61.5 ± 0.2
H-Telo DNA **6**	48.0 ± 1.0	63.0 ± 1.0
H-Telo DNA **7**	49.2 ± 0.5	62.8 ± 0.5
control H-Telo DNA **9**	52.2 ± 0.9	62.0 ± 0.7
TERRA **8**	44.9 ± 0.9	71.1 ± 0.7
control TERRA **10**	46.9 ± 0.1	73.0 ± 0.5

2.2.3.1 Detection of H-Telo DNA GQ structure in aqueous buffer

DNA ONs **5**–**7** in the presence of K^+ ions, which predominantly favor the formation of hybrid-type mixed parallel-antiparallel GQ structures, displayed significant enhancement in fluorescence intensity (9–14-fold) as compared to the respective duplexes formed by hybridization with complementary DNA ON **11** (Figure 11A, 11B, 11C). Interestingly, in Na^+ ionic condition, which is known to induce antiparallel GQ structure, the ONs exhibited further enhancement in fluorescence intensity as compared to GQs (2–3-fold) and duplexes (17–40-fold) in K^+ and Na^+ ionic conditions, respectively. Antiparallel structure of ONs **5**–**7** in the presence of Na^+ ions displayed discernibly higher excited-state lifetime as compared to the hybrid-type structures formed in the presence of K^+ ions (Table 5). Further, the minimally perturbing nature of the nucleoside probe as confirmed by CD and thermal melting measurements indicate that the observed fluorescence is a true reflection of the respective GQ structures formed in different ionic conditions.

2.2.3.2 Detection of TERRA RNA GQ structure in aqueous buffer

RNA ON **8**, in the presence of solutions of Na^+ and K^+, showed a highly intense fluorescence band (~4 fold), relative to that of its duplex **8•11** (Figures 11D). Furthermore, through time-resolved fluorescence analysis, the GQ structure of **8** showed lifetimes of 2.8 and 2.9 ns in the presence of Na^+ and K^+ ions, respectively (Table 6). Comparable fluorescence properties exhibited by the GQs of **8** under both Na^+ and K^+ ionic conditions are consistent with the parallel GQ structure confirmed by CD analysis (Figure 8).
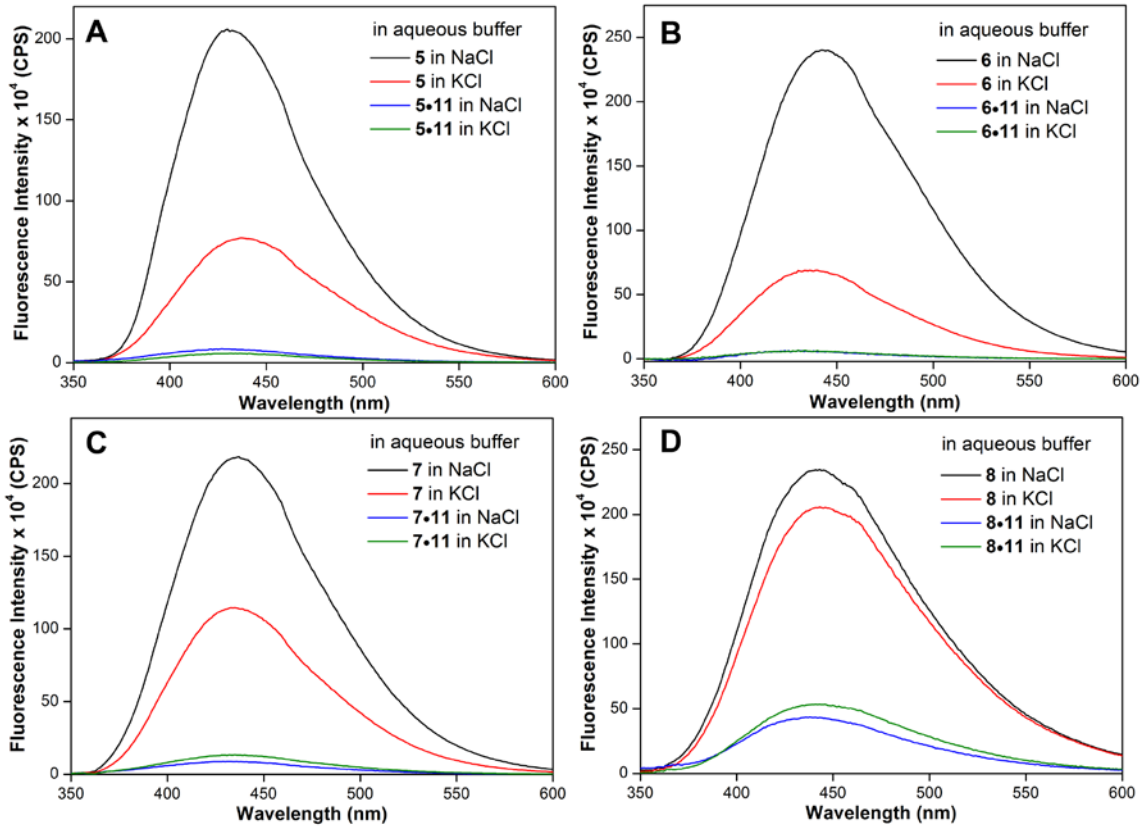
Figure 11. Steady-state fluorescence spectra of (**A**) H-Telo DNA ON **5** and corresponding duplex **5•11**, (**B**) H-Telo DNA ON **6** and corresponding duplex **6•11**, (**C**) H-Telo DNA ON **7** and corresponding duplex **7•11** and (**D**) TERRA RNA ON **8** and corresponding duplex **8•11** in Tris-HCl buffer (pH 7.5) containing 50 mM NaCl or 50 mM KCl. ON DNA samples (1 µM) were excited at 322 nm with an excitation and emission slit width of 3 nm and 4 nm, respectively. ON RNA samples (0.26 µM) were excited at 322 nm with an excitation and emission slit width of 6 nm and 8 nm, respectively.

Table 5. Fluorescence properties of GQs of H-Telo DNA ONs 5–7 and respective duplexes in aqueous buffer and AOT RM at $w_0 = 20$.

	ON sample	λ_{em} (nm)	$\tau_{av}{}^a$ (ns)	ON sample	λ_{em} (nm)	$\tau_{av}{}^a$ (ns)	ON sample	λ_{em} (nm)	$\tau_{av}{}^a$ (ns)
aqueous Buffer	**5** in NaCl	430	2.07	**6** in NaCl	442	1.65	**7** in NaCl	437	1.78
	5 in KCl	436	0.72	**6** in KCl	436	1.15	**7** in KCl	435	1.41
	5•11 in NaCl	427	nd	**6•11** in NaCl	430	nd	**7•11** in NaCl	433	nd
	5•11 in KCl	436	nd	**6•11** in KCl	430	nd	**7•11** in KCl	432	nd
AOT RM	**5** in NaCl	430	2.13	**6** in NaCl	434	1.67	**7** in NaCl	432	1.86
	5 in KCl	430	2.15	**6** in KCl	435	1.68	**7** in KCl	434	1.85
	5•11 in NaCl	427	nd	**6•11** in NaCl	435	nd	**7•11** in NaCl	441	nd
	5•11 in KCl	429	nd	**6•11** in KCl	436	nd	**7•11** in KCl	433	nd

[a] Standard deviation for lifetime (τ_{av}) is ≤ 0.09. nd = not determined. Excited-state lifetime of duplexes could not be determined as they displayed very low fluorescence

Table 6. Fluorescence properties of TERRA RNA 8 and respective duplexes in aqueous buffer and AOT RM at $w_0 = 20$.

	ON sample	λ_{em} (nm)	$\tau_{av}{}^a$ (ns)
	8 in NaCl	441	2.83
aqueous	**8** in KCl	443	2.94
buffer	**8•11** in NaCl	438	nd
	8•11 in KCl	441	nd
	8 in NaCl	435	2.27
AOT RM	**8** in KCl	432	2.29
	8•11 in NaCl	436	nd
	8•11 in KCl	436	nd

^{a} Standard deviation for lifetime (τ_{av}) is ≤ 0.07. nd = not determined. Excited-state lifetime of duplexes could not be determined as they displayed very low fluorescence

2.2.3.3 Detection of H-Telo DNA GQ structure in AOT RM

The water pool of AOT RM at $w_0 \sim 20$ is known to mimic the compartmentalized environment of the cell, and hence, has been used to study the conformation dynamics and hybridization rates of ONs by NMR and absorption spectroscopy techniques.[33-37] Appropriate amount of the preformed stock solution of GQs and duplexes made using H-Telo DNA ONs **5**–**7** in Tris-HCl buffer, containing NaCl or KCl, was added to AOT RM such that the concentration of AOT and ON was fixed at 200 mM and 1.0 µM respectively and w_0 value was maintained at 20. respectively. In al Upon excitation, the GQ forming ONs **5**–**7** displayed very high fluorescence intensity, whereas the corresponding duplexes were found to show very weak fluorescence (Figure 13). Unlike in aqueous buffer, where the ONs showed noticeable difference in their fluorescence intensity in Na$^+$ and K$^+$ ionic conditions, they exhibited almost similar fluorescence profile in AOT RM, irrespective of added salt (Figure 13). The excited-state lifetime of H-Telo DNA ONs in RM containing NaCl or KCl was similar and matched well with the lifetime of the antiparallel GQ structure formed in aqueous buffer containing NaCl (Table 5). These results suggest that ONs **5**–**7** in RM adopt antiparallel GQ structure irrespective of the type of added monovalent cations. This notion was further supported by CD experiments, wherein the GQ forming ONs produced similar CD profiles in AOT RM containing either Na$^+$ or K$^+$ ions, which were characteristic of an antiparallel GQ structure (Figure 14).[16a]
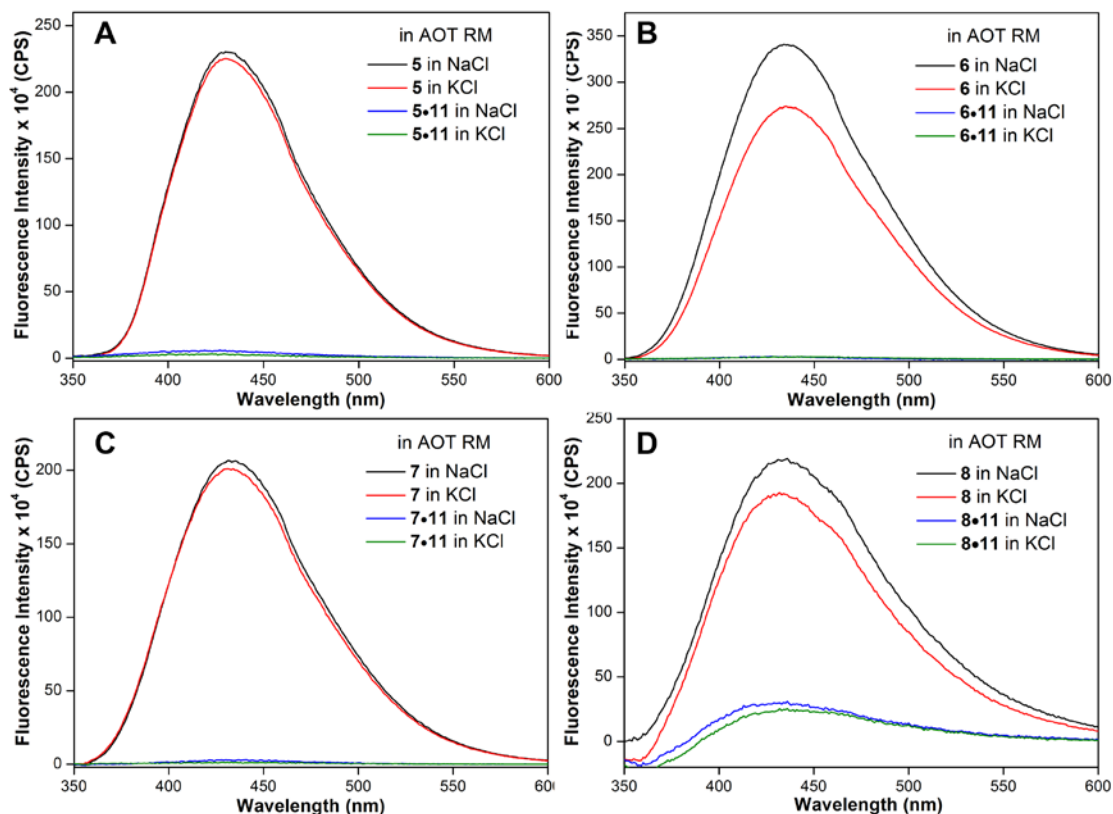
Figure 13. Steady-state fluorescence spectra of H-Telo DNA ON (**A**) **5** and corresponding duplex **5•11**, (**B**) **6** and corresponding duplex **6•11**, (**C**) **7** and corresponding duplex **7•11** and **(D)** H-Telo RNA **8** and corresponding duplex **8•11** in AOT RM at $w_0 = 20$ containing 50 mM NaCl or 50 mM KCl. DNA ON samples (1 µM) were excited at 322 nm with an excitation and emission slit width of 3 nm and 4 nm, respectively. H-Telo RNA ON samples (0.26 µM) were excited at 322 nm with excitation and emission slit width 6 and 8 nm respectively.
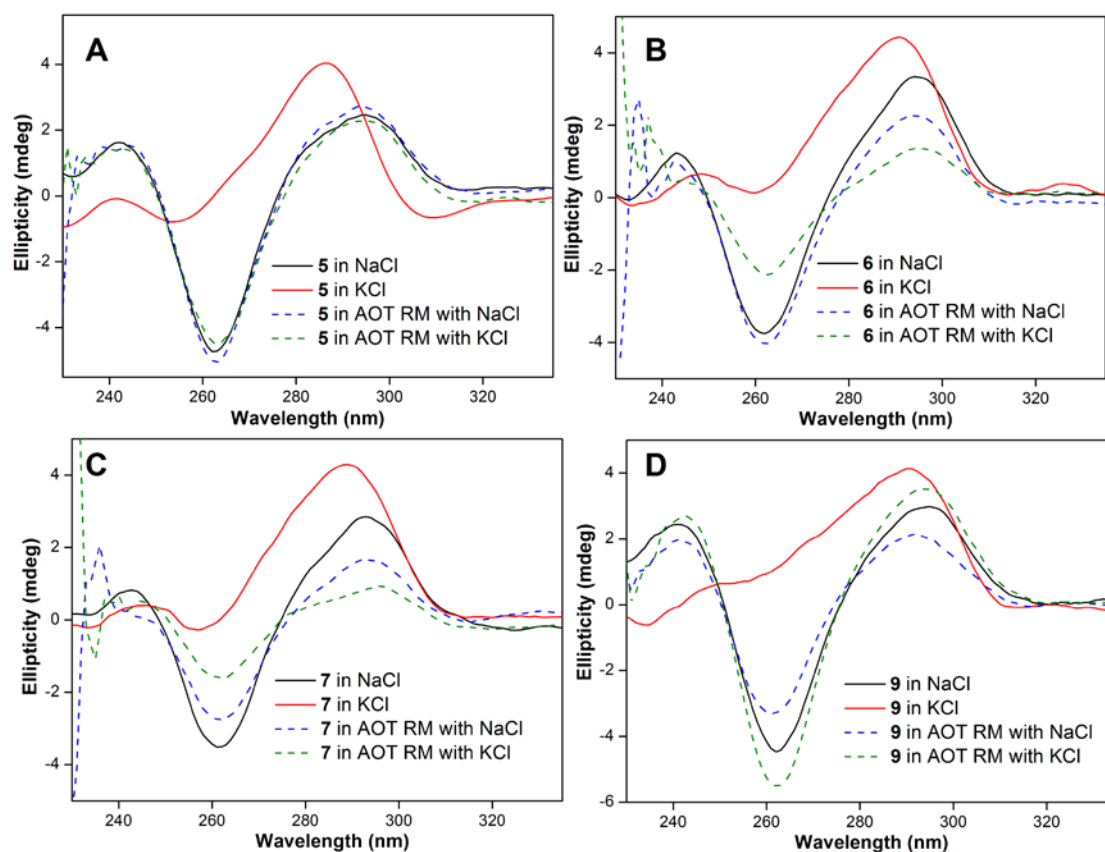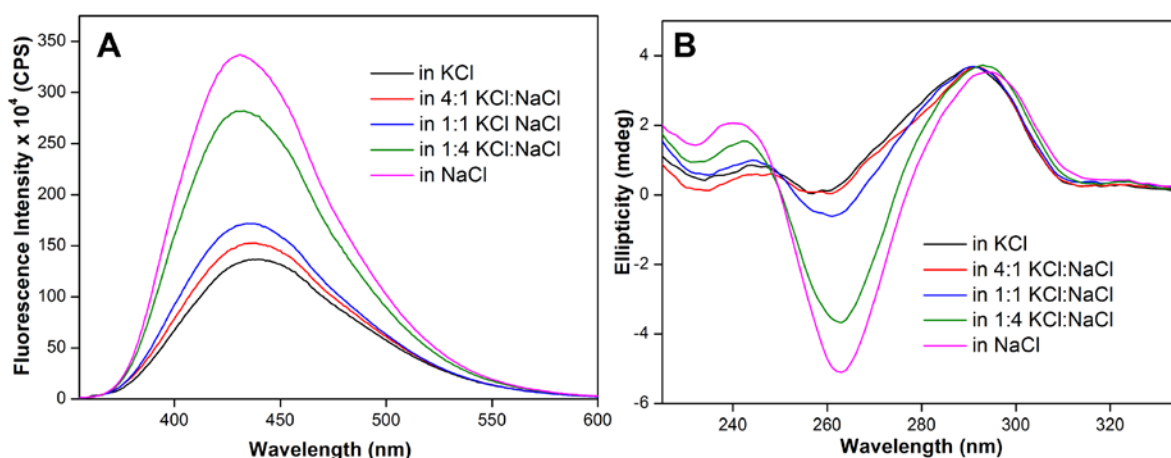
Figure 14. CD spectra of fluorescent H-Telo DNA ONs **5–7** and control H-Telo DNA ON **9** in aqueous buffer (solid lines) and AOT RM (dashed lines).

Similar to other studies, we have also used the readily available sodium salt of AOT (200 mm) to form stable RMs. The conversion of hybrid structures to an antiparallel GQ structure in RM is possibly due to the exchange of K^+ ions of the ONs with Na^+ ions of the AOT head group. This structural conversion would depend on the ratio of Na^+ to K^+ ions. Such interconversion of telomeric GQ structures as a function of different Na^+ to K^+ ratios has been analyzed by CD experiments.[16b, 48]

In order to ascertain the structural transformation happening in the RM core, we performed fluorescence and CD studies by using H-Telo DNA ON **5** in an aqueous buffer containing different ratios of Na^+/K^+ ions. The hybrid form of ON **5** in the presence of K^+ ions gave a reasonably intense fluorescence band as before; this did not change upon increasing the amount of Na^+ ions up to a ratio of 1:1 (Figure 15 A). Further addition of Na^+ ions ($Na^+/K^+ = 4:1$) resulted in a significant increase in fluorescence intensity, as a result of the conversion of hybrid to antiparallel structure. Similar results were obtained from CD experiments (Figure 15 B). In the experiments performed with RMs, the concentration of Na^+/K^+ was maintained at 4:1, and hence, ON **5** would have adopted an antiparallel GQ

structure. Collectively, these results are consistent with reports in the literature that the interconversion of telomeric GQ structures of similar sequences relies on Na$^+$ and K$^+$ ion exchange.[16b,48]

Attempts to induce the formation GQ structures supported by K$^+$ ions by increasing the KCl concentration or by using the potassium salt of AOT failed because these conditions collapsed the RMs. Nevertheless, the emissive nucleoside, irrespective of the position of modification, effectively signalled the formation of GQ structure with enhancement in fluorescence intensity in aqueous buffer and in a confined environment. This feature of the nucleoside probe is advantageous because most fluorophores, when placed in the vicinity of a guanine residue, exhibit drastically quenched emission, which hampers their practical application.[49]


Figure 15. (A) Steady state fluorescence spectra (0.26 µM) and **(B)** CD spectra (8 µM) of H-Telo DNA **5** in aqueous buffer having 50 mM total salt concentration with different ration of KCl:NaCl.

2.2.3.4 Detection of GQ structures of TERRA in AOT RM

Samples were prepared by adding appropriate amounts of preformed stock solution of TERRA ON **8** and corresponding duplex (**8•11**) in Tris·HCl buffer, containing NaCl or KCl, to AOT RMs such that the concentrations of AOT and ON and the value of w_0 were fixed at 200 mm, 0.26 mm, and 20, respectively. Upon excitation, TERRA **8** displayed an intense fluorescence band in AOT, which was similar under both NaCl and KCl ionic conditions (Figure 13D). The fluorescence intensity of the GQ of **8** was significantly higher than that of the corresponding duplex **8•11** in AOT. Furthermore, the excited-state decay kinetics of **8** in RMs containing NaCl or KCl revealed similar lifetimes of about 2.3 ns (Table 6). These results suggest that, similar to experiments in aqueous buffer, irrespective of the ionic

conditions, TERRA ON **8** adopts a parallel structure in AOT RMs. This observation was confirmed by recording the CD profile of **8** and control, unmodified TERRA ON **10** in AOT RMs containing either 50 mm NaCl or 50 mm KCl. Consistent with earlier reports,[24a, 46] the CD profiles of both modified and unmodified TERRA ONs showed a distinct positive peak at 265 nm and a negative peak at 240 nm; these were characteristic of a parallel GQ structure (Figure 16).
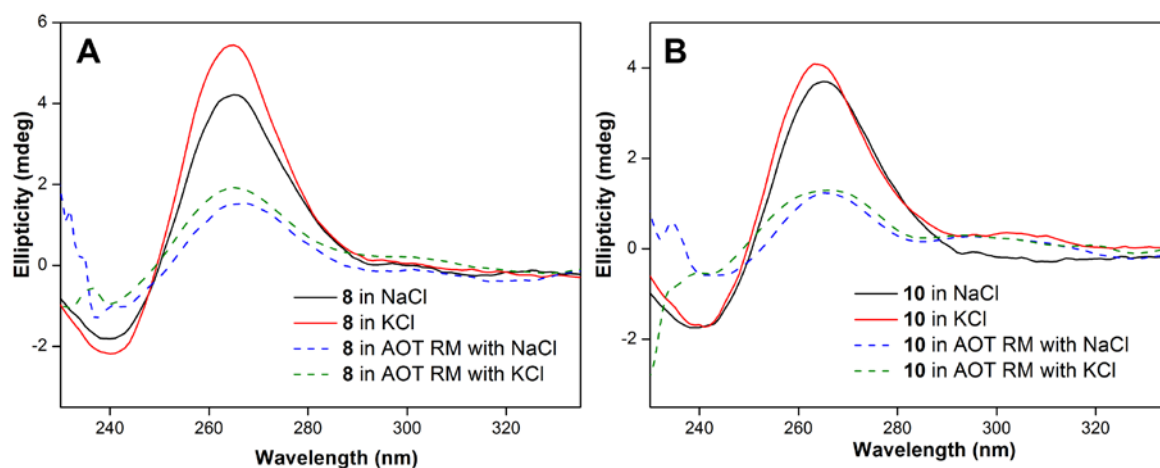

Figure 16. CD spectra of fluorescent H-Telo RNA ON **8** and control H-Telo RNA ON **10** aqueous buffer (solid lines) and AOT RM (dashed lines).

2.2.4 Probing ligand binding in aqueous buffer and RM

Biochemical and structural investigations document that ligands bind individual domains of a GQ structure with different affinities and induce significant conformational changes, particularly in the loop residues.[50,51] A simple fluorescence binding assay was designed in aqueous buffer and AOT RMs to evaluate the ability of the nucleoside analogues to signal ligand-induced conformational changes in the H-Telo DNA and TERRA GQ structures. DNA and RNA ONs **5**–**8** were titrated with pyridostatin (PDS); a known ligand that binds and alters the function of certain GQ-forming clusters of human genomic loci, including telomeres and proto-oncogenes (Figure 17A).[52] In order to draw a comparison between bulk solution and RM binding data, ONs were annealed in the presence of NaCl, which would induce the formation of an antiparallel GQ for HTelo DNA ONs and a parallel GQ for TERRA ON in both aqueous buffer and RMs (Figure 17B). Upon addition of PDS to ON **5** containing a modification in the diagonal loop, a dose-dependent quenching in fluorescence intensity (~15-fold), corresponding to an apparent $K_d = 0.38 \pm 0.02$ µM, was obtained in aqueous buffer (Figures 17B and 18A). Whereas ONs **6** and **7**, containing modifications in

the lateral loops, signaled the binding of PDS with significant quenching in fluorescence intensity, the apparent K_d values (0.53 ± 0.02 µM and 0.68 ± 0.002 µM, respectively) were higher than that of PDS binding to ON **5** (Figures 17B, 18B, 18C, and 18D). TERRA ON **8**, in which the modification is in the propeller loop, showed about a fivefold quenching in fluorescence intensity upon increasing the PDS concentration (K_d = 0.36 ± 0.02 µm; Figure 19). The differential binding affinity exhibited by PDS is likely to be due to differences in the physicochemical environment of the G-tetrad near the diagonal, lateral, and propeller loops.[51] In the absence of a change in emission maximum, fluorescence quenching upon ligand binding could be due to derigidification of the fluorophore or a proximal effect, wherein the close vicinity of a polyaromatic ligand with the fluorophore could induce nonradiative dissipation of energy.[51,53]
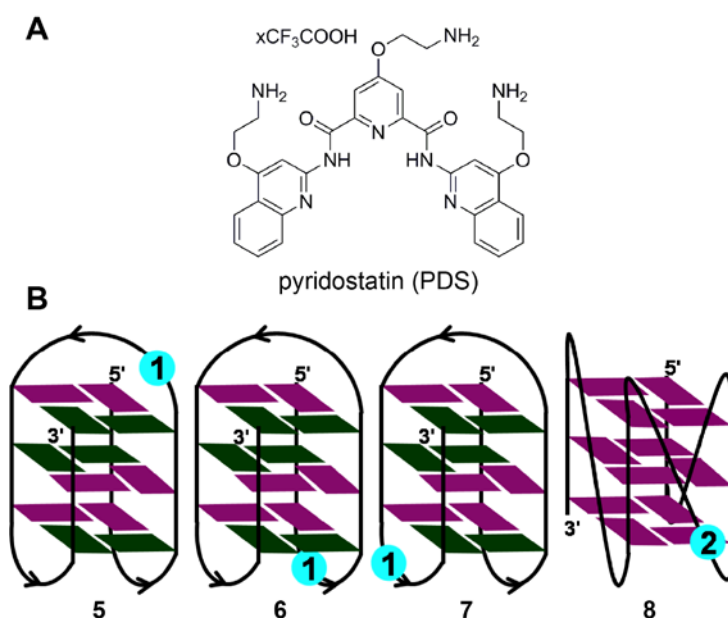
Figure 17. (**A**) Chemical structure of GQ binder, pyridostatin (PDS), used in this study. (**B**) The position of nucleoside **1** in different loops of the antiparallel GQ structure of ONs **5**–**7** is shown. In ON **5**, nucleoside **1** is placed in the diagonal loop. In ONs **6** and **7**, nucleoside **1** is placed in the lateral loops. *Syn* and *anti* guanosines are colored in green and purple, respectively. The position of nucleoside **2** in propeller loop of parallel GQ structure of RNA ON **8**.
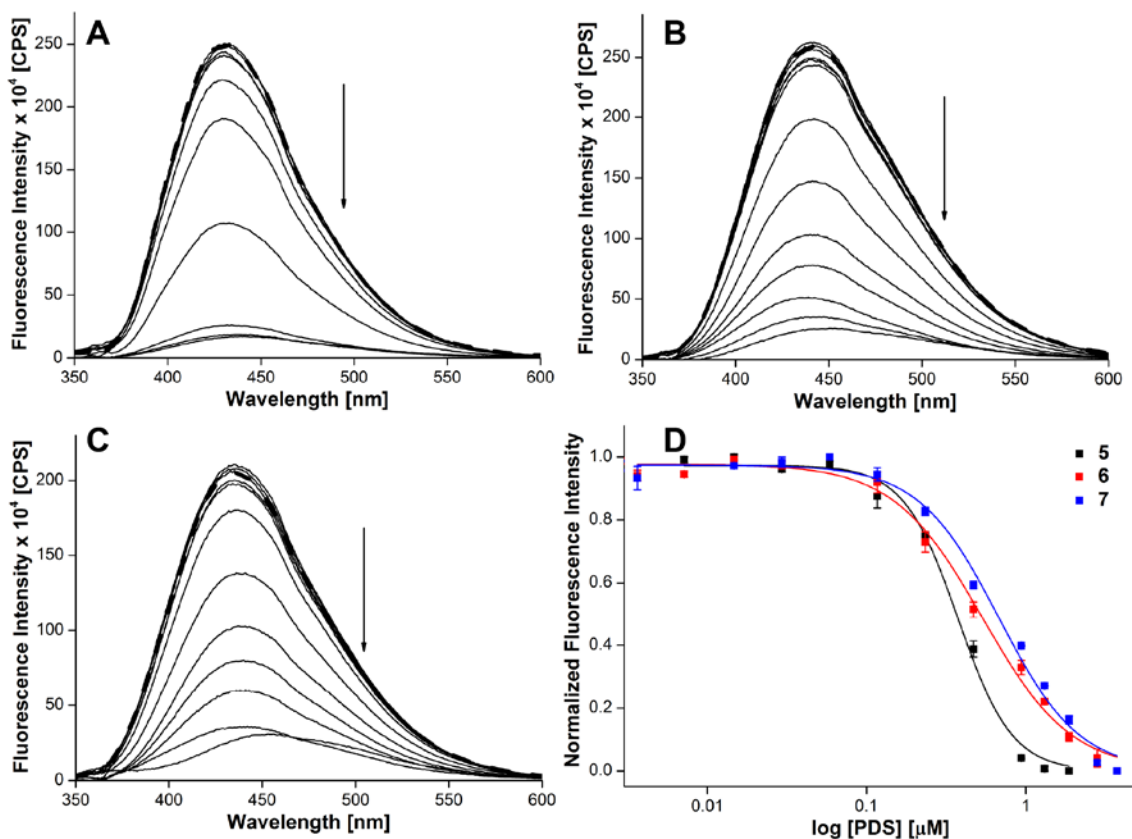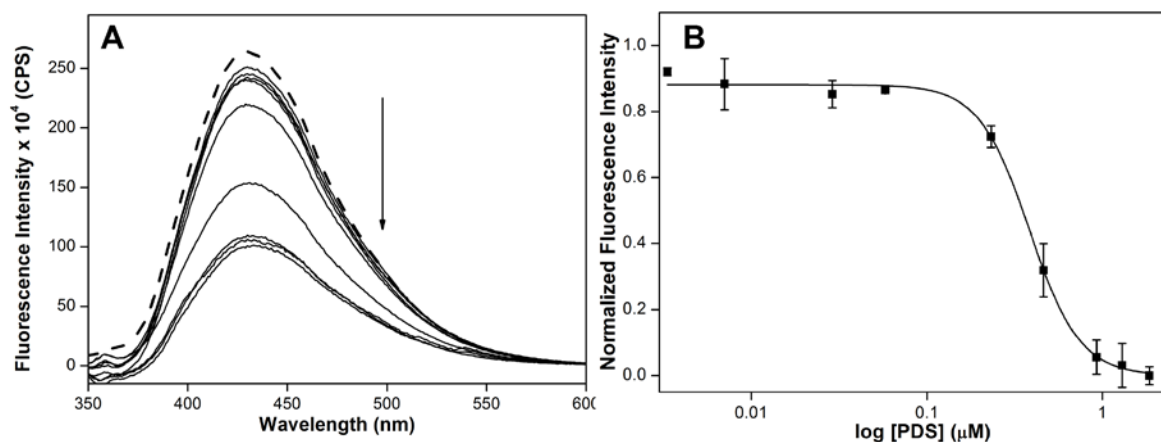
Figure 18. Emission spectra of H-Telo DNA ONs (**A**) **5**, (**B**) **6** and (**C**) **7** in aqueous buffer (pH 7.5) containing NaCl (50 mM) as a function of increasing concentration of PDS. Dashed line represents fluorescence spectrum of GQ ONs in the absence of PDS. The ON (0.28 μM) samples were excited at 322 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. (**D**) Curve fit for the binding of PDS to H-Telo DNA ONs **5–7** in aqueous buffer containing NaCl (50 mM). Normalized fluorescence intensity at respective emission maximum (Table 5) is plotted against log [PDS].
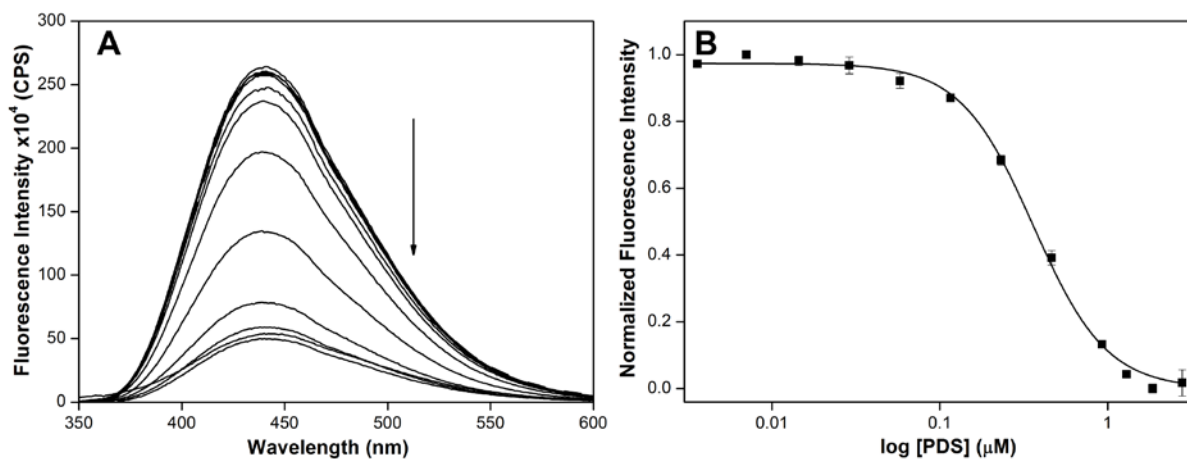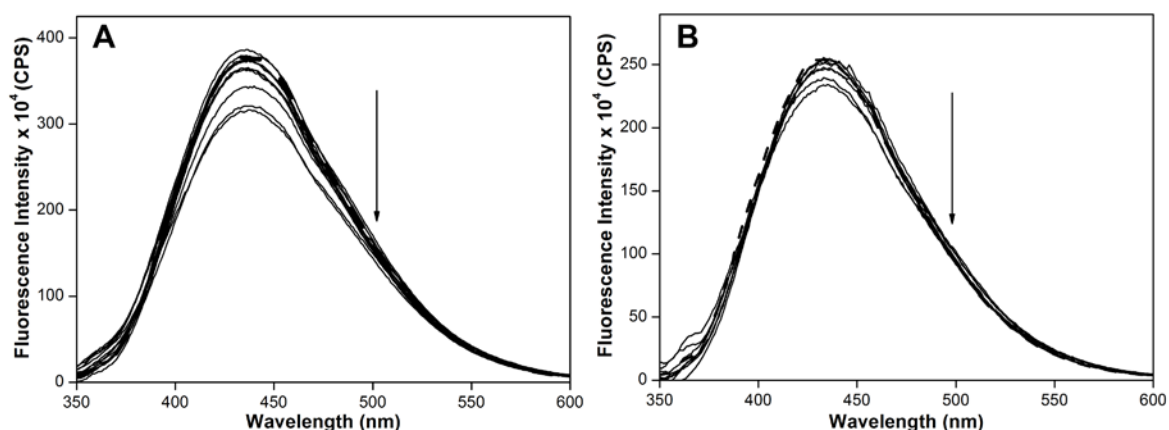
Figure 19. (**A**) Emission spectra of H-Telo RNA ON **8** (0.26 µM) in in aqueous buffer (pH 7.5) containing NaCl (50 mM) as a function of increasing concentration of PDS as a function of increasing concentration of PDS. Dashed line represents fluorescence spectrum of **8** in the absence of PDS. The samples were excited at 322 nm with an excitation and emission slit width of 6 nm and 8 nm, respectively. (**B**) Curve fit for the binding of PDS to H-Telo RNA ON **8** in AOT RM. Normalized fluorescence intensity at 441 nm is plotted against log [PDS].

The applicability of nucleoside probes **1** and **2** in estimating the binding of ligand to H-Telo DNA and TERRA GQs in confined environments was studied by preparing stock solutions of DNA and RNA ONs **5**–**8** containing increasing concentrations of PDS. The individual stock solution was then added to AOT RMs, such that w_0 was maintained at 20. The 2'-deoxyuridine analogue **1**, placed in the diagonal loop of the antiparallel GQ structure of ON **5**, showed significant quenching (~ 3 fold) in fluorescence intensity as a function of increasing PDS concentration (Figure 20). The extent of quenching observed in RMs was less than that in aqueous buffer, which could be due to the more viscous nature of the water pool compared with that of bulk water.[32c] Nevertheless, the dose-dependent quenching in fluorescence intensity enabled the determination of K_d in RMs (0.38 ± 0.05 µm), which was very close to that of the K_d in aqueous buffer. DNA ONs **6** and **7**, wherein the modification is in the lateral loops, exhibited only minor changes in fluorescence intensity upon increasing the ligand concentration, which did not yield reliable K_d values (Figure 21). This observation is interesting because the relatively stronger binding event, that is, PDS binding to the tetrad near the diagonal loop, is not affected by the cell-like confined environment, whereas the binding of PDS to the tetrad near the lateral loop, which is weaker, is affected by confinement. TERRA ON **8**, in which the modification is in the propeller loop, showed noticeable quenching (~ 2 fold) in fluorescence intensity in a dose dependent manner upon addition of PDS, which gave a K_d value of 0.39 ± 0.05 µm (Figure 22). This K_d value is close to that observed in aqueous buffer. Taken together, these results reveal that

62

compartmentalization and the chemical environment of different domains of the GQ structure determine the binding preference and affinity of ligands to DNA and RNA GQs.
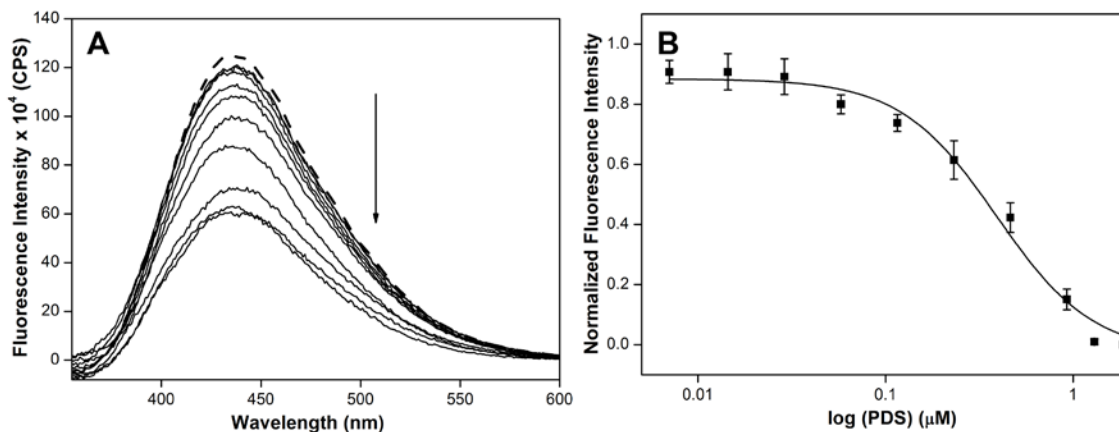

Figure 20. (**A**) Emission spectra of H-Telo DNA ON **5** (0.28 μM) in AOT RM containing NaCl (50 μM) as a function of increasing concentration of PDS. Dashed line represents fluorescence spectrum of **5** in the absence of PDS. The samples were excited at 322 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. (**B**) Curve fit for the binding of PDS to H-Telo DNA ON **5** in AOT RM. Normalized fluorescence intensity at 430 nm is plotted against log [PDS].

Figure 21. Emission spectra of H-Telo DNA ON (**A**) **6** (0.28 μM) and (**B**) **7** (0.28 μM) in AOT RM containing NaCl (50 mM) as a function of increasing concentration of PDS. Dashed line represents fluorescence spectrum of ON in the absence of PDS. The samples were excited at 322 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively.

Figure 22. (**A**) Emission spectra of H-Telo RNA ON **8** (0.26 μM) in AOT RM containing NaCl (50 mM) as a function of increasing concentration of PDS. Dashed line represents fluorescence spectrum of **8** in the absence of PDS. The samples were excited at 322 nm with an excitation and emission slit width of 6 nm and 8 nm, respectively. (**B**) Curve fit for the binding of PDS to H-Telo RNA ON **8** in AOT RM. Normalized fluorescence intensity at 435 nm is plotted against log [PDS].

2.3 Conclusions

The conformation sensitivity and ability of benzofuran-modified fluorescent nucleoside analogues **1** and **2** to sense the microenvironment of confined water of AOT RMs have been aptly utilized in investigating the GQ structures of H-Telo DNA and RNA ONs in a synthetic cellular model by means of fluorescence spectroscopy. These novel GQ sensors have essentially facilitated a direct comparison of the structure and ligand binding affinity of H-Telo DNA and RNA ON repeats in aqueous buffer and in the cell-like confined environment of RMs. The present platform, based on these nucleoside probes and RMs, could be highly suitable for setting up screening formats that mimic cellular environment to discover efficient GQ binders with therapeutic potential. In particular, an investigation of the structure and ligand-binding ability of G-rich sequences, which form similar GQ topologies, irrespective of the ionic conditions (e.g., TERRA, *C-myc*, *NRAS*) will greatly benefit from this platform.

2.4 Experimental Section

2.4.1 Materials

Dioctyl sodium sulfosuccinate (AOT) and pyridostatin (PDS) were obtained from Sigma-Aldrich. AOT was dried under vacuum for 48 h before use. *n*-heptane (HPLC grade) was purchased from RANKEM, India. Nucleoside **1** and **2** were synthesized by Sarangamath Sangamesh following the previously reported procedure.[29,30] Synthesis of phosphoramidite **3** and DNA ONs **5–7** were performed by Dr. Cornelia Panse following the previously reported metods.[29] Synthesis of phosphoramidite **4** and RNA ON **8** was performed by Dr. Vyankat

Sontakke following reported synthetic procedure.[40] *N*-benzoyl-protected dA, *N,N*-dimethylformamide-protected dG and *N*-acetyl-protected dC phosphoramidite substrates for solid-phase DNA synthesis were obtained from Proligo Reagents. Solid supports for DNA and RNA synthesis and TBDMS-protected ribonucleoside phosphoramidite substrates were purchased from GLEN Research. All other reagents for solid phase oligonucleotide synthesis were obtained from either ChemGenes corporation or Sigma Aldrich. Oligonucleotides (ONs) **9** and **11** were purchased from Integrated DNA Technologies. Custom ONs were purified by polyacrylamide gel electrophoresis (PAGE) under denaturing conditions and desalted on Sep-Pak Classic C18 cartridges (Waters Corporation). Custom synthesized oligoribonucleotide **10** was purchased from Dharmacon RNAi Technologies, PAGE-purified and desalted using Sep-Pak Classic C18 cartridge. Milipore water was autoclaved and used in all biophysical experiments.

2.4.2 Instrumentation

Mass measurements were done on an Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer. HPLC analysis of ONs was performed using Agilent Technologies 1260 Infinity HPLC. UV-thermal melting analysis of ONs was performed on a Cary 300Bio UV-Vis spectrophotometer. CD spectra were recorded on a JASCO-J-815 CD spectrometer. Steady-state and time-resolved fluorescence experiments were carried out in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on Fluorolog-3 and TCSPC instrument (Horiba Jobin Yvon, Fluorolog-3), respectively.

2.4.3 Fluorescence study of nucleoside 2 in AOT RM

To study the fluorescence properties of nucleoside **2** in AOT RM as a function of increasing w_0 values, samples were prepared by adding appropriate volumes of **2** in water to AOT in *n*-heptane such that the final ON and AOT concentration was maintained at 1 µM and 200 mM, respectively. Samples were sonicated for 30 s and equilibrated at room temperature for 3 h before use. For steady-state experiments, samples were excited at 322 nm with an excitation and emission slit width of 3 nm and 4 nm, respectively. All measurements were done in triplicate in a micro fluorescence cuvette (Hellma, path length 1.0 cm) at 20 °C. To study excited-state decay kinetics of nucleoside **2** in AOT RM at different w_0 values, samples (1 µM) were excited using 339 nm diode laser source (IBH, UK, NanoLED-339L) and emission signal was collected at the respective emission maximum. All experiments were done in

triplicate and lifetimes were calculated by fitting the decay profiles in IBH DAS6 software. The χ^2 value for all the curve fits was found to be nearly unity. Steady-state anisotropy measurements were performed by exciting the samples at 322 nm. The anisotropy value (*r*) was determined by analyzing the data using software provided with the instrument. Anisotropy measurements were performed in triplicate and the values reported in this study are an average of ten successive measurements for each sample.

2.4.4 Solid-phase synthesis of modified DNA and RNA ONs

Benzofuran-modified H-Telo DNA ONs **5**–**7** were synthesized on a 1.0 µmol scale (1000 Å CPG solid support) by a standard DNA ON synthesis protocol using phosphoramidite **3**. The solid support after final detritylation step was treated with 30% aqueous ammonium hydroxide solution for 24 h at 55 °C. The solution was evaporated to dryness using speedvac. Benzofuran-modified TERRA ON **8** was synthesized on a 1.0 mmol scale (1000 Å CPG solid support) by means of a standard RNA ON synthesis protocol with phosphoramidite **4**.[40] After trityl deprotection on the synthesizer, the solid support was treated with a 1:1 solution (1.5 mL) of methylamine (10m) in ethanol and water for 12 h. The mixture was then centrifuged, and the supernatant was evaporated to dryness on a SpeedVac. The residue was dissolved in DMSO (100 mL) and TEA·3HF (150 mL; TEA=triethylamine) was added. The resulting solution was heated at 65 °C for 2.5 h and slowly cooled to RT. The solution of deprotected RNA ON was lyophilized to dryness. The DNA and RNA ON residues were purified by 20% PAGE under denaturing conditions. The band corresponding to the full-length modified ON product was identified by means of UV shadowing, and cut and transferred to a Poly-Prep column (Bio-Rad). The gel pieces were crushed with a sterile glass rod, and the ON was extracted by using ammonium acetate buffer (0.5m, 3 mL) for 12 h and desalted with Sep-Pak classic C18 cartridges (Waters). The purity of ONs **5**–**8** was examined by reversedphase (RP) HPLC and characterized by MALDI TOF MS analyses (Figures 5 and 6, Table 3).

2.4.5 Steady-state and time-resolved fluorescence studies of H-Telo DNA and RNA ONs in buffer and AOT RM

2.4.5.1 Sample preparation in aqueous buffer

Fluorescent H-Telo DNA ONs **5–7** (1 µM) and RNA ON **8** (0.26 µM) were annealed at 90 °C for 3 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 50 mM NaCl or 50 mM KCl. The samples were slowly cooled to RT and were kept at ~4 °C overnight. DNA (1 µM) and RNA (0.26 µM) duplexes were made by heating a 1:1 mixture of H-Telo DNA or RNA ONs and complementary ON **11** at 90 °C for 3 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 50 mM NaCl or 50 mM KCl. The solutions were slowly cooled to RT and were kept at ~4 °C overnight.

2.4.5.2 Sample preparation in AOT RM

Stock solutions of H-Telo DNA ONs and their corresponding duplexes were prepared in 10 mM Tris-HCl buffer (pH 7.5) containing either 50 mM NaCl or KCl as mentioned above. Appropriate amount of the annealed DNA ON solution was mixed with AOT in n-heptane RM such that the final concentration of DNA ON, AOT and w_0 was 1 µM, 200 mM and 20, respectively. Final volume of the DNA sample was 486 µL. Sample for RNA ON was made similarly maintaining the final concentration of RNA ON, AOT and w_0 at 0.26 µM, 200 mM and 20, respectively. Final volume of RNA sample was 243 µL. The samples were sonicated for 30 s and equilibrated at RT for 3 h.

2.4.5.3 Fluorescence analysis

Steady-state fluorescence measurements of DNA ON samples in aqueous buffer and AOT RM were performed by exciting the samples at 322 nm with an excitation and emission slit width of 3 nm and 4 nm, respectively. Steady-state fluorescence measurements of RNA ON samples in aqueous buffer and AOT RM were performed by exciting the samples at 322 nm with an excitation and emission slit width of 6 nm and 8 nm, respectively. Excited-state decay kinetics study of ONs was carried out by exciting the samples using a 339 nm diode laser source (IBH, UK, NanoLED-339L). The fluorescence signal was collected at respective emission maximum and analyzed using IBH DAS6 analysis software. Lifetime measurements of DNA ONs samples were performed in triplicate and Lifetime measurements of RNA ONs

samples were performed in duplicate. χ^2 values for all curve fits were found to be close to unity.

2.4.6 CD measurements

ON samples were annealed and prepared for CD analysis in aqueous buffer and AOT RM as above. The CD spectrum of fluorescently modified H-Telo DNA and RNA ONs **5–8** (8 µM) and control unmodified H-Telo DNA and RNA ONs **9–10** (8 µM) in Tris-HCl buffer (10 mM, pH 7.5) and in AOT RM containing either 50 mM NaCl or KCl was recorded from 200 to 350 nm on a J-815 CD spectropolarimeter (Jasco, USA) using 1 nm bandwidth at 20 °C. Each CD profile is an average of three scans collected at a scan speed of 100 nm min^{-1}. CD measurements were performed in duplicate and all spectra were corrected using an appropriate blank solution in the absence of ONs.

2.4.7 Thermal melting analysis

Fluorescently modified H-Telo DNA and RNA ONs **5–8** (1 µM), and control unmodified H-Telo DNA and RNA ONs **9,10** (1 µM) were annealed by heating at 90 °C for 3 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 50 mM KCl or 50 mM NaCl. The samples were cooled to RT and kept in an ice bath for at least 1 h. Thermal melting analysis was performed using Cary 300Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 295 nm.

2.4.8 Fluorescence binding assay: PDS binding to H-Telo DNA and RNA ONs 5–8 in aqueous buffer and AOT RM

The GQ structure of H-Telo DNA and RNA ONs **5–8** was formed by annealing the ONs in Tris-HCl buffer (pH 7.5) containing 50 mM NaCl at 90 °C for 3 min. The samples were cooled to RT and kept at ~4 °C for overnight.

2.4.8.1 Binding studies in aqueous buffer

A series of DNA ON samples (0.28 µM) in Tris-HCl buffer (10 mM, 50 mM NaCl) containing increasing concentration of pyridostatin (PDS, 4 nm to 4 µM) was prepared and incubated at RT for 30 min. Samples were excited at 322 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. A series of RNA ON samples (0.26 µM) in Tris-

HCl buffer (10 mM, 50 mM NaCl) containing increasing concentration of pyridostatin (PDS, 4 nm to 3 µM) was prepared and incubated at RT for 30 min. Samples were excited at 322 nm with an excitation and emission slit width of 6 nm and 8 nm, respectively. Fluorescence experiments were performed in duplicate in micro fluorescence cuvette at 20 °C. Appropriate blank in absence of ONs, but containing respective concentration of the ligand, was subtracted from the individual spectrum.

2.4.8.2 Binding studies in AOT RM

A series of stock solutions of H-Telo DNA ONs (3.8 µM) was prepared in Tris-HCl buffer (pH 7.5, 50 mM NaCl) containing increasing concentration of PDS. The samples were incubated at RT for 30 min. Individual DNA-PDS stock solution (36 µL) was added to AOT RM (450 µL) such that the final concentration of the ON, AOT and w_0 was 0.28 µM, 200 mM and 20, respectively. Final volume of the sample was 486 µL and the concentration range of PDS was between 4 nm to 4 µM. A series of stock solutions of H-Telo RNA ONs (3.5 µM) was prepared in Tris-HCl buffer (pH 7.5, 50 mM NaCl) containing increasing concentration of PDS. The samples were incubated at RT for 30 min. Individual RNA-PDS stock solution (18 µL) was added to AOT RM (225 µL) such that the final concentration of the ON, AOT and w_0 was 0.26 µM, 200 mM and 20, respectively. Final volume of the sample was 243 µL and the concentration range of PDS was between 4 nm to 3 µM. The samples were sonicated for 30 s and equilibrated at RT for 3 h. Fluorescence spectra were recorded as above.

From the dose-dependent quenching curves the apparent dissociation constants (K_d) for the binding of PDS to H-Telo DNA and RNA ONs **5**–**8** in aqueous buffer and AOT RM were determined by fitting normalized fluorescence intensity (F_N) versus log of PDS concentration plot to Hill equation (Origin 8.5).[54, 40]

$$F_N = \frac{F_i - F_s}{F_0 - F_s}$$

F_i is the fluorescence intensity at each titration point. F_0 and F_s are the fluorescence intensity in the absence of ligand (L) and at saturation, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$F_N = F_0 + (F_s - F_0)\left(\frac{[L]^n}{[K_d]^n + [L]^n}\right)$$

2.5 References

1. (a) Meyn, J.; Ratliff, R. L.; Moyzis, R. K. *Proc. Natl. Acad. Sci. U. S. A.* **1989***, 86,* 7049–7053. (b) Harris, M. L.; Merrick, C. J. *PLoS Pathog* **2015**, *11*, e1004562.
2. Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
3. Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev*. **2011**, *40*, 5867–5892.
4. Bugaut, A.; Balasbramanian, S. *Nucleic Acids Res.* **2012**, *40*, 4727–4741.
5. Rhodes, D.; Lipps, H. J. *Nucleic Acids. Res.* **2015***, 43*, 8627–8637.
6. Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429–7455.
7. Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discov.* **2011**, *10*, 261–275.
8. Ohnmacht, S. A.; Neidle, S. *Bioorg. Med. Chem. Lett*. **2014**, *24*, 2602–2612.
9. Zhang, S.; Wu, Y.; Zhang, W. *ChemMedChem* **2014**, *9*, 899–911.
10. (a) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. *Methods* **2012**, *57*, 64–75. (b) Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11–14. (c) Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. *ChemBioChem* **2013**, *14*, 540–558. (d) Neidle, S. *J. Med. Chem.* **2016**, *59*, 5987–6011.
11. (a) Rache, A. D.; Mergny, J.-L. *Biochimie* **2015**, *115*, 194–202. (b) Darby, R. A. J.; Sollogoub, M.; McKeen, C.; Brown, L.; Risitano, A.; Brown, N.; Barton, C.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2002***, 30*, e39.
12. (a) Largy, E.; Granzhan, A.; Hamon, F.; Verga, D.; Teulade-Fichou, M.-P. *Top. Curr. Chem.* **2013***, 330*, 111–177. (b) Bhasikuttan, A. C.; Mohanty, J. *Chem. Commun.* **2015**, *51*, 7581–7597. (c) Laguerre, A.; Stefan, L.; Larrouy, M.; Genest, D.; Novotna, J.; Pirrotta, M.; Monchaud, D. *J. Am. Chem. Soc.* **2014***, 136*, 12406–12414. (d) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. *J. Am. Chem. Soc.* **2013***, 135*, 367–376.
13. (a) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem. Commun.* **2006***, 401–402. (b) Gray, R. D.; Petraccone, L.; Trent, J. O.; Chaires, J. B. *Biochemistry* **2010**, *49*, 179–194. (c) Dumas, A.; Luedtke, N. W. *J. Am. Chem. Soc.* **2010**, *132*, 18004–18007. (d) Nadler, A.; Strohmeier, J.; Diederichsen, U. *Angew. Chem. Int. Ed.* **2011**, *50*, 5392–5396. (e) Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A. *ACS Chem. Biol*. **2015**, *10*, 1311–1318.
14. (a) Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14579–14584. (b) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186. (c) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh Jr, R. M.; Sen, D.; Lansdrop, P. M. *Nucleic Acids Res.* **2013**, *42*, 860–869. (d) Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M.; Monchaud, Y. D. *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525. (e) Chen, S.-B.; Hu, M.-H.; Liu, G.-C.; Wang, J.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Tan, J.-H. *J. Am. Chem. Soc.* **2016**, *138*, 10382–10385. (f) Shivalingam, A.; Izquierdo, M. A.; Marois, A. L.; Vyšniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. *Nat. Commun.* **2015**, *6*, 8178.
15. (a) Petraccone, L.; Pagano, B.; Giancola, C. *Methods* **2012**, *57*, 76–83. (b) Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. *Biochemistry* **2010**, *49*, 4554–4563.

16. (a) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282. (b) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.

17. Chen, Z.; Zheng, K.-W.; Hao, Y.-H.; Tan, Z. *J. Am. Chem. Soc.* **2009**, *131*, 10430–10438.

18. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. *J. Am. Chem. Soc.* **2010**, *132*, 17105–17107.

19. Miyoshi, D.; Fujimoto, T.; Sugimoto, N. *Top. Curr. Chem.* **2012**, *330*, 87–110.

20. Lannan, F. M.; Mamajanov, I.; Hud, N. V. *J. Am. Chem. Soc.* **2012**, *134*, 15324–15330.

21. Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.

22. Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2013**, *41*, 7934–7946.

23. Hänsel, R.; Löhr, F.; Trantirek, L.; Dötsch, V. *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.

24. (a) Xu, Y.; Kaminaga, K.; Komiyama, M. *J. Am. Chem. Soc.* **2008**, *130*, 11179–11184. (b) Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570–2578.

25. Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantirek, L.; Dötsch, V. *Angew. Chem. Int. Ed.* **2014**, *53*, 10300–10314.

26. Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. *Chem. Sci.* **2015**, *6*, 3314–3320.

27. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y.; *Nucleic Acids Res* **2017**, 45, 5501–5511.

28. Levinger, N. E. *Science* **2002**, *298*, 1722–1723.

29. Tanpure, A. A.; Srivatsan, S. G. *ChemBioChem* **2012**, *13*, 2392–2399.

30. Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J.* **2011**, *17*, 12820–12827.

31. De, T. K.; Maitra, A. *Adv. Colloid Interface Sci.* **1995**, *59*, 95–193.

32. (a) Levinger, N. E.; Swafford, L. A. *Annu. Rev. Phys. Chem.* **2009**, *60*, 385–406. (b) Karukstis, K. K.; Frazier, A. A.; Martula, D. S.; Whiles, J. A. *J. Phys. Chem.* **1996**, *100*, 11133–11138. (c) Hasegawa, M.; Sugimura, T.; Suzaki, Y.; Shindo, Y. *J. Phys. Chem.* **1994**, *98*, 2120–2124. (d) Mukherjee, T. K.; Panda, D.; Datta, A. *J. Phys. Chem. B* **2005**, *109*, 18895–18901.

33. (a) O'Connor, C. J.; Lomax, T. D.; Ramage, R. E. *Adv. Colloid Interface Sci.* **1984**, *20*, 21–97. (b) Fiori, S.; Renner, C.; Cramer, J.; Pegoraro, S.; L. Moroder, L. *J. Mol. Biol.* **1999**, 291, 163–175. (c) Nicot, C.; Vacher, M.; Vincent, M.; Galla, J.; Waks, M. *Biochemistry* **1985**, *24*, 7024–7032. (d) Yeung, P. S.-W.; Eskici, G.; Axelsen, P. H. *Biochim. Biophys. Acta Biomembr.* **2013**, 1828, 2314–2318. (e) Yeung, P. S.-W.; Axelsen, P. H. *J. Am. Chem. Soc.* **2012**, 134, 6061–6063.

34. Pawar, M. G.; Srivatsan, S. G.; *J. Phys. Chem. B* **2013**, *117*, 14273–14282.

35. Workman, H.; Flynn, P. F. *J. Am. Chem. Soc.* **2009**, *131*, 3806–3807.

36. Park, L.-C.; Maruyama, T.; Goto, M. *Analyst* **2003**, *128*, 161–165.

37. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Chem. Commun.* **2010**, *46*, 1700–1702.

38. Pramanik, S.; Nagatoishi, S.; Sugimoto, N. *Chem. Commun.* **2012**, *48*, 4815–4817.

39. Ho, M.-C.; Chang, C.-W. *RSC Adv.* **2014**, *4*, 20531–20534.

40. Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149.

41. Bru, R.; Sánchez-Ferrer, A.; García-Carmona, F. *Biochem. J.* **1995**, *310*, 721–739.

42. Noé, M. S.; Sinkeldam, R. W.; Tor, Y. *J. Org. Chem.* **2013**, *78*, 8123–8128.

43. Hazra, P.; Sarkar, N. *Chem. Phy. Lett.* **2001**, *342*, 303–311.

44. Sengupta, B.; Guharay, J.; Sengupta, P. K. *Spectrochim. Acta A* **2000**, *56*, 1433–1441.

45. (a) Schoeftner, S.; Blasco, M. A. *Nat. Cell Biol.* **2008**, *10*, 228–236. (b) Rippe, K.; Luke, B. *Nat. Struct. Mol. Biol.* **2015**, *22*, 853–858. (c) Deng, Z.; Norseen, J.; Wiedmer, A.; Riethman, H.; Lieberman, P. *Mol Cell.* **2009**, *35*, 403–413.

46. Biffi, G.; Tannahill, D.; Balasubramanian, S. *J. Am. Chem. Soc.* **2012**, *134*, 11974–11976.

47. (a) Tran, P. L. T.; Mergny, J.-L.; Alberti, P. *Nucleic Acids Res.* **2011**, *39*, 3282–3294. (b) Rachwal, P. A.; Fox, K. R. *Methods* **2007**, *43*, 291–301.

48. Wang, Z.-F.; Li, M.-H.; Hsu, S.-T. D.; Chang, T.-C. *Nucleic Acids Res.* **2014**, *42*, 4723–4733.

49. Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2009**, *10*, 1389–1398.

50. Collie, G. W.; Campbell, N. H.; Neidle, S. *Nucleic Acids Res.* **2015**, *43*, 4785–4799.

51. Le, D. D.; Antonio, M. D.; Chan, L. K. M.; Balasubramanian, S. *Chem Commun.* **2015**, *51*, 8048–8050.

52. Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759.

53. Cservenyi, T. Z.; Riesen, A. J. V.; Berger, F. D.; Desoky, A.; Manderville, R. A. *ACS Chem. Biol.* **2016**, *11*, 2576–2582.

54. (a) Tam, V. K.; Kwong, D.; Tor, Y.; *J. Am. Chem. Soc.* **2007**, *129*, 3257–3266. (b) Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. *Angew. Chem. Int. Ed.* **2004***, 43,* 3177–3182.

Chapter 3

# A Dual-App Nucleoside Probe Provides Structural Insights into the Human Telomeric Overhang in Live Cells
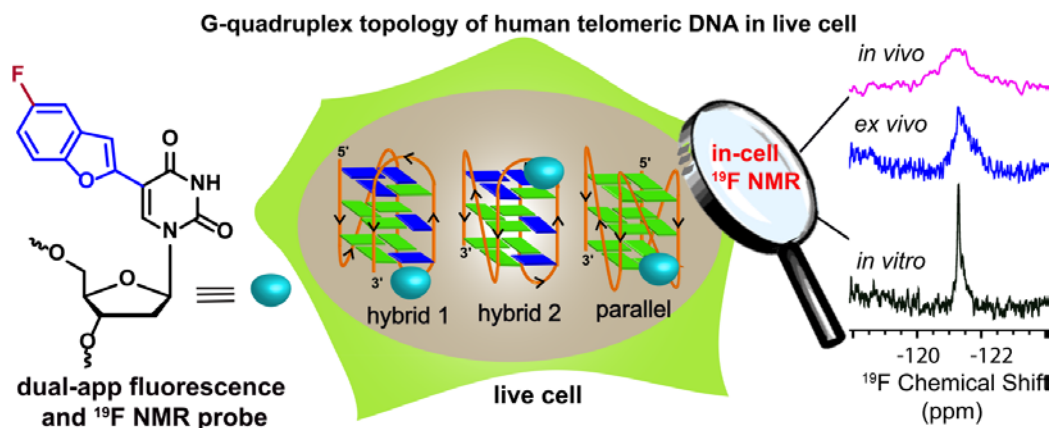
3.1 Introduction

Non-canonical four-stranded GQ-forming sequences are prevalent in the telomeres and in the promoter DNA and untranslated regions of mRNA of several genes that are known to cause cancer.[1,2] The location and conservation of putative GQ-forming sequences across various eukaryotes,[3,4] and extensive biophysical investigations point out that GQ is an important structural element in maintaining the stability of genome and in the regulation of core cellular processes like replication, transcription and translation.[1,5,6] Further, small molecule ligands that stabilize GQ structure have provided compelling evidence linking expression levels of proto-oncogenes and GQ structure.[7-11] Hence, controlling the ensuing function of GQ structure by using ligands is viewed as an alternative therapeutic strategy for cancer and ageing related diseases.[12,13] Despite a flurry of interest in this direction,[12-16] discovery of clinically viable GQ-binders remains a major challenge.[17,18] This is because GQ exhibits a high degree of structural polymorphism (e.g., parallel, antiparallel, hybrid-type mixed parallel-antiparallel stranded GQs), which depends on the sequence, ionic conditions, confinement and molecular crowding,[19-23] and most GQ binders/probes poorly differentiate different GQs topologies.[24] Hence, recent studies are directed towards developing biophysical platforms to understand the topology adopted by a G-rich sequence inside the cell and identify binders that specifically target a GQ motif amongst other such motifs in the genome.[25]

Circular dichroism (CD), fluorescence and NMR techniques are commonly used to study the structure, dynamics and binding affinities of GQs in synthetic models mimicking the crowded and confined environment of the cell.[26-29] One of the well studied system is the human telomeric (H-Telo) DNA repeat $(TTAGGG)_n$. In an aqueous buffer containing K^+ ions H-Telo DNA oligonucleotide (ON) repeat forms multiple GQ structures,[30] whereas in the presence of crowding agents like PEG it adopts a parallel conformation, which exhibits slower folding dynamics and reduced ligand-binding affinities.[28,31] In contrast, by using optical tweezers it was identified that H-Telo DNA in a confined environment of DNA origami nanocages formed a GQ with higher stability and faster folding rates.[32] While these studies provide valuable structural information, non-native conditions bias the ONs to adopt a certain GQ structure as in the case of PEG due to its dehydrating effect.[26,28] Alternatively, electron microscopy,[33] and GQ-specific antibodies[34-37] and fluorescent ligands[38-41] have facilitated the visualization of GQ structures in different cell types. Although these methods

are quantitative, they do not provide the much-needed information on the topology of GQs. Very recently, by using a combination of antibodies D1 and BG4, which specifically bind to parallel GQs and "all types" of GQ structures, respectively, it was identified that ~77% of H-Telo repeats exist in parallel conformation and the remaining presumed to adopt other GQ structures.[42] EPR analysis of frozen *Xenopus laevis* oocytes microinjected with nitroxide radical-labeled H-Telo DNA ON repeat also showed the presence of a mixture of parallel- and antiparallel-stranded GQs.[43] However, an important limitation of EPR and immunofluorescence GQ detection methods is that they are restricted to frozen or fixed and permeabilized cells wherein morphological integrity is compromised.[38a] On the contrary, NMR analysis in oocyte egg extract suggested that parallel conformation is not the preferred conformation of the telomeric repeat.[44] However, severe broadening of NMR signal hampered the structural elucidation in live oocytes.[45]

Amidst these conflicting results and limitations in currently available tools, it is hypothesized that understanding the structure of individual GQ-forming sequences in native cellular environment and devising discovery platforms to identify topology-specific GQ binders are of high priority. This, we envisioned, can be achieved by developing a GQ-sensing biophysical platform, which (i) is compatible to both cell-free and in-cell analysis, (ii) reports the formation as well as distinguishes different GQ topologies in cell-free and cellular systems, and (iii) supports discovery assays to identify topology-specific GQ binders. Such a platform would not only allow the profiling of the topology of various GQ-forming sequences in cells but also would have a profound impact on harnessing the potential of GQ-directed therapeutic strategies. Here, we describe an innovative approach to investigate GQ structures and their ligand binding *in vitro* and in live cells by using a conformation-sensitive dual-purpose nucleoside analog probe (Figure 1). The nucleoside probe is composed of a microenvironment-sensitive fluorophore and ^{19}F NMR label, which is derived by attaching fluorobenzofuran at the 5-position of 2′-deoxyuridine. This minimally perturbing nucleoside, when incorporated into the loop region of H-Telo DNA ON repeat, serves as a common probe to detect different GQ topologies and quantify topology-specific binding of ligands to GQ structures by fluorescence and NMR techniques. Importantly, through systematic ^{19}F NMR studies in intraoocyte buffer, *Xenopus laevis* oocyte lysate, egg extract and live oocytes we successfully determined the GQ topologies adopted by the H-Telo overhang in live cells.
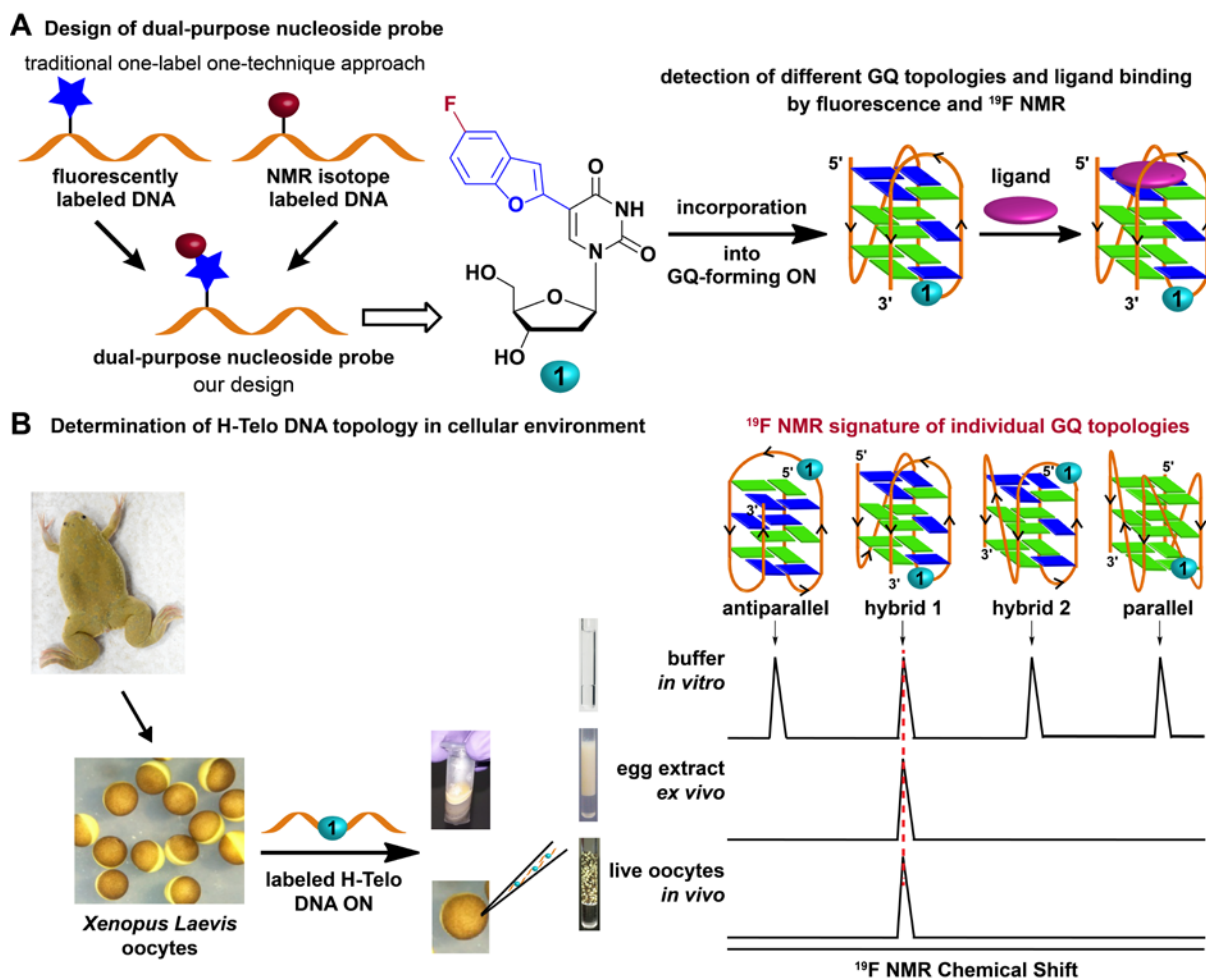
Figure 1. Probe design and determination of GQ topologies in live cells. (**A**) The dual-purpose nucleoside probe **1** is designed by combining the environment-sensitivity of a fluorescent 5-benzofuran-modified nucleoside analog and high responsiveness of ^{19}F NMR signal to its local environment. Nucleoside **1**, incorporated into a G-rich ON sequence, reports and distinguishes different GQ conformations and facilitates the estimation of topology-specific binding of ligands to different GQs by fluorescence and NMR spectroscopy techniques. (**B**) A schematic diagram showing the experimental design for the determination of GQ structure in cells by using *Xenopus laevis* oocyte as the model. The sensitivity of ^{19}F-labeled nucleoside **1** to subtle differences in the conformation produces distinct NMR signature for different GQ conformations. A comparison of signatures *in vitro* (buffer), *ex vivo* (egg extract) and *in vivo* (live oocytes) conditions provides insights into the GQ topologies adopted by H-Telo DNA overhang in cells.

3. 2 Results and Discussion

3.2.1 Design and synthesis of dual-purpose nucleoside probe

We recently developed a conformation-sensitive fluorescent nucleoside probe by conjugating benzofuran moiety at the 5-position of uracil.[46] The nucleoside analog is structurally minimally invasive, and importantly, serves as an excellent GQ sensor, wherein it

photophysically discriminates duplex and various GQ forms of H-Telo DNA ON repeat. While being highly sensitive to subtle differences in conformations, its fluorescence properties, particularly excitation maximum in the UV region (λ_{max} = 330 nm), precluded its use in cell-based analysis. So, in order to retain the high conformation sensitivity of the fluorophore and still determine the structure of GQ in cells, we came up with the idea of introducing a ^{19}F atom as an NMR label to expand the proficiency of benzofuran-modified nucleoside analog in probing the structure and recognition of GQ structures in both cell-free and cellular environments (Figure 1). The choice of ^{19}F isotope label was based on the following reasons.[47] It is 100% naturally abundant and is highly sensitive to changes in its local environment. Its signal is highly dispersed as compared to proton signal, and importantly, absence of endogenous ^{19}F label eliminates the background signal from the cells. Further, ^{19}F NMR spectroscopy, which has emerged as a powerful tool in discovering protein binders, has recently been used to study non-canonical nucleic acid structures.[48,49] Akin to antibodies, these studies predicted the formation of GQs, but did not provide information on the topology as these systems were not designed to distinguish different GQ topologies.

Based on the responsiveness of benzofuran-modified nucleoside analog and useful properties of ^{19}F isotope, a dual-purpose nucleoside probe (**1**) was assembled by attaching fluorobenzofuran moiety at the 5-position of uracil (Figure 1). 5-Fluorobenzofuran was first stanylated, which upon cross-coupling with 5-iodo-2′-deoxyuridine in the presence of a palladium catalyst gave 5-fluorobenzofuran-2′-deoxyuridine **1** (Scheme 1). Corresponding phosphoramidite **2** was synthesized as depicted in scheme 1.

An important aspect of this design is that the fluorophore and ^{19}F atom are intentionally integrated into the same electronic system so that the effect of microenvironment will be similar on both the labels. Hence, the signature of individual GQ structures obtained using fluorescence and NMR techniques could be efficiently correlated with one another.
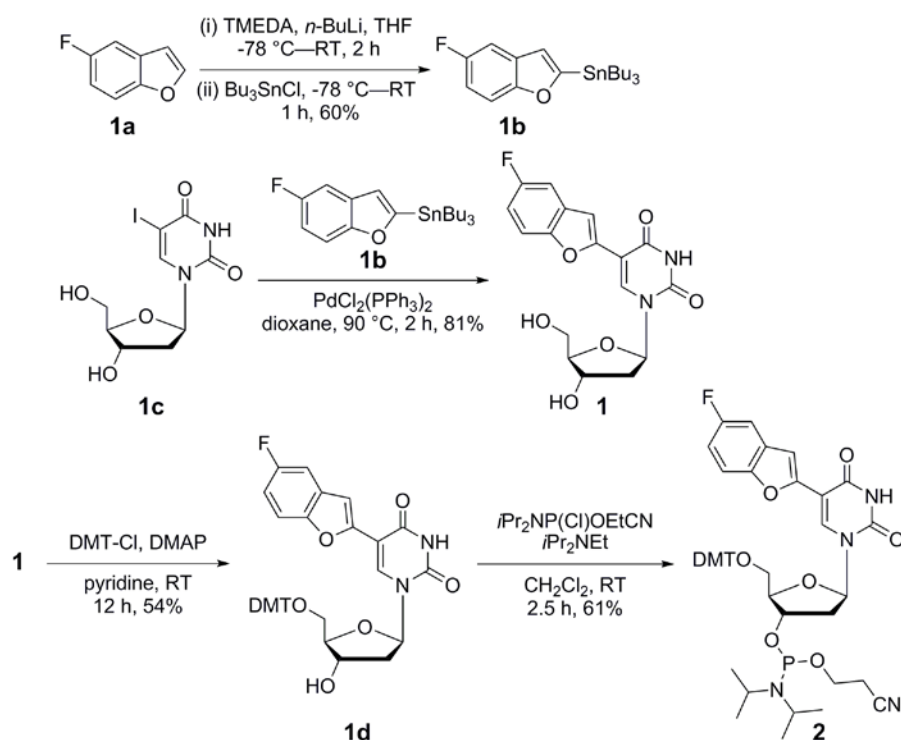
Scheme 1. Synthesis of 5-fluorobenzofuran-2′-deoxyuridine **1** and corresponding phosphoramidite **2**. TMEDA: *N,N,N′,N′*-tetramethylethylenediamine; THF: tetrahydrofuran; DMT-Cl = 4,4′-dimethoxytrityl chloride, DMAP = 4-dimethylaminopyridine.

### 3.2.2 Nucleoside 1 is highly sensitive to its microenvironment

Ground-state electronic spectrum of nucleoside **1** in solvents of different polarity and viscosity revealed only minor changes in absorption maxima (Figure 2A, Table 1). However, excited-state properties like emission maximum, quantum yield and lifetime were significantly affected. The nucleoside in water exhibited a strong emission band centered at 437 nm with a quantum yield and lifetime of 0.11 and 0.84 ns, respectively (Figure 2A, Figure 3 and Table 1,). In less polar solvents like methanol and dioxane, the emission band was considerably blue shifted and the quantum yield and lifetime were found to be significantly lower as compared to in water. A plot of Stokes shift in solvents of different polarity *versus* $E_T(30)$ (Reichardt's microscopic solvent polarity parameter) gave nearly a linear correlation, which further confirmed the responsiveness of the nucleoside to changes in microenvironment (Figure 2B). By design the nucleoside analog contains a rotatable aryl-aryl bond between fluorobenzofuran and uracil rings. This could invoke conformation sensitivity as the relative conformation of two rings will influence π-conjugation, and hence, the

fluorescence properties. As the viscosity of the medium was increased from water to ethylene glycol to glycerol, the nucleoside exhibited a progressive increase in quantum yield and lifetime due to rigidification of the fluorophore in a more viscous medium (Figure 2C, Table 1 and Figure 3). Further, significantly higher fluorescence anisotropy displayed by the nucleoside in a viscous medium confirmed the conformation sensitivity of the system (Table 1).
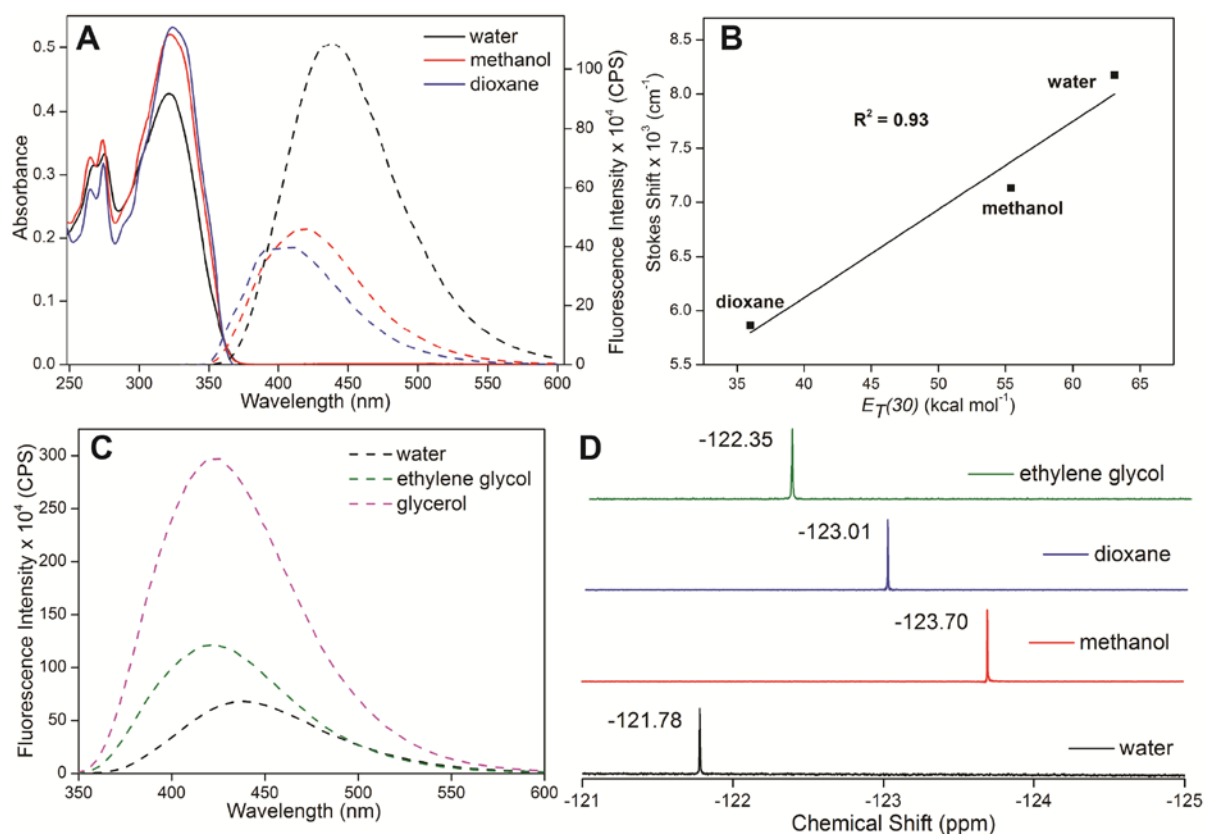
Figure 2. Effect of microenvironment on the fluorescence and ^{19}F NMR signal of nucleoside **1**. (**A**) UV absorption (25 µM, solid lines) and fluorescence (5 µM, dashed line) spectra of nucleoside **1** in solvents of different polarity. In fluorescence study, samples were excited at respective lowest energy absorption maximum (Table 1) with excitation and emission slit widths of 2 nm and 4 nm, respectively. (**B**) Stokes shift *versus* $E_T(30)$ plot of nucleoside **1** in solvents of different polarity. (**C**) Fluorescence spectra (5 µM, dashed line) of nucleoside **1** in solvents of different viscosity. Excitation and emission slit widths were kept at 2 nm and 3 nm, respectively. (**D**) ^{19}F NMR spectra of nucleosides **1** in solvents of different polarity and viscosity. All samples contained 15% d_6-DMSO and each spectrum was referenced relative to an external standard (Trifluorotoluene (TFT) -63.72 ppm).
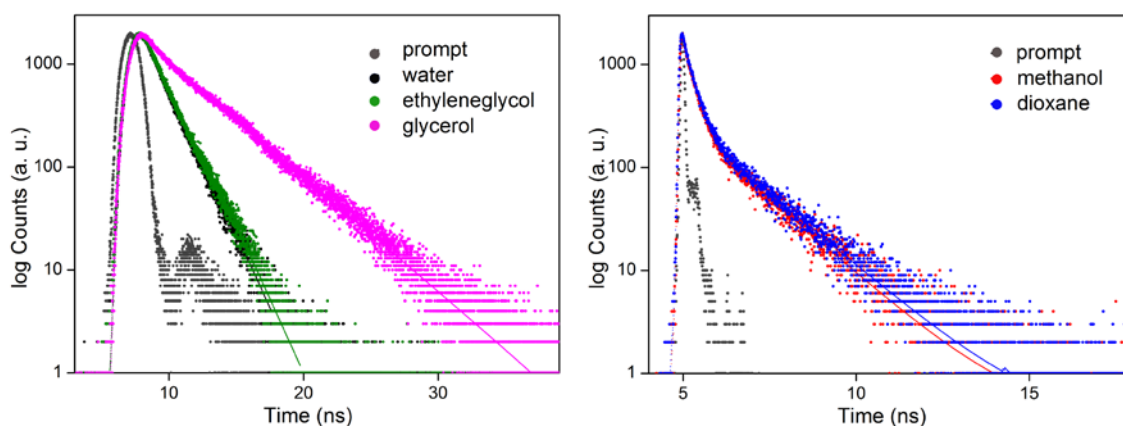
Figure 3. Excited-state decay profile of nucleoside **1** in solvents of different polarity and viscosity. Instrument response (prompt) is shown in gray dots. Curve fits are shown in solid lines.

Akin to fluorescence, the nucleoside exhibited a well resolved ^{19}F chemical shift in solvents of different polarity and viscosity (Figure 2D). Depending on the solvent polarity, the dipole of a fluorinated molecule can polarize the surrounding solvent molecules or align itself with the dipole of the solvent molecules. This could either shield or deshield the fluorinated molecule and alter the ^{19}F chemical shift.[50a] Similarly, changes in solvent viscosity can alter the relative conformation between the fluorobenzofuran and uracil rings, which can also lead to shielding or deshielding of the ^{19}F label.[50b] Hence, distinct NMR signal displayed by nucleoside **1** in different solvents is due to a combination of polarity and viscosity effects. Collectively, these studies indicate that environment-sensitive nucleoside **1** could serve as a two-channel probe to study nucleic acids by fluorescence and ^{19}F NMR techniques.

Table 1. Photophysical properties of nucleoside **1** in different microenvironments.

| Solvent	$\lambda_{max}{}^a$ (nm)	λ_{em} (nm)	Stokes Shift (cm^{-1})	$I_{rel}{}^b$	Φ^c	$\tau_{av}{}^c$ (ns)	r^c
water	322	437	8.17 x 10^3	1.00	0.11	0.84	0.03
methanol	322	418	7.13 x 10^3	0.43	0.04	0.33	n.d.
dioxane	324	400	5.86 x 10^3	0.36	0.03	0.31	n.d.
ethylene glycol	325	420	6.96 x 10^3	1.79	0.20	0.94	0.20
glycerol	326	424	7.09 x 10^3	4.39	0.52	2.36	0.34

^{a}Lowest absorption energy maximum is given. ^{b}Relative emission intensity at λ_{em} is given with respect to intensity in water. ^{c}Standard deviation for quantum yield (Φ), average lifetime (τ_{av}) and anisotropy (r) in different solvents is ≤ 0.005, ≤ 0.02 ns and ≤ 0.003 respectively. n.d. = not determined.

### 3.2.3 Incorporation of nucleoside 1 into H-Telo GQs

The nucleoside probe was incorporated into the loop regions formed by TTA residues for two reasons—(i) The conformation of loop residues is distinctly different in different H-Telo GQ structures,[19] and (ii) modifications on the G-tetrads could potentially affect the formation as well as the stability of GQ.[51] One of the T residues in each of the three loops was replaced with the dual-purpose nucleoside analog **1** by solid-phase ON synthesis protocol using phosphoramidite substrate **2** to produce H-Telo DNA ONs **3–5** (Figure 4, Scheme 1). HPLC and MALDI-TOF mass analyses of PAGE purified ONs confirmed the purity and integrity of the modified ONs (Figure 5, Figure 6 and Table 2).

3 5' AGGGTTAGGG **1** TAGGGTTAGGG 3'

4 5' AGGG **1** TAGGGTTAGGGTTAGGG 3'

5 5' AGGGTTAGGGTTAGGG **1** TAGGG 3'

6 5' AGGGTTAGGGTTAGGGTTAGGG 3'

7 5' CCCTAACCCTAACCCTAACCCT 3'

Figure 4. Sequence of fluorobenzofuran-labeled H-Telo DNA ONs **3–5**. One of the T residues of loop 2 (**3**), loop 1 (**4**) and loop 3 (**5**) was replaced with nucleoside **1**. Sequence of control unmodified H-Telo DNA ON **6** and complementary ON **7** is shown.
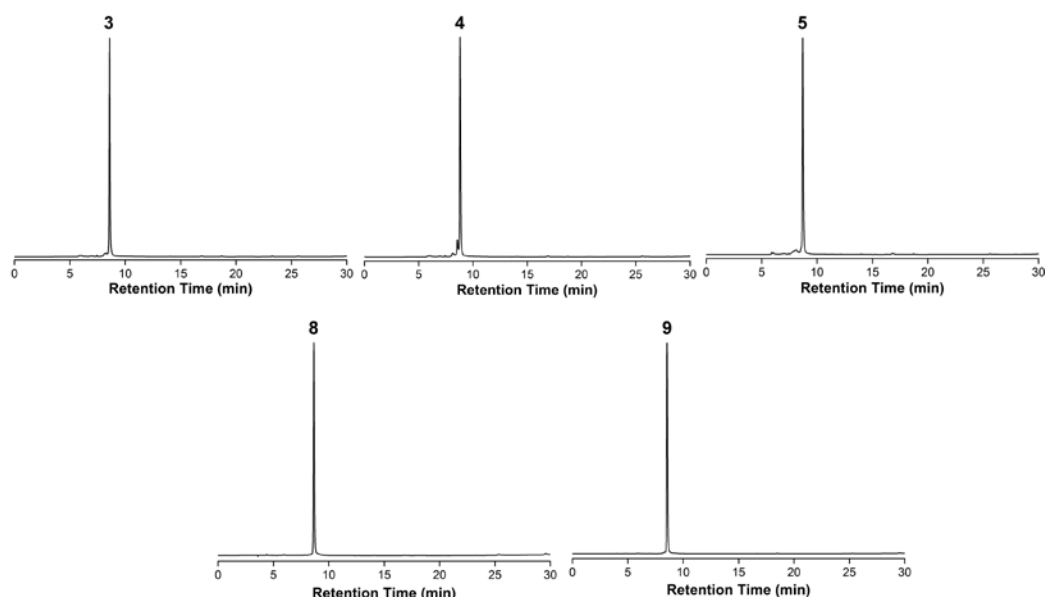
Figure 5. HPLC chromatogram of PAGE purified ONs **3–5**, **8** and **9** at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0−100% B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).
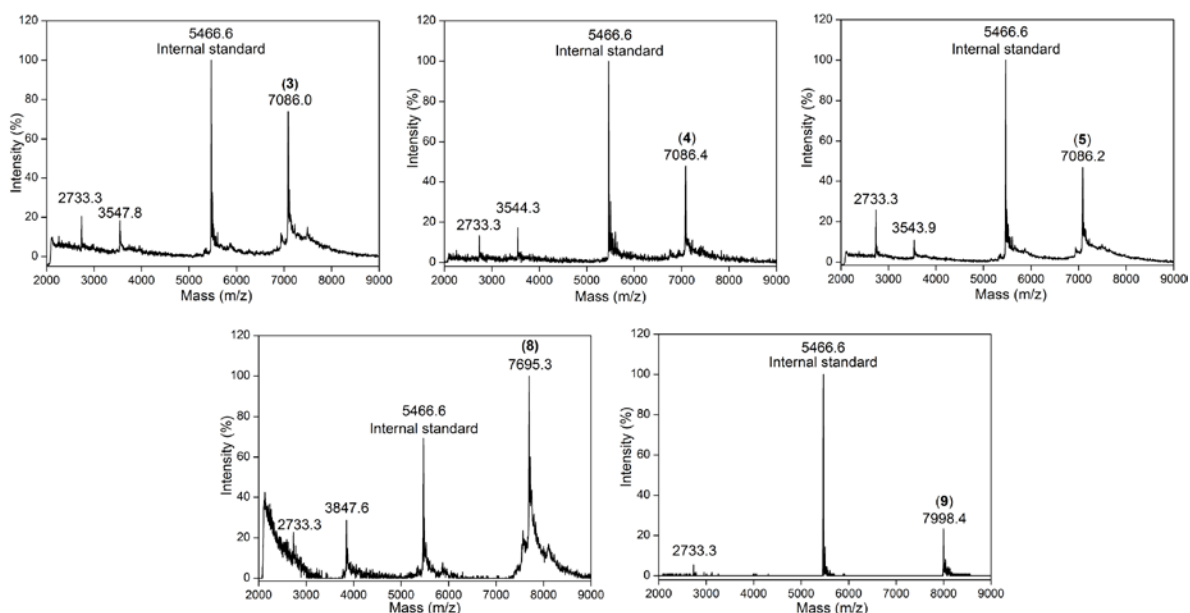
Figure 6. MALDI-TOF spectrum of modified H-Telo DNA ONs calibrated relative to the +1 and +2 ions of an internal 18-mer DNA ON standard (m/z for +1 and +2 ions are 5466.6 and 2733.3, respectively). See Table 2 for details.

Table 2. Molar extinction coefficient and mass of modified H-Telo DNA ONs.

H-Telo DNA ON	$\varepsilon_{260}{}^{a}$ $(M^{-1}cm^{-1})$	Calculated mass	Observed mass
3	230.4×10^3	7086.6	7086.0
4	230.4×10^3	7086.6	7086.4
5	230.4×10^3	7086.6	7086.2
8	246.2×10^3	7695.0	7695.3
9	255.0×10^3	7999.2	7998.4

^{a}Molar absorption coefficient ε_{260} of the modified ONs was determined by using OligoAnalyzer 3.1, The extinction coefficient of nucleoside **1** ($\varepsilon_{260} = 10.31 \times 10^3$ $M^{-1}cm^{-1}$) was used in the place of thymidine

The effect of modification on the formation and stability of GQ structure was examined by CD and thermal-melting experiments. Modified (**3**–**5**) and control unmodified (**6**) H-Telo DNA ONs were annealed in different ionic conditions using either potassium phosphate buffer containing 100 mM KCl or sodium phosphate buffer containing 100 mM NaCl. Both control and modified ONs in K^{+} ionic conditions showed similar CD profiles characteristic of hybrid type mixed parallel-antiparallel stranded GQ structures (a positive peak at ~290 nm and a shoulder at ~270 nm, Figure 7).[30] In the presence of NaCl, the ONs exhibited a positive peak at ~293 nm and a strong negative peak at ~263 nm characteristic of

an antiparallel GQ structure (Figure 7B).[30] UV-thermal melting analysis of the modified and native ONs in different ionic conditions gave a characteristic melting profile for the GQ structure at 295 nm with similar T_m values (Figure 8 and Table 3).[52] Consistent with the literature reports, these results clearly prove that the modified H-Telo DNA ONs form respective GQ structures in different ionic conditions and the fluorobenzofuran modification has only a minor impact on the GQ stability.
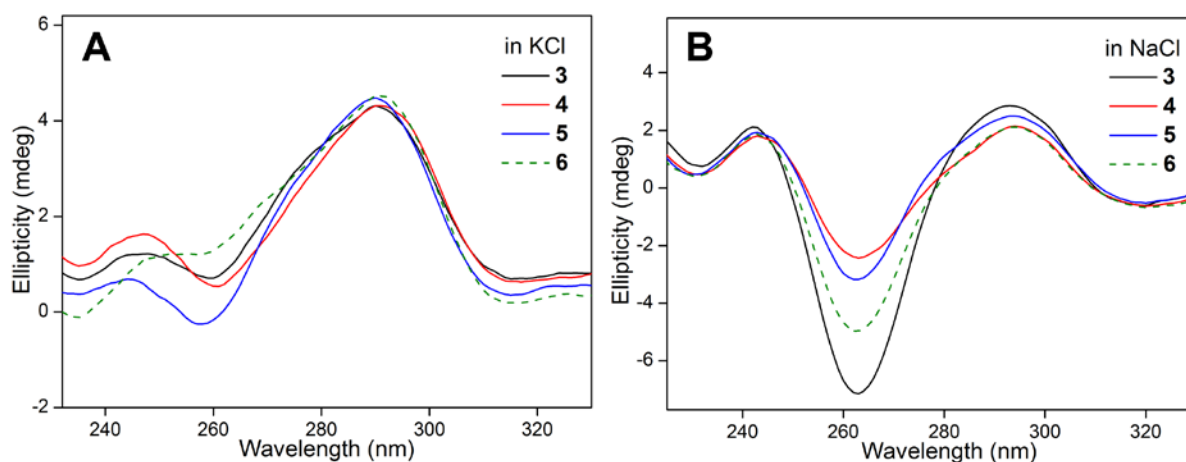
Figure 7. CD spectrum of 5-fluorobenzofuran-modified H-Telo DNA ONs **3**–**5** and unmodified control H-Telo DNA ON **6**. (**A**) In potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl and (**B**) in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl.
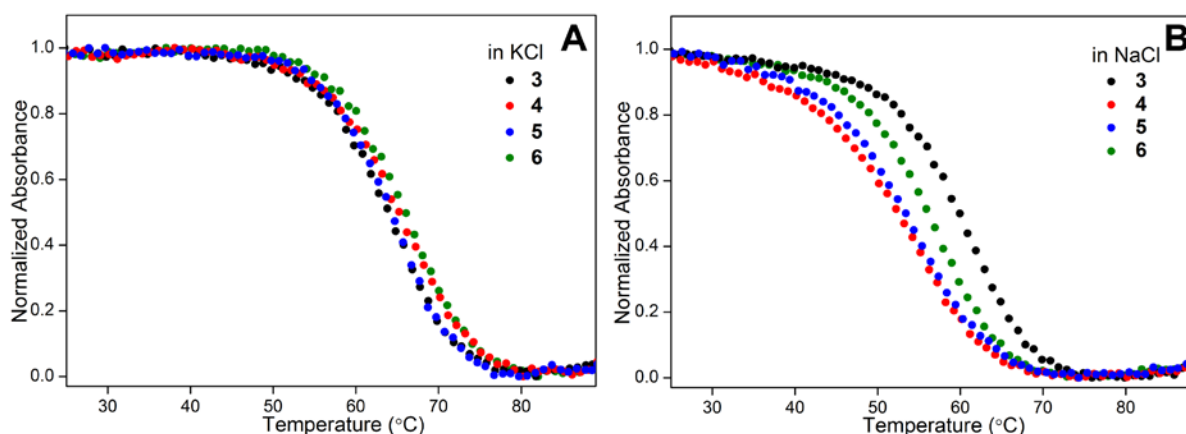
Figure 8. UV-thermal melting profile of 5-fluorobenzofuran-modified H-Telo DNA ONs **3**–**5** and unmodified H-Telo DNA ON **6**. (**A**) In potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl and (**B**) in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl. For T_m values see Table 3.

Table 3. T_m values of modified (**3**–**5**) and control unmodified (**6**) H-Telo DNA ONs.

H-Telo DNA ON	T_m (°C) in KCl	T_m (°C) in NaCl
3	66.5 ± 0.4	61.5 ± 0.4
4	65.8 ± 0.4	55.5 ± 0.5
5	65.5 ± 0.7	55.0 ± 0.4
control ON **6**	66.5 ± 0.7	57.8 ± 0.4

3.2.4 Fluorescence detection of H-Telo GQ topologies

Samples of labeled ONs, annealed into GQs and duplexes (**3•7**, **4•7** and **5•7**) in a buffer containing K^+ or Na^+ ions were excited at 330 nm. Steady-state fluorescence of duplexes in both the ionic conditions showed similar emission bands, which were very weak (Figure 9). The telomeric DNA ONs, which form multiple structures with hybrid-type GQs as the predominant conformation in K^+ conditions, displayed significantly higher fluorescence intensity (6–11 fold) as compared to respective perfect duplexes. Remarkably, in the presence Na^+ ions, which favour the antiparallel conformation, the ONs exhibited further enhancement in fluorescence intensity as compared to the hybrid-type GQs (Figure 9). Excited-state decay kinetic analysis showed that the antiparallel conformation has discernibly higher lifetime as compared to the hybrid type GQs formed in K^+ ionic conditions (Figure 10 and Table 4). It is important to mention here that changes in ionic conditions did not affect the fluorescence profile of the free nucleoside probe (Figure 11). Hence, the changes in fluorescence properties (intensity, quantum yield and lifetime) are due to the differences in the microenvironment of the nucleoside probe in different G-quadruplex conformations.
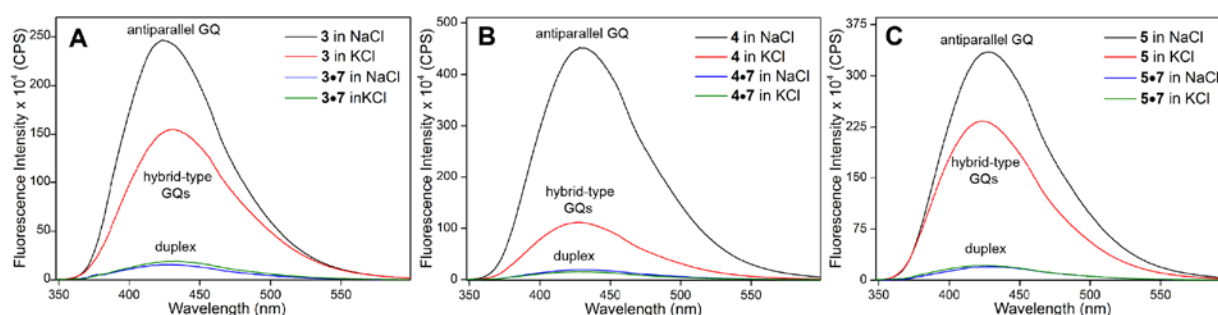
Figure 9. Fluorescence spectrum of GQs of ON **3**–**5** and the corresponding duplexes **3•7**, **4•7** and **5•7** in different ionic conditions. Samples (0.5 µM) were excited at 330 nm with excitation and emission slit widths of 4 nm and 6 nm, respectively.
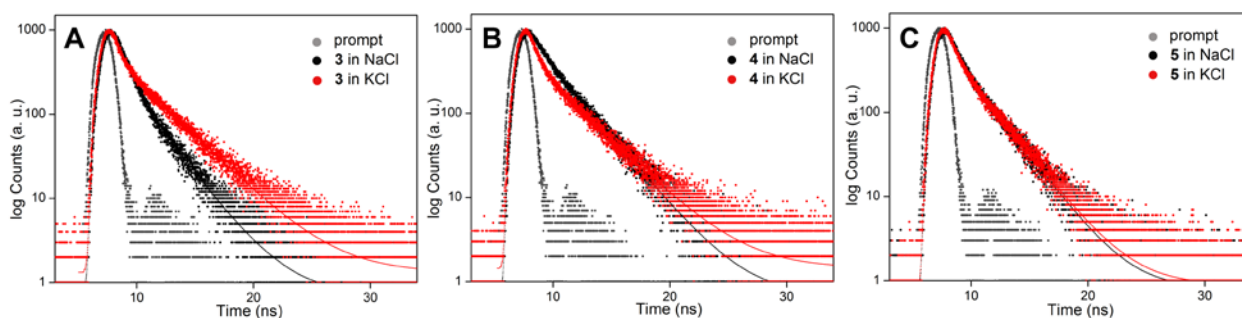
**Figure 10.** Time-resolved fluorescence spectrum of H-Telo DNA ONs (0.5 μM) (**A**) **3** (**B**) **4** and (**C**) **5** in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl or in potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl. Instrument response is shown in gray dots. Curve fits are shown in solid lines.

Table 4. Fluorescence properties of H-Telo DNA ONs **3–5** and respective duplexes in aqueous buffer

ON sample	λ_{em} (nm)	Φ^a	Φ_r^b	τ_{av}^c (ns)	ON sample	λ_{em} (nm)	Φ^a	Φ_r^b	τ_{av}^c (ns)	ON sample	λ_{em} (nm)	Φ^a	Φ_r^b	τ_{av}^c (ns)
3 in NaCl	424	0.053	27	1.09	**4** in NaCl	428	0.093	93	1.40	**5** in NaCl	428	0.062	31	0.89
3 in KCl	429	0.030	15	0.60	**4** in KCl	428	0.022	11	0.79	**5** in KCl	423	0.043	22	0.73
3•7 in NaCl	426	0.002	1	n.d.	**4•7** in NaCl	426	0.001	1	n.d.	**5•7** in NaCl	428	0.002	1	n.d.
3•7 in KCl	432	0.002	1	n.d.	**4•7** in KCl	428	0.002	1	n.d.	**5•7** in KCl	426	0.002	1	n.d.

[a]Standard deviation for quantum yield (Φ) of GQs is ≤ 0.006 and that of duplexes is ≤ 0.0002. [b]Φ_r = relative quantum yield with respect to the quantum yield of the duplex in the respective ionic conditions. [b]Standard deviation for lifetime (τ_{av}) is ≤ 0.06 ns. n.d. = not determined. Reliable lifetime values could not be obtained for duplexes as they were very weakly fluorescent.
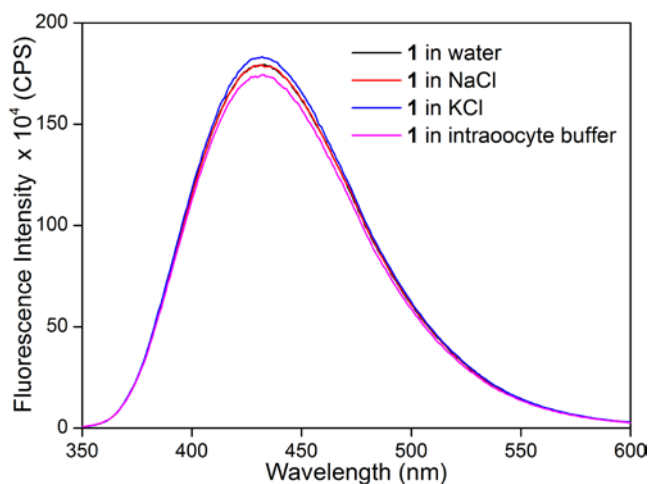
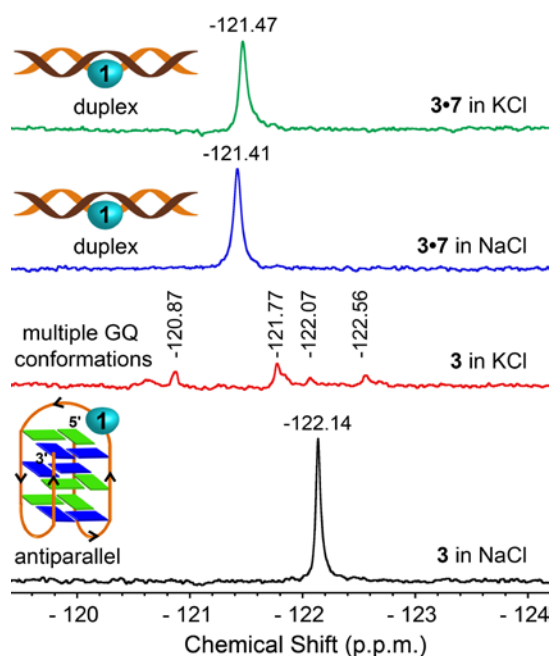**Figure 11.** Fluorescence spectra of nucleoside **1** (5 µM) in water, sodium phosphate buffer containing 100 mM NaCl, potassium phosphate buffer containing 100 mM KCl and intraoocyte buffer (25 mM HEPES (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl₂, 1 mM MgCl₂, 0.1 mM EDTA). Changes in ionic conditions did not affect the fluorescence profile of the free nucleoside probe.

The ability of nucleoside probe **1** to photophysically distinguish different GQ topologies from the duplex form is due to distinct conformation and microenvironment of the probe in these structures. In the duplex state, the base-paired nucleoside analog is likely to experience a strong stacking interaction with adjacent bases, which is known to promote non-radiative decay pathway.[53] The presence of guanine residue next to the emissive analog can further quench the fluorescence as guanine is very well known to quench the fluorescence of several dyes by electron transfer process.[54] However, in the GQ structures, the nucleoside analog placed in the loop region is not H-bonded and is also away from the G-tetrad. Hence, enhancement in fluorescence exhibited by different GQs is likely due to reduced (i) stacking interaction between the fluorophore and adjacent bases and (ii) electron transfer process between the fluorophore and guanine residues. Further, solvation-desolvation and rigidification-derigidification of the fluorophore could have also influenced the fluorescence outcome of the probe in duplex and GQ structures.

3.2.5 ^{19}F label exhibits a distinct and resolved signature for different GQ topologies

Unlike ^{1}H NMR spectrum of a GQ structure, which shows multiple signals for the imino protons between 10 and 12 ppm,[55] the proton-decoupled ^{19}F NMR spectrum of modified H-Telo DNA ONs should give a single and distinct peak for each and every GQ structure if the ^{19}F-labeled nucleoside **1** is sensitive to subtle differences in the conformations (Figure 1B). H-Telo DNA ON **3**, containing modified nucleoside in the middle loop was chosen as the study model. In consensus with fluorescence data, the duplex form (**3•7**) in the presence of

NaCl or KCl gave very similar chemical shifts (Figure 12). Interestingly, ON **3** in K$^+$ ionic conditions produced at least four well resolved peaks corresponding to potentially four different GQ topologies as has been predicted in literature for this sequence.[30] Although ON **3** in the presence of K$^+$ ions showed higher fluorescence and lifetime as compared to the duplex form, we could not ascertain the formation of multiple structures as in the case of ^{19}F NMR measurements. ON **3**, which adopts only an antiparallel conformation in the presence of Na$^+$ ions,[56] gave only one peak with a distinct chemical shift. Collectively, these results clearly underscore the bifunctionality of the nucleoside probe in faithfully reporting the formation as well as discriminating different GQ topologies by fluorescence and NMR techniques.

Figure 12. ^{19}F label of nucleoside **1** gives a distinct NMR signal for different H-Telo DNA GQ topologies. ^{19}F NMR spectrum (100 µM) of H-Telo DNA ON **3** and corresponding duplex (**3•7**) in sodium phosphate buffer containing 100 mM NaCl or potassium phosphate buffer containing 100 mM KCl at 25 °C. All samples (100 µM) contained 20% D$_2$O and each spectrum was referenced relative to an external standard (TFT).

### 3.2.6 Estimation of ligand binding to different GQ topologies

The conformation-sensitivity of the dual-purpose probe was put to use in estimating the binding affinity of ligands to different H-Telo GQ structures of ON **3** by titrating with pyridostatin (PDS) and BRACO19, which are known functional GQ binders (Figure 13).[57,58] As the concentration of the ligand was increased a dose-dependent quenching in fluorescence intensity, with no apparent change in emission maximum, was observed (Figure 14). The

fluorescence quenching upon ligand binding could be due to the distinct conformation attained by the nucleoside probe, which is less rigid and favours nonradiative decay pathway as a result of its proximity to the polyaromatic ligands.[59] The apparent K_d values determined from fluorescence experiment revealed that PDS binds to the antiparallel conformation with higher binding affinity as compared to hybrid type structures formed in KCl conditions (Figure 15 and Table 5). On the contrary, BRACO19 shows higher binding affinity for hybrid type structures as compared to the antiparallel GQ topology. Further, the suitability of ^{19}F label in detecting the ligand binding event was tested by using H-Telo DNA ON **3** in NaCl, wherein it forms only an antiparallel conformation with an intense peak at -122.14 ppm. Upon increasing PDS concentration, the peak intensity of free GQ structure reduced and a new broad peak at -121.69 ppm corresponding to the ligand-bound GQ appeared (Figure 16). Small molecule ligands tested against ONs in screening assays are usually soluble in DMSO. To test the amount of DMSO that would not affect the prevalent GQ topology in solution, CD, fluorescence and ^{19}F NMR spectrum of ON **3** was recorded in a buffer containing increasing percentage of DMSO. The results indicate that up to 2.5 % of DMSO can be used in binding assays without affecting the GQ topology (Figure 17).
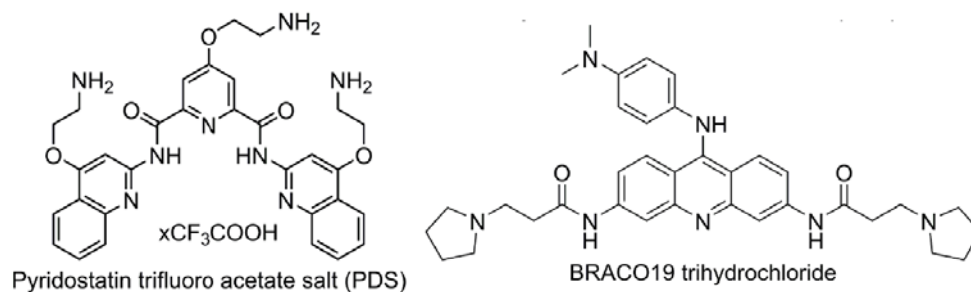
**Figure 13.** Chemical structure of GQ binder, pyridostatin (PDS) and BRACO19
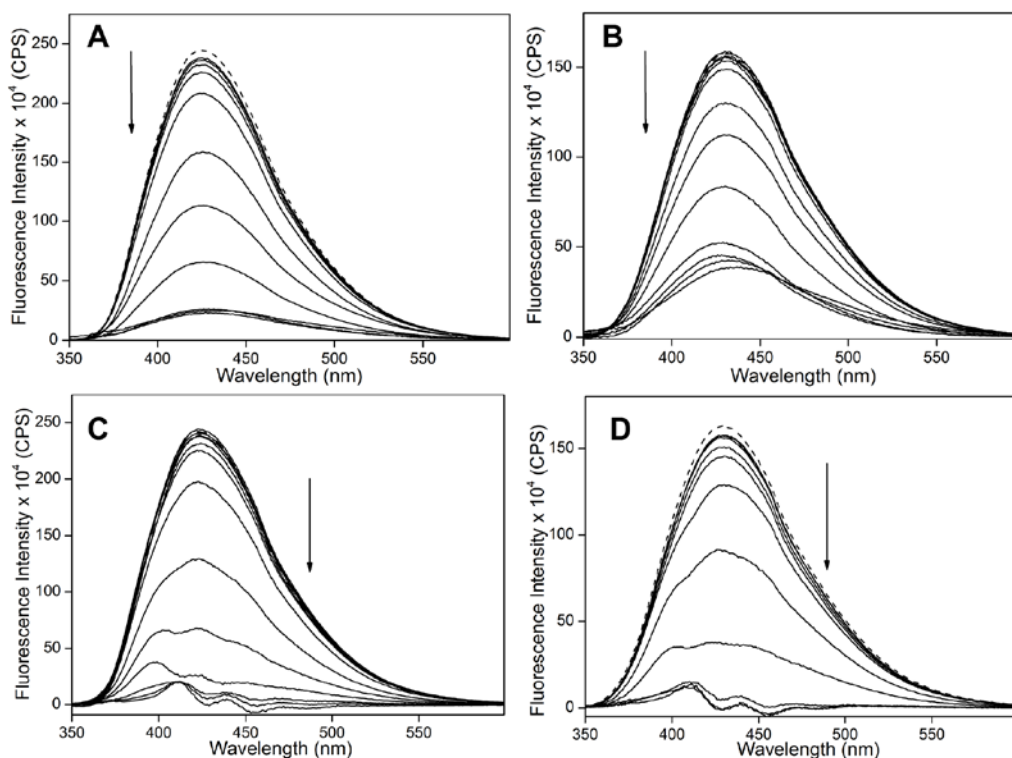
Figure 14. Emission spectra for the titration of labeled H-Telo DNA ON **3** (0.5 µM) with increasing concentration of PDS (**A**) in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl and (**B**) potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl. Emission spectra for the titration of labeled H-Telo DNA ON **3** (0.5 µM) with increasing concentration of BRACO19 (**C**) in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl and (**D**) potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl. Samples were excited at 330 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. The dashed line represents the spectrum of ON **3** without ligand.
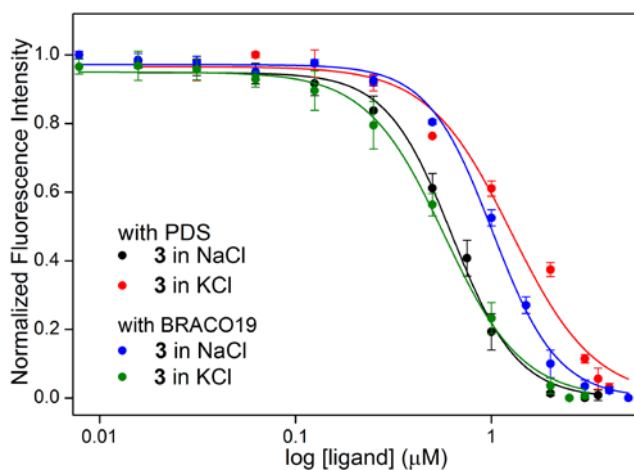

Figure 15. Curve fits for the binding of PDS and BRACO19 to GQ structures of H-Telo DNA ON **3** in different ionic conditions.

Table 5. Dissociation constant (K_d) of PDS and BRACO19 binding to H-Telo DNA ON **3**.

ligand	K_d (μM)	
	in NaCl	in KCl
PDS	0.63 ± 0.03	1.26 ± 0.06
BRACO19	1.01 ± 0.03	0.57 ± 0.05
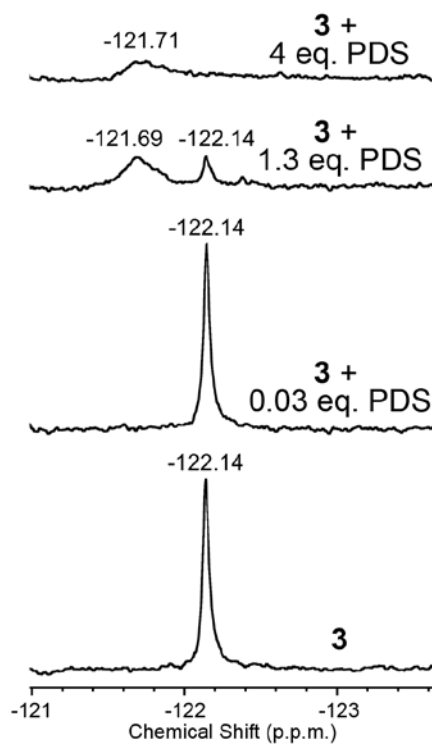
Figure 16. Changes in ^{19}F NMR signal of the antiparallel GQ structure of ON **3** (100 μM) in sodium phosphate buffer containing 100 mM NaCl as a function of increasing PDS concentration.
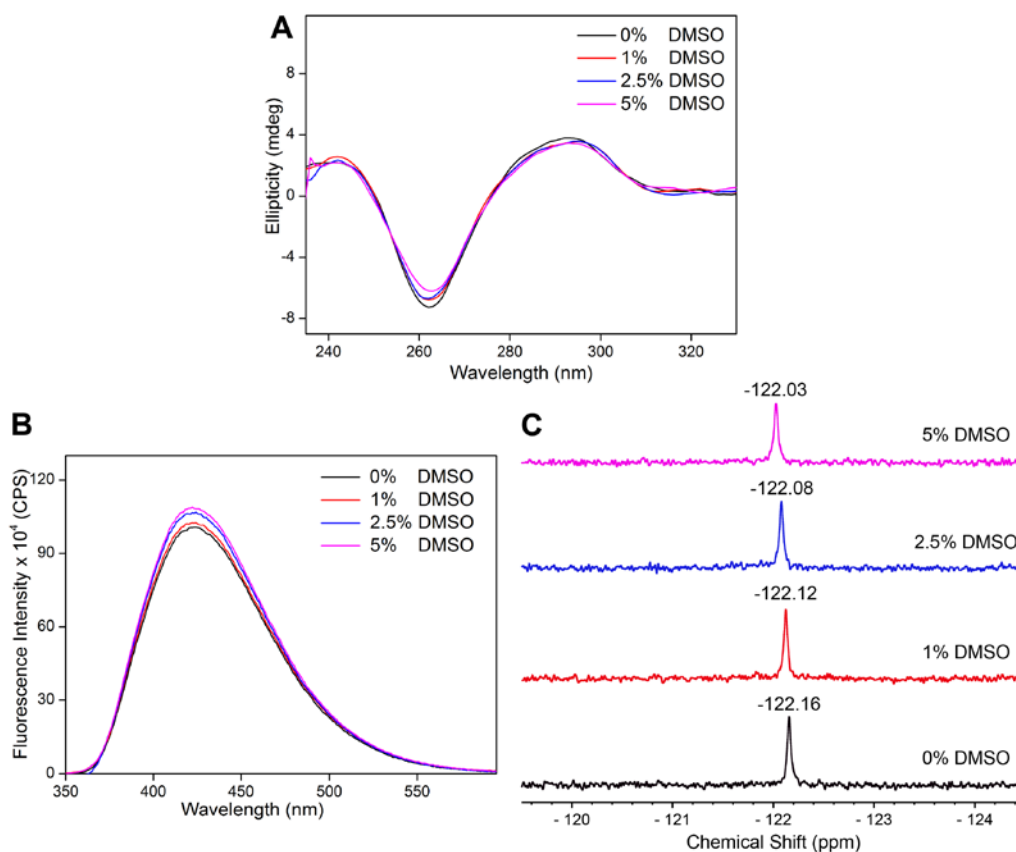
Figure 17. CD (8 µM), fluorescence (0.5 µM) and ^{19}F NMR spectra (100 µM) of ON **3** in sodium phosphate buffer containing 100 mM NaCl with increasing percentage of DMSO.

### 3.2.7 Structural insights into the H-Telo DNA overhang in live cells

We took advantage of the sensitivity of ^{19}F label, which provides distinct NMR signatures for different GQs, in determining the topology of H-Telo overhang in live cells. *Xenopus laevis* oocyte was chosen as the model cell as it is large in size, easily injectable and is also commonly used in in-cell NMR analysis of proteins[60] and more recently, nucleic acids.[45,61] These NMR studies rely on transfection or microinjection of isotope enriched proteins and oligonucleotides as endogenous proteins and nucleic acids do not contain intrinsic isotope labels. Similarly, a notable number of GQ-specific light-up and fluorogenic small molecule probes require transfection of ON sequence of interest into the cells for efficient visualization.[38b,40,62] Unlike proteins, isotope labeling of nucleic acids with ^{15}N/^{13}C is laborious and very expensive. In this context, the fluorobenzofuran-modified nucleoside analog has an added advantage as it can be easily incorporated into various DNA ON sequences by solid-phase method.

To obtain a progressive understanding of the GQ structure in cell-free and cellular environments, ^{19}F NMR spectrum was recorded in intraoocyte buffer conditions, oocyte clear lysate, egg extract (*ex vivo* model) and live oocyte (*in vivo* model). Particularly for these experiments, fluorobenzofuran-labeled H-Telo DNA ONs **8** and **9**, which predominantly form hybrid type 1 and 2 GQ structures, respectively, in K^+ ionic conditions were additionally synthesized (K^+ concentration is significantly higher than Na^+ concentration in intraoocyte environment).[63,64] While both these sequences form a 3+1 GQ structure (three strands parallel and one strand antiparallel), they show difference in loop arrangement as the double chain reversal loop in hybrid 1 and hybrid 2 is located at the 5′-end and 3′-end, respectively (Figure 18A).
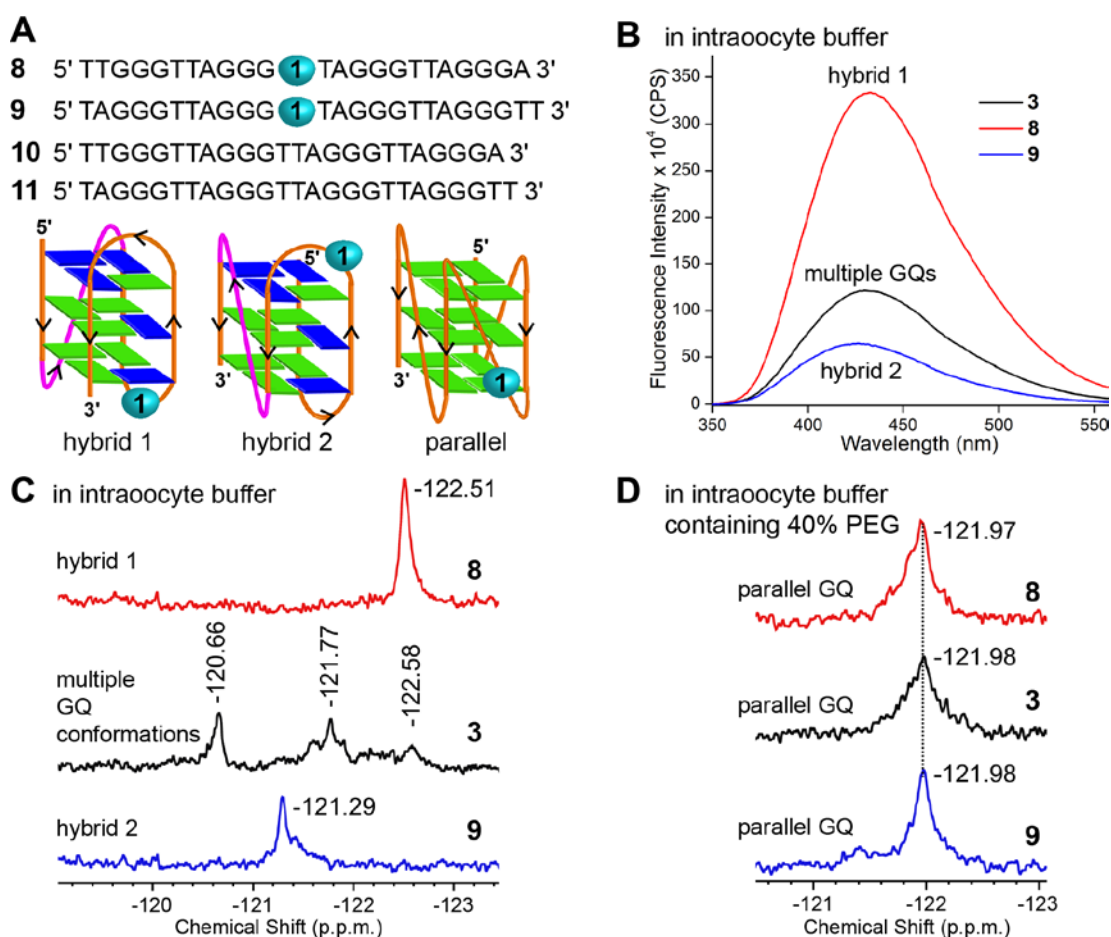

Figure 18. (**A**) Sequence of labeled H-Telo DNA ONs **8** and **9** and respective control unmodified ONs **10** and **11**, which predominantly adopt hybrid 1 and hybrid 2 GQ conformation, respectively. (**B**) Nucleoside analog **1** photophysically distinguishes hybrid 1 and hybrid 2 GQ structures of ONs **8** and **9**. ON samples (0.5 µM) were excited at 330 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. (**C**) ^{19}F NMR signature of H-Telo DNA ONs **3**, **8** and **9** (100 µM) in intraoocyte buffer at 18 °C. (**D**) ^{19}F NMR signature of the parallel GQ conformation of ONs **3**, **8** and **9** (100 µM). ONs **3**, **8** and **9** convert into parallel GQ conformation in the presence of PEG (Figure 19).[26]

Fluorescence and NMR analyses in intraoocyte buffer (25 mM HEPES, pH = 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl$_2$, 1 mM MgCl$_2$, 0.1 mM EDTA)[45a] clearly revealed the ability of the nucleoside probe to distinguish different GQ structures of H-Telo ONs **3**, **8** and **9** (Figure 18). Hybrid 1 conformation of **8** exhibited noticeably higher fluorescence intensity (~5-fold) as compared to hybrid 2 conformation of **9** (Figure 18B). Similarly, these GQs showed sharp and distinct ^{19}F NMR signals (Figure 18C). Consistent with the structural polymorphism of ON **3**, we observed intermediate fluorescence and multiple ^{19}F signals. In intraoocyte buffer, containing 40% of a synthetic crowding agent (PEG), all the three ONs gave a similar chemical shift corresponding to the parallel conformation (Figure 18D and Figure 19). PEG, due to its dehydrating effect, is known to bias the H-Telo DNA ONs to adopt a parallel GQ structure in solution as well as in solid-state.[28] It is important to mention here that imino proton signals of labeled ONs were marginally affected by the 5-fluorobenzofuran label (Figure 20). Further, CD and T_m measurements revealed that the control unmodified and modified ONs formed respective GQ structures with little difference in CD and melting profiles (Figure 21 and Table 6). From these studies, we successfully determined the ^{19}F NMR signature of individual H-Telo GQ structures for further conformational analysis in oocytes (Figure 22).
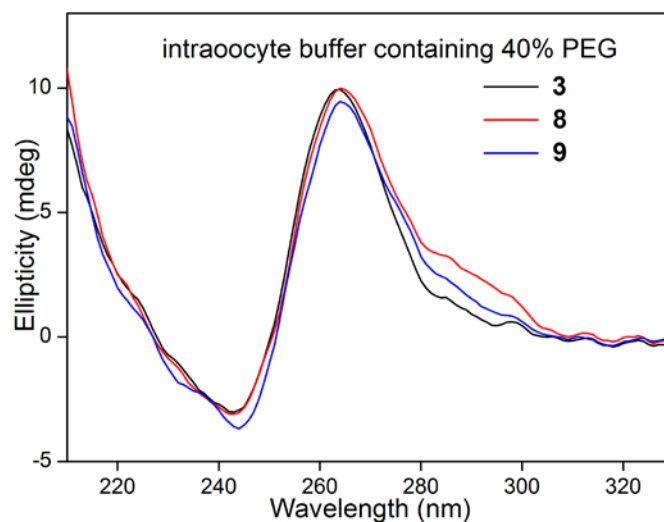
Figure 19. CD spectra of 5-fluorobenzofuran modified H-Telo DNA **3**, **8** and **9** (8 µM) in intraoocyte buffer (25 mM HEPES, pH = 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl$_2$, 1 mM MgCl$_2$, 0.1 mM EDTA) containing 40% PEG. In presence of PEG all three modified ONs adopted a parallel GQ conformation.
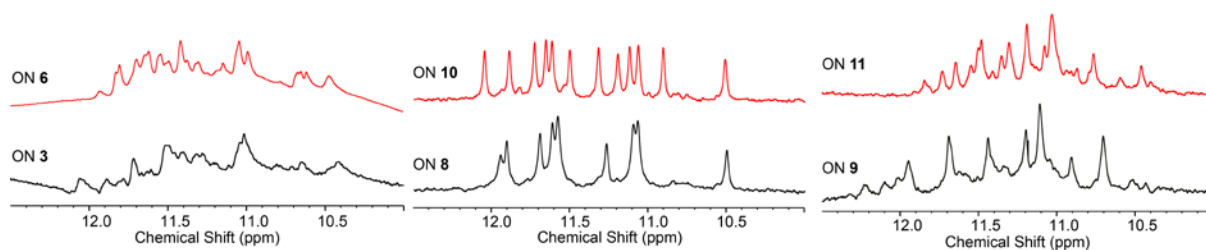
Figure 20. ^{1}H NMR spectrum of 5-fluorobenzofuran-labeled (black) and control unmodified (red) H-Telo DNA ONs in intraoocyte buffer conditions. All samples (100 µM) contained 20% D_2O and all spectra were recorded at 18 °C.
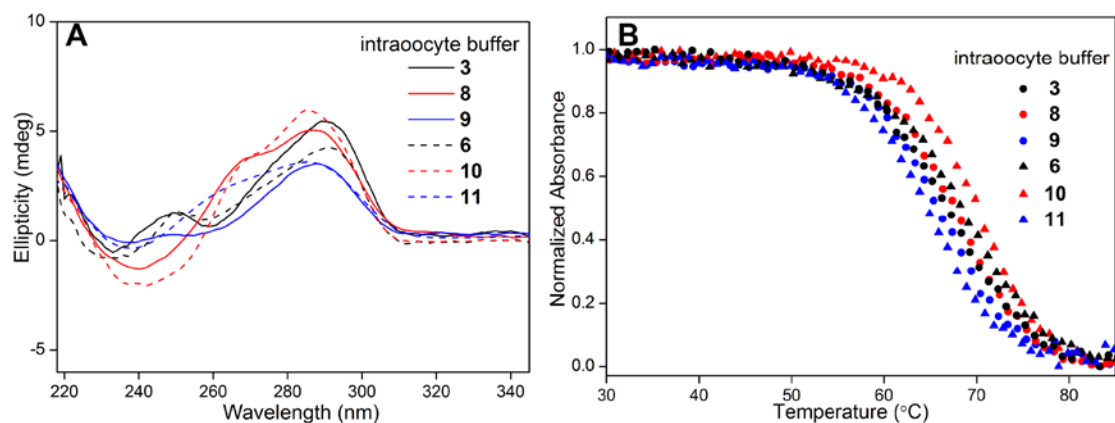
Figure 21. (**A**) CD spectra of 5-fluorobenzofuran-modified H-Telo DNA ONs **3**, **8** and **9** (8 µM) and respective unmodified H-Telo DNA ONs **6**, **10** and **11** in intraoocyte buffer at 20 °C. (**B**) UV-thermal melting profile of above ONs (1 µM) in intraoocyte buffer. For T_m values see Table 6.

Table 6. T_m values of modified H-Telo DNA ONs (**3**, **8** and **9)** and corresponding unmodified H-Telo DNA ONs (**6**, **10** and **11)** in intraoocyte buffer.

Modified H-Telo DNA ON	T_m (°C)	Control H-Telo DNA ON	T_m (°C)
3	68.1 ± 0.6	**6**	69.7 ± 0.5
8	68.1 ± 0.4	**10**	70.9 ± 0.1
9	66.7 ± 0.5	**11**	66.4 ± 0.9
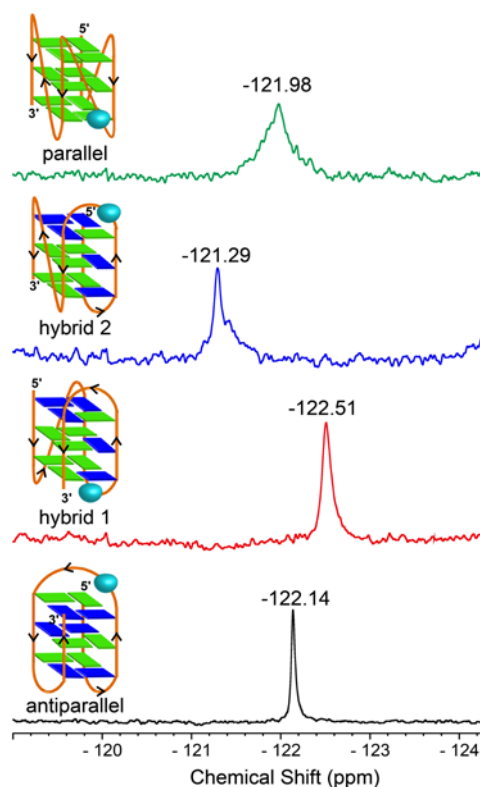
Figure 22. ^{19}F NMR signature for different GQ topologies of H-Telo DNA ON repeats.

Mechanically crushed oocyte suspension was heated at 95 °C to denature the proteins, and the clear lysate thus obtained after centrifugation was used in NMR experiments. In case of measurements in egg extract, the oocytes were allowed to mature, which were then crushed and without further manipulations the mixture was centrifuged. The crude inter-phase egg extract, apart from providing molecular crowding, is also known to maintain most of the biological activities of an intact cell.[44] The ^{19}F NMR spectrum of ON **3** in clear lysate almost resembled the spectra obtained in intraoocyte buffer conditions (Figure 23). Although in egg extract there was an indication for the formation of multiple GQ structures, the ^{19}F signals were broad and not well resolved. This observation is not surprising as molecular crowding and inhomogeneity of the extract can reduce the tumbling rate and increase the relaxation process.[65] It is important to mention here that ^{1}H NMR spectrum of **3** in lysate and egg extract was poorly resolved to provide structural information, which is in consensus with the reported NMR spectrum of the native H-Telo DNA ON of the same sequence.[45b] However, ^{19}F NMR experiment indicates that telomeric ON repeat **3** is not completely converted into the parallel topology in cellular environment as has been predicted by using synthetic molecular crowding agents and cosolvents (Figure 23).[26,28]
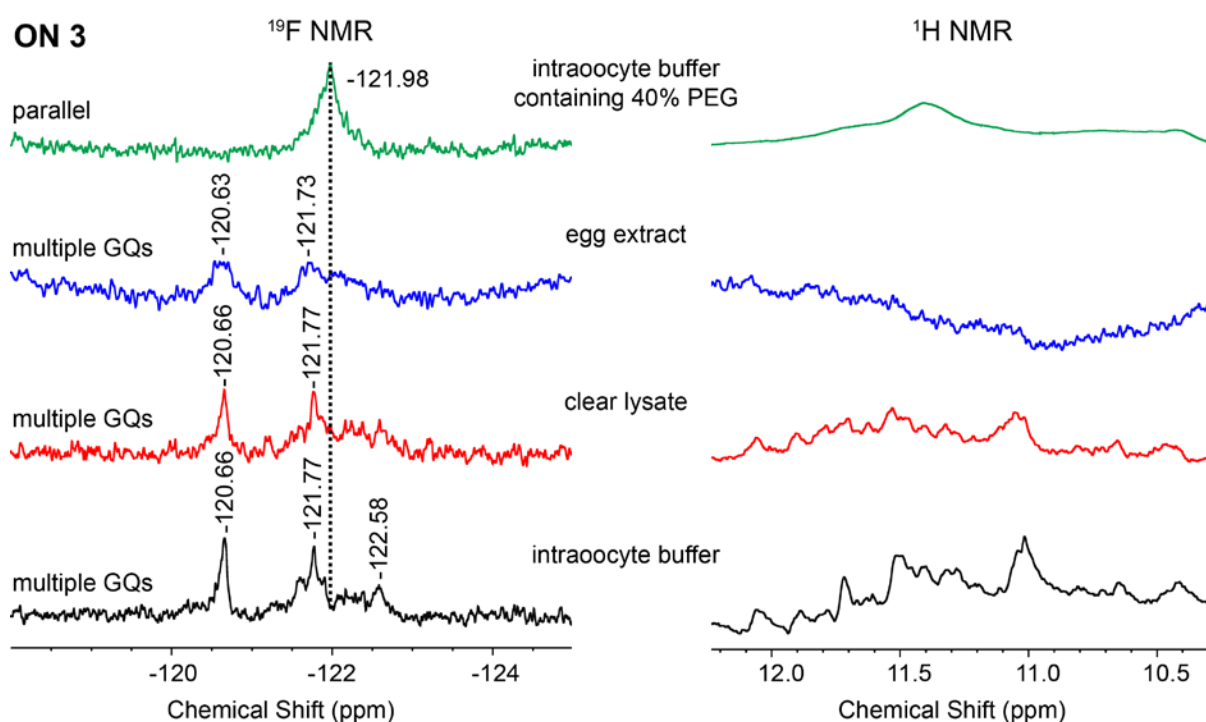
Figure 23. ^{19}F and ^{1}H NMR spectra of H-Telo DNA ON **3** in intraoocyte buffer, oocyte clear lysate, egg extract and intraoocyte buffer containing 40% PEG at 18 °C.

A spectrum of ON **8** in clear lysate displayed a sharp peak at -122.47 ppm corresponding to the signature of hybrid 1 GQ structure obtained in intraoocyte buffer conditions (Figure 24A). Interestingly, a small but visible peak at -121.97 ppm potentially corresponding to the parallel GQ conformation was also observed. Further analysis in egg extract produced a broader peak most probably encompassing both hybrid 1 and parallel conformations. In the absence of experimental proof, it is believed that the merging of signals could be due to dynamic inter-conversion between hybrid 1 and parallel conformations.[66,67] Unfortunately, ^{19}F NMR signatures could not be matched in live oocytes due line broadening. ^{1}H NMR spectrum of this ON sequence indicated the formation of hybrid 1 structure in clear lysate, but failed to provide structural information in egg extract due to severe line broadening (Figure 24A). These results support the notion that parallel GQ is not the only conformation adopted by H-Telo DNA repeat in cellular environment.
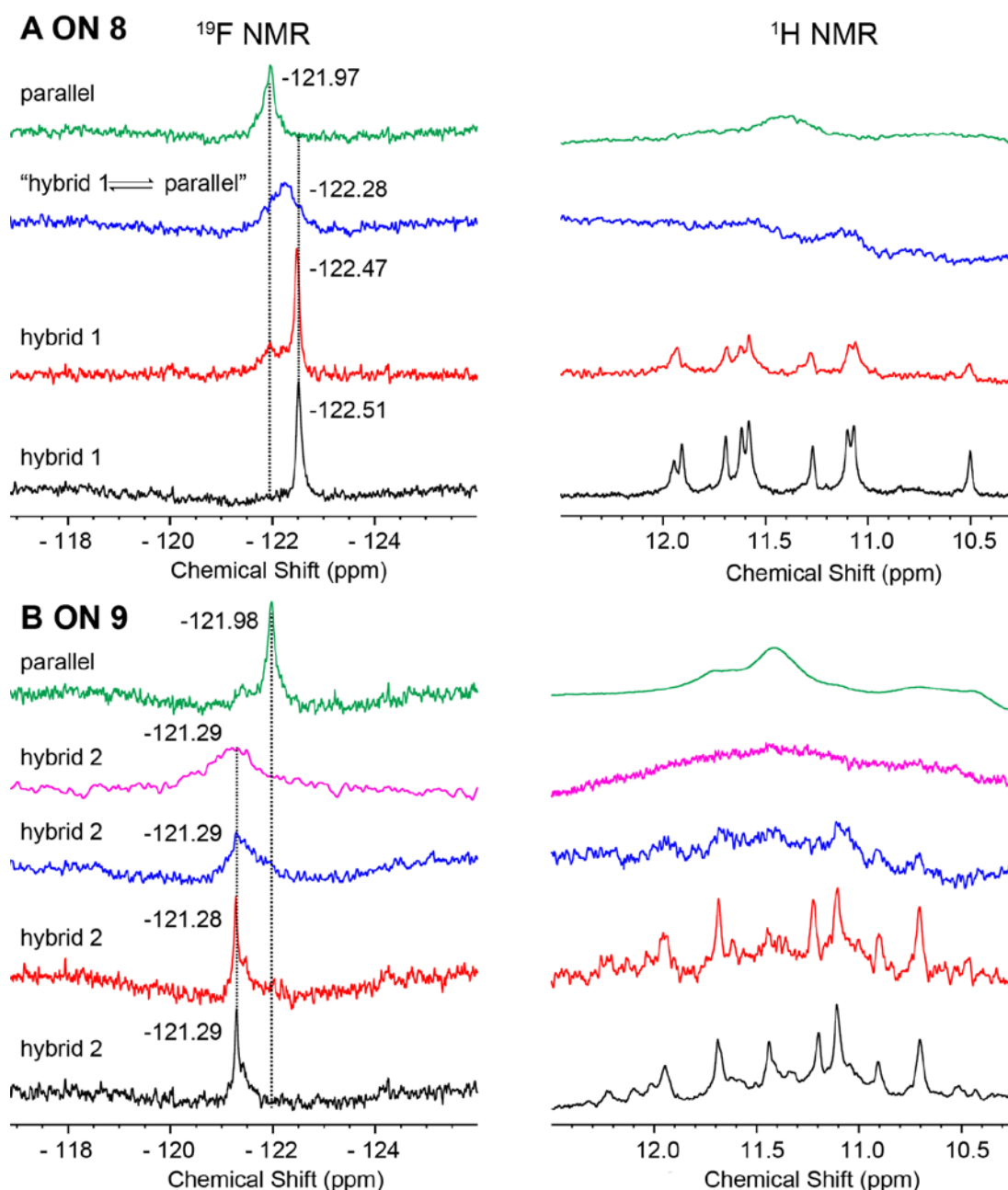
Figure 24. [19]F label of nucleoside **1** serves as a useful tool to determine the GQ structure of H-Telo overhang in cellular environment. (**A**) ^{19}F and ^{1}H NMR spectra of ON **8** in intraoocyte buffer, oocyte clear lysate, egg extract and intraoocyte buffer containing 40% PEG. (**B**) ^{19}F and ^{1}H NMR spectra of ON **9** in intraoocyte buffer, oocyte clear lysate, egg extract, live oocytes and intraoocyte buffer containing 40% PEG. NMR measurements were performed at 18 °C.

ON **9** presented an identical ^{19}F NMR signal (-121.29 ppm) in clear lysate and egg extract corresponding to the hybrid 2 conformation obtained in intraoocyte buffer conditions, albeit little line broadening in egg extract (Figure 24B). Following this observation, a concentrated solution of ON **9** (3.6 mM) in water was microinjected (50 nL) into live oocytes. Nearly 200 oocytes each containing nearly 180 μM of the DNA ON were used in the

in-cell NMR experiments. Rewardingly, a broad peak at -121.29 ppm characteristic of a hybrid 2 GQ conformation was obtained in live oocytes (Figure 24B). Unlike ON **8**, hybrid 2 conformation of ON **9** did not show detectable conversion into other structures. Notably, under these conditions there was no apparent cell death. Due to severe line broadening, ^{1}H NMR analysis in egg extract and live oocytes could not be used in deducing the structure, which again substantiates the limited use of ^{1}H NMR in in-cell structural analysis (Figure 24B). To confirm if the observed ^{19}F signal associated with the folded GQ structure of ON **9** is indeed emanating from inside the live cells, the following control experiments were performed. After recording the NMR spectrum, the extracellular buffer was subjected to ^{19}F NMR analyses. No signal was seen in ^{19}F NMR spectrum. The leakage of ON **9** was also tested by a fluorescence leakage assay (Figure 25A). By comparing the area under the fluorescence curve of lysate obtained after 0 h (sample A), 10 h of incubation after microinjection of ON **9** (sample C) and extracellular buffer after 10 h of incubation (sample B), it was found that only a small fraction (6–10%) of the microinjected sample leaked into the extracellular buffer during the NMR acquisition time (~10 h, Figure 25B) (See experimental section for more details). The integrity of labeled ON **9** in egg extract and live oocytes after NMR measurement was also examined. The egg extract and live oocytes were heat denatured and the clear lysate obtained was subjected to HPLC and mass analyses. The retention time of extracted ON **9** and its mass as compared to the authentic sample confirmed the intactness of the ON in the cellular environment (Figure 26 and Table 7). It is important to mention here that in-cell concentration of the dual-app probe-labeled ON is considerably higher than the concentrations of nucleic acids in cells. However, performing NMR analysis using ultrasensitive cryogenic ^{19}F probe will significantly bring down the working concentration of labeled DNA ONs in cells.[68]
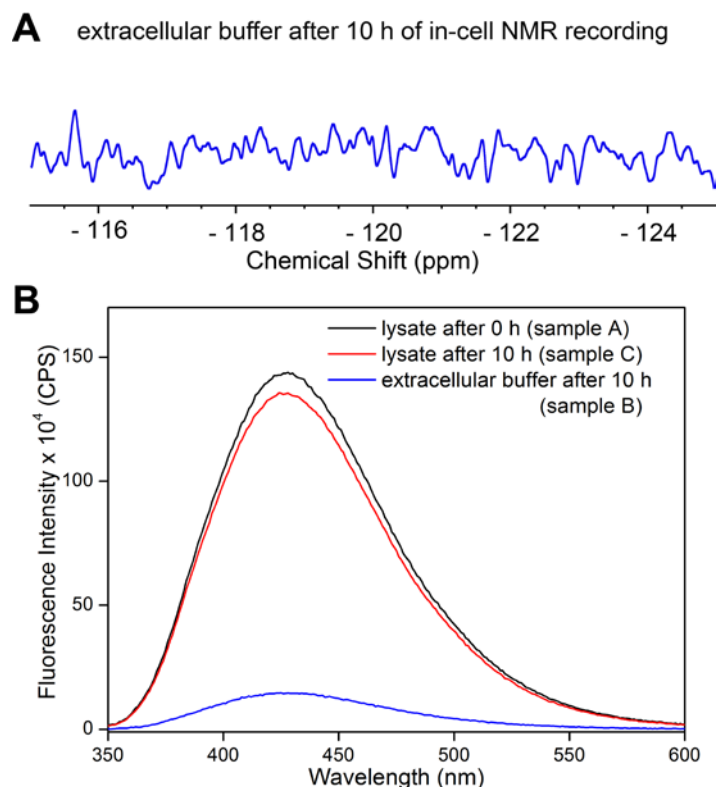
Figure 25. (**A**) ^{19}F NMR spectrum of extracellular buffer after 10 h of in-cell NMR analysis. (**B**) Fluorescence leakage assay: Emission spectrum of lysates obtained after 0 h (sample A) and 10 h of incubation after microinjection of ON **9** (sample C) and extracellular buffer after 10 h of incubation (sample B). Samples were excited at 330 nm with excitation and emission slit width 4 nm and 6 nm, respectively. See experimental section for details.
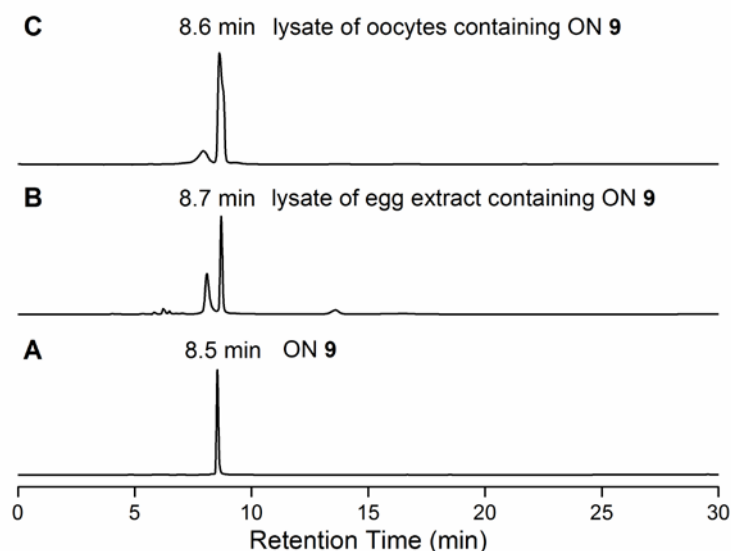
Figure 26. HPLC chromatogram of (**A**) neat ON **9**, (**B**) lysate of egg extract containing H-Telo DNA ON **9** after NMR analysis, (**C**) lysate of oocytes microinjected with H-Telo DNA ON **9** after NMR analysis at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0−100% B in 30 min. HPLC analysis was performed using Luna C18 column (250 x 4.6 mm, 5 micron).

Table 7. Mass analysis of HPLC fractions obtained from NMR samples of ON **9**.

Sample	Calculated mass of ON **9**	Observed mass
lysate of egg extract containing H-Telo DNA ON **9**	7999.2	7999.4
lysate of oocytes microinjected with H-Telo DNA ON **9**	7999.2	7998.9

Among the various GQ-forming sequences predicted to be present in the genome, the ones present in proto-oncogenes have attracted significant attention.[1,2] Many of these sequences have been individually studied *in vitro* to assess their propensity to adopt different GQ structures.[12] However, it is very important to understand what GQ topology or topologies an individual GQ-forming sequence will form in the cell and how they interact with ligands so that screening platforms can be designed to identify GQ motif-specific binders. In this regard, fluorobenzofuran-modified nucleoside probe is highly beneficial as its fluorescence and NMR labels could be used to set up screening assays and detect GQ topology in cells, respectively. Further, high throughput parallel and conventional solid-phase DNA synthesizers have similar reaction cycles. Hence, accessibility to different GQ-forming sequences labeled with the dual-app probe should not be difficult.

3.3 Conclusions

We have used an innovative yet a simple probe design approach to develop a new nucleoside probe functioning both as conformation-sensitive fluorescent and NMR labels. The fluorophore component of the dual-purpose probe provides an *in vitro* platform to photophysically distinguish different GQ topologies and estimate topology-specific binding of ligands. The ^{19}F-label of the nucleoside exhibits unique signatures for individual GQ conformations, thereby facilitating the direct correlation of GQ structures formed *in vitro* and in native cellular environment. Our findings using the dual-purpose nucleoside probe indicate that H-Telo DNA ON repeat adopts multiple conformations namely, hybrid-type and parallel GQ structures in cellular environment. The telomeric repeat did not outrightly get converted into a parallel conformation in cellular environment. This is contrary to the GQ studies using immunofluorescence staining in fixed cells, synthetic crowding agents and under crystallization conditions, which strongly support the predominant formation of a parallel conformation.[28,42] The utility of our probe and new findings demonstrated using telomeric DNA repeat are highly important on two counts. The conformations adopted by a nucleic

acid motif in non-native conditions need not be favoured under complex cellular environment and the structural information obtained using synthetic cell models cannot be straightforwardly extended to native cells. Hence, understanding the topology adopted by individual GQ-forming sequences in cells and identify small molecule ligands that can specifically target a GQ motif amongst others in the genome will have a profound impact on GQ-directed therapeutic strategies. In this context, our bifunctional nucleoside probe represents a unique tool, which will not only enable the determination of GQ structures adopted by individual GQ-forming sequences in cells by NMR, but also harness GQ-directed therapeutic strategies by supporting fluorescence-based discovery platforms to identify topologic-specific binders.

3.4 Experimental Section

3.4.1. Materials

N,N,N′,N′-tetramethylethylenediamine, *n*-butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) dichloride, 5-iodo-2′-deoxyuridine, 4,4′-dimethoxytrityl chloride, *N,N*-diisopropylethylamine, pyridostatin trifluoroacetate salt, BRACO19 hydrochloride, progesterone, PEG 200, Ficoll 400 and all reagents (Bio-Ultra grade) used for buffer preparation were purchased from Sigma-Aldrich. 5-fluorobenzofuran (**1a**) was synthesized following a previously reported procedure.[69] 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite was purchased from Alfa Aesar. *N*-benzoyl-protected dA, *N,N*-dimethylformamide-protected dG, *N*-acetyl-protected dC and dT phosphoramidite substrates for solid-phase synthesis were procured from Proligo Reagents. Solid supports for DNA synthesis were obtained from Glen Research. All other reagents, required for solid-phase oligonucleotide (ON) synthesis were purchased either from ChemGenes Corporation or from Sigma-Aldrich. DNA ONs **6**, **7**, **10** and **11** were purchased from Integrated DNA Technology, and purified by denaturing polyacrylamide gel electrophoresis (PAGE). ONs were eluted and desalted using Sep-Pak Classic C18 cartridges (Waters Corporation). Amicon Ultra (0.5 mL, 3K) centrifugal filters were obtained from Merck Millipore. Autoclaved Millipore water was used in all biophysical analysis.

3.4.2 Instrumentation

Modified ONs were synthesized on a ABI applied Biosystems 392 DNA/RNA synthesizer. Mass measurements were conducted either on an Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer or on a Water Synapt G2 High Definition mass spectrometers. HPLC analysis was done using Agilent Technologies 1260 Infinity HPLC. Reverse-phase flash chromatographic (C18 RediSepRf column) purifications were carried out using Teledyne ISCO, Combi Flash Rf. Absorption spectra were recorded on a Shimadzu UV-2600 spectrophotometer. Steady-state and time-resolved fluorescence spectra were recorded on a TCSPC Fluorolog-3 fluorescence spectrometer (Horiba Jobin Yvon). UV-thermal melting analysis of ONs was carried out on Cary 300 Bio UV-Vis spectrophotometer. CD analysis was performed on a JASCO J-815 CD spectrometer. NMR spectra of small molecules were recorded on a Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. NMR spectra of ONs were recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe and processed using Bruker TopSpin Software. Microinjection of oocytes was performed by using a NARISHIGE micromanipulator equipped with a manual oocyte microinjection pipette (Drummond Scientific Company).

3.4.3 Synthesis of 5-fluorobenzofuran-2′-deoxyuridine 1 and corresponding phosphoramidite 2

3.4.3.1 Tributyl(5-fluorobenzofuran-2-yl)stannane (1b)

5-Fluorobenzofuran **1a**[69] (0.91 g, 6.7 mmol, 1.0 equiv) was dissolved in dry THF (30 mL) and was cooled at -78 °C. TMEDA (0.93 g, 8.0 mmol, 1.2 equiv) was added to the reaction mixture and the mixture was allowed to stir at -78 °C for 30 min. Subsequently, *n*-BuLi (4.0 mL of 2 M solution in hexane, 8.0 mmol, 1.2 equiv) was added dropwise and the reaction mixture was allowed to stir for 1 h under N_2 atmosphere. The mixture was brought to room temperature over a period of 1 h and again cooled to -78 °C. Bu_3SnCl (2.17 mL, 8.0 mmol, 1.2 equiv) was added dropwise to the reaction mixture and stirred for 1 h. Afterwards, the reaction was quenched with ammonium chloride solution (50 mL) and extracted two times with diethyl ether (2 × 50 mL). The organic layer was dried over sodium sulphate and evaporated to give an oily residue. The residue was purified by silica gel flash column chromatography (hexane) to afford clear oil (1.7 g, 60%). TLC (5% EtOAc in hexane); R_f = 0.67; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 7.42–7.38 (m, 1H), 7.19 (dd, J_1 = 8.6 Hz, J_2 = 2.6 Hz, 1H), 6.93 (td, J_1 = 9.0 Hz, J_2 = 2.7 Hz, 1H), 6.86–6.84 (m, 1H), 1.63–1.55 (m, 6H), 1.40–

1.31 (m, 6H), 1.20–1.11 (m, 6H), 0.90 (t, J = 7.4 Hz, 9H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 168.2, 160.3, 157.9, 155.1, 129.0, 128.9, 118.2, 118.2, 111.4, 111.3, 111.1, 110.9, 105.7, 105.5, 29.1, 27.3, 13.8, 10.3; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -123.70; MALDI-TOF: m/z Calculated for C$_{20}$H$_{31}$FNaOSn [M+Na]$^+$ = 449.13, found = 449.14.

3.4.3.2 5-Fluorobenzofuran-2′-deoxyuridine (1)

A mixture of 5-iodo-2′-deoxyuridine **1c** (0.7 g, 1.98 mmol, 1.0 equiv) and *bis*(triphenylphosphine)-palladium(II) dichloride (0.07 g, 0.10 mmol, 0.05 equiv) was dissolved in degassed anhydrous dioxane (30 mL) and tributyl(5-fluorobenzofuran-2-yl)stannane **1b** (2.5 g, 5.90 mmol, 3.0 equiv) was added. The mixture was heated at 90 °C for 2 h under N$_2$ atmosphere and filtered through Celite pad. The Celite pad was washed with methanol (2 × 15 mL). The filtrate was evaporated and the resulting residue was purified by reverse-phase column chromatography (C18 Redi*SepR$_f$* column, 60% MeOH in H$_2$O) to afford product **1** as white solid (0.57 g, 81%). TLC (15% MeOH in CH$_2$Cl$_2$); R_f = 0.48; ^{1}H NMR (400 MHz, d_6-DMSO): δ (ppm) 11.78 (s, 1H), 8.79 (s, 1H), 7.56 (dd, J_1 = 9.0 Hz, J_2 = 4.2 Hz, 1H), 7.43 (dd, J_1 = 8.8 Hz, J_2 = 2.8 Hz, 1H), 7.33 (br, 1H), 7.11 (td, J_1 = 9.2 Hz, J_2 = 2.8 Hz, 1H), 6.22 (t, J = 6.4 Hz, 1H), 5.30 (d, J = 4 Hz, 1H), 5.26–5.24 (m, 1H), 4.36–4.30 (m, 1H), 3.89–3.87 (m, 1H), 3.75–3.65 (m, 2H), 2.25–2.22 (m, 2H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 160.3, 159.8, 157.5, 151.1, 149.4, 149.3, 137.6, 129.9, 129.8, 111.7, 111.6, 111.3, 106.6, 106.3, 104.3, 103.9, 103.9, 87.6, 85.1, 70.0, 60.8, 40.5; ^{19}F NMR (376.6 MHz, d_6-DMSO): δ (ppm) -121.77; HRMS: m/z Calculated for C$_{17}$H$_{15}$FN$_2$NaO$_6$ [M+Na]$^+$ = 385.0812, found = 385.0806; λ$_{max}$ (H$_2$O) = 267, 275 and 322 nm, ε$_{267}$ = 12.6 x 10^3 M^{-1} cm^{-1}, ε$_{275}$ = 13.31 x 10^3 M^{-1} cm^{-1}, ε$_{322}$ = 17.11 x 10^3 M^{-1} cm^{-1}, ε$_{260}$ = 10.31 x 10^3 M^{-1} cm^{-1}.

3.4.3.3 5-Fluorobenzofuran-modified 5′-*O*-DMT-2′-deoxyuridine (1d)

A mixture of 5-fluorobenzofuran modified deoxyuridine **1** (0.30 g, 0.84 mmol, 1.0 equiv), DMAP (0.010 g, 0.084 mmol, 0.1 equiv), DMT-Cl (0.37 g, 1.1 mmol, 1.3 equiv) and anhydrous pyridine (10 mL) was stirred at room temperature for 12 h under N$_2$ atmosphere. Then pyridine was evaporated under vacuum and the resulting residue was purified by silica gel column chromatography (1% MeOH in CH$_2$Cl$_2$ containing 0.5% Et$_3$N) to afford compound **1d** as an off white foam (0.3 g, 54%). TLC (5% MeOH in CH$_2$Cl$_2$ containing few drops of Et$_3$N); R_f = 0.4; ^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 8.51 (s, 1H), 7.50 (br, 1H), 7.49 (br, 1H), 7.38–7.35 (m, 5H), 7.22–7.18 (m, 2H), 7.13–7.09 (m, 2H), 6.72 (d, J = 8.4 Hz,

4H), 6.65–6.59 (m, 1H), 6.43 (t, J = 6.6 Hz, 1H), 5.99 (dd, J_1 = 8.8 Hz, J_2 = 4 Hz, 1H), 4.51–4.49 (m, 1H), 4.10–4.09 (m, 1H), 3.65–3.61 (m, 7H), 3.31 (dd, J_1 = 10.8 Hz, J_2 = 2.8 Hz, 1H) , 2.55–2.50 (m, 1H), 2.41–2.34 (m, 1H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 160.4, 160.0, 158.7, 158.1, 149.8, 149.4, 149.3, 144.6, 135.9, 135.7, 135.6, 130.2, 130.2, 128.3, 128.0, 127.1, 113.3, 113.3, 111.9, 111.7, 111.5, 111.4, 106.7, 106.5, 106.2, 105.9, 105.9, 87.0, 86.5, 85.7, 72.1, 63.3, 55.3, 41.6; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -122.32; HRMS: m/z Calculated for C$_{38}$H$_{33}$FN$_2$NaO$_8$ [M+Na]$^+$ = 687.2119, found = 687.2119.

3.4.3.4 5-Fluobenzofuran modified 2′-deoxyuridine phosphoramidite substrate (2)

To a solution of **1d** (0.2 g, 0.3 mmol, 1.0 equiv) in anhydrous dichloromethane (2.7 mL) was added DIPEA (0.26 mL, 1.5 mmol, 5 equiv) and stirred for 10 min. To this solution 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (0.10 mL, 0.45 mmol, 1.5 equiv) was added and the mixture was stirred for 2.5 h. Subsequently, the reaction mixture was evaporated to dryness and the residue was redissolved in ethyl acetate (20 mL). The organic layer was washed with 5% sodium bicarbonate solution (20 mL) and brine solution (20 mL) successively, dried over sodium sulphate and evaporated to dryness. The residue was purified by silica gel column chromatography (40% EtOAc in hexane containing 0.5% Et$_3$N) to afford the product **2** as a white foam (0.16 g, 61%). TLC (hexane:EtOAc= 50:50 containing 1% Et$_3$N); R_f = 0.56, 0.68 for two diastereomers; ^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 8.59 (s, 1H), 7.52–7.51 (m, 2H), 7.40–7.37 (m, 5H), 7.22–7.18 (m, 2H), 7.12–7.07 (m, 2H), 6.72–6.70 (m, 4H), 6.58–6.53 (m, 1H), 6.43–6.39 (m, 1H), 5.77 (dd, J_1 = 8.8 Hz, J_2 = 4 Hz, 1H), 4.61–4.56 (m, 1H), 4.23–4.22 (m, 1H), 3.69 (dd, J_1 = 10.8 Hz, J_2 = 2.4 Hz, 1H), 3.65 (s, 3H), 3.64 (s, 3H), 3.62–3.52 (m, 5H), 3.24 (dd, J_1 = 10.6 Hz, J_2 = 2.6 Hz, 1H), 2.60–2.54 (m, 1H), 2.41–2.38 (m, 2H), 1.17–1.14 (m, 12H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 160.4, 160.1, 158.7, 158.1, 149.7, 149.4, 149.2, 144.6, 136.0, 135.7, 135.7, 130.3, 130.2, 129.8, 129.7, 128.4, 128.0, 127.1, 117.5, 113.3, 113.2, 111.8, 111.5, 111.5, 111.4, 106.7, 106.3, 106.1, 105.9, 105.8, 86.8, 86.2, 86.2, 85.7, 73.2, 73.1, 62.8, 58.4, 58.2, 55.3, 43.5, 43.4, 40.9, 40.8, 24.8, 24.7, 24.7, 24.6, 20.3, 20.3; ^{31}P NMR (162 MHz, CDCl$_3$): δ (ppm) 149.23; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -122.48; HRMS: m/z Calculated for C$_{47}$H$_{51}$FN$_4$O$_9$P [M+H]$^+$ = 865.3378, found = 865.3379.

3.4.4 Photophysical analysis of 5-fluorobenzofuran-modified nucleoside analog 1 in different solvents.

3.4.4.1 UV absorption and steady-state fluorescence studies

UV absorption spectrum of nucleoside **1** (25 µM) in various solvents was recorded on a Shimadzu UV-2600 spectrophotometer in quartz cuvette (Hellma, path length 1 cm). Each sample contained 2.5% of DMSO and measurements were performed in triplicate. For steady-state fluorescence study, nucleoside **1** (5 µM) was excited in different solvents at their respective lowest energy absorption maximum (Table 1). Fluorescence experiments were performed in triplicate in micro fluorescence cuvette (Hellma, path length 1 cm) on a Fluorolog-3 spectrophotometer. Fluorescence samples contained 0.5% of DMSO. Anisotropy values (r) of nucleoside **1** in different solvents were determined by analysing the data using software provided with the instrument. Anisotropy measurements were performed in triplicate and each anisotropy value was an average of 10 successive measurements.

3.4.4.2 Time-resolved fluorescence study

Time-resolved fluorescence study was performed on a TCSPC instrument (Horiba Jobin Yvon, Fluorolog-3). Nucleoside **1** (5 µM) in water, ethylene glycol and glycerol was excited by using 339 nm LED source (IBH, UK, NanoLED-339L). In dioxane and methanol nucleoside **1** (400 µM) was excited by using 375 nm diode laser source (IBH, UK, NanoLED-375L). Fluorescence signal at respective emission maxima was collected. All studies were done in triplicate and lifetimes were calculated by fitting the decay profile using IBH DAS6 software. The χ^2 value for all the curve fits was found to be nearly one.

3.4.4.3 Quantum yield calculation

Quantum yield of nucleoside **1** in different solvents was determined relative to 2-aminopurine as the standard. Following equation was used to calculate the quantum yield.[70]

$$\Phi_{F(x)} = \left(A_s/A_x\right)\left(F_x/F_s\right)\left(n_x/n_s\right)^2 \Phi_{F(s)}$$

Where s is the standard, x is the modified nucleoside, A is the absorbance at excitation wavelength, F is the area under the emission curve, n is the refractive index of the solvent, and Φ_F is the quantum yield. Quantum yield of 2-aminopurine in water is 0.68. Quantum yield of fluorescent H-Telo DNA ONs and corresponding duplexes was determined relative to the quantum yield of nucleosides **1**. Quantum yield of nucleoside **1** in water is 0.11.

3.4.5 Solid phase DNA synthesis

5-fluorobenzofuran-modified H-Telo DNA ONs **3**–**5**, **8** and **9** were synthesized (1 µmole scale, 1000 Å CPG solid support) by a standard ON synthesis protocol using phosphoramidite **2**.[71] In DNA synthesis, the solid support after final detritylation step was treated with 30% aqueous ammonium hydroxide solution for 16 h at 55 °C and 30 min at 70 °C. The solution was evaporated to dryness using speed vac and the residue was purified by denaturing PAGE (20% gel). The band corresponding to the full-length modified ON product was identified by UV shadowing, which was cut and transferred to a Poly-Prep column (Bio-Rad). The gel pieces were crushed with a sterile glass rod, and the ON was extracted using ammonium acetate buffer (0.5 M, 3 mL) for 12 h and desalted using Sep-Pak classic C18 cartridges (Waters). The purity of ONs **3**–**5**, **8** and **9** was confirmed by RP-HPLC and characterized by MALDI TOF mass analysis (see Figure 5, Figure 6 and Table 2).

3.4.6 MALDI-TOF mass analysis of ONs

A mixture of 2 µL of the modified ON (~200 µM), 4 µL of a 9:1 solution of 3-hydroxypicolinic acid and ammonium citrate buffer (100 mM, pH 9) and 2 µL of an internal DNA standard (100 µM) was desalted using an ion-exchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form). The sample was incubated at 55 °C for 10 min and cooled to RT. The sample was then spotted on a MALDI plate, air dried and subjected to mass analysis. The resultant spectra were calibrated relative to the internal DNA standard (Figure 6 and Table 2).

3.4.7 Circular dichroism (CD) studies

Modified and unmodified ONs (8 µM) were heated at 90 °C for 3 min in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl, or in potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl or in intraoocyte buffer (25 mM HEPES (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA). The samples were allowed to come to room temperature over a period of 2 h and were kept at 4 °C for 30 min before recording the spectrum. CD spectra were recorded from 200 to 350 nm in a quartz cuvette (Starna Scientific, path length 5 mm) on a J-815 CD spectropolarimeter (Jasco, USA) using 1 nm bandwidth at 20 °C. Each CD profile is an average of three scans collected at a

scan speed of 100 nm min^{-1}. CD measurements were performed in duplicate and all spectra were corrected using an appropriate blank solution in the absence of ONs.

3.4.8 Thermal melting analysis of DNA ONs

Modified and unmodified DNA ONs (1 µM) were annealed similarly like CD samples and thermal melting analysis was performed using Cary 300 Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 295 nm.

3.4.9 Steady-state and time-resolved fluorescence of labeled H-Telo DNA ONs

GQs of ONs **3–5** (0.5 µM) were formed by heating the ONs at 90 °C for 3 min either in sodium phosphate buffer (10 mM, pH 7.0) containing 100 mM NaCl or potassium phosphate buffer (10 mM, pH 7.0) containing 100 mM KCl. The samples were allowed to come to room temperature over a period of 2 h and were kept at 4 °C for 30 min. ON duplexes were made by heating a 1:1.1 mixture of ONs **3–5** and complementary DNA ON **7** at 90 °C for 3 min in buffers as mentioned above. Similarly, ONs **3**, **8** and **9** were annealed to form respective GQ structures in intraoocyte buffer (25 mM HEPES (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl$_2$, 1 mM MgCl$_2$, 0.1 mM EDTA). Steady-state fluorescence analysis of GQs (0.5 µM) and duplexes (0.5 µM) was performed by exciting the samples at 330 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively, at 20 °C. Time-resolved fluorescence analysis was performed by exciting the samples (0.5 µM) using a 339 nm LED source (IBH, UK, NanoLED-339L) on TCSPC instrument (Horiba Jobin Yvon, Fluorolog-3) at 20 °C, and collecting the fluorescence signal at respective emission maximum. All measurements were performed in triplicate and lifetimes were calculated by fitting the decay profile using IBH DAS6 software. The χ^2 value for all the curve fits was found to be nearly one. Quantum yield determination is provided in experimental section 3.4.4.3.

3.4.10 Fluorescence binding assay

H-Telo DNA ON **3** was annealed to form respective GQs in different buffers as mentioned above. A series of ON **3** samples (0.5 µM) containing increasing concentration of the ligand (PDS or BRACO19; 8 nM to 5 µM) was prepared and incubated for 30 min before fluorescence analysis. Samples were excited at 330 nm with an excitation and emission slit width of 4 nm and 6 nm, respectively. The measurements were made in triplicate at 20 °C.

The spectrum corresponding to a blank without any ON but containing the particular ligand concentration was subtracted from each spectrum. From the dose-dependent quenching curves, the apparent dissociation constants (K_d) for the binding of ligands to the H-Telo DNA ON were determined by fitting normalized fluorescence intensity (F_N) *versus* log of ligand concentration plot to Hill equation (Origin 8.5).[46]

$$F_N = \frac{F_i - F_s}{F_0 - F_s}$$

F_i is the fluorescence intensity at each ligand concentration. F_0 and F_s are the fluorescence intensity in the absence of ligand (L) and at saturation point, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$F_N = F_0 + (F_s - F_0)\left(\frac{[L]^n}{[K_d]^n + [L]^n}\right)$$

3.4.11 Sample preparation for ^{19}F NMR analysis of H-Telo DNA ONs

3.4.11.1 Phosphate and intraoocyte buffers

H-Telo DNA ONs (100 µM) were annealed into GQs in phosphate buffers (containing NaCl or KCl) and intraoocyte buffer by heating the samples at 90 °C for 3 min. All samples contained 20% D_2O. The samples were cooled to room temperature and were kept at 4 °C for 30 min. Samples were transferred into a Shigemi tube (5 mm advance NMR microtube) for NMR analysis.

3.4.11.2 Clear lysate[45]

Around 250 healthy stage VI oocytes were selected and kept in a Petri dish containing Ori-Ca^{2+} buffer (5 mM HEPES (pH = 7.6), 110 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$ and 1 mM $MgCl_2$). The Petri dish was kept on an ice bath for 15 min, and the oocytes were transferred to another Petri dish containing ice cold intraoocyte buffer. Finally oocytes were transferred into an Eppendorf tube, buffer above the oocytes was removed carefully and 200 µL ice cold intraoocyte buffer supplemented with 20% D_2O was added. Oocytes were mechanically crushed in ice and insoluble mixture was removed by centrifuging at 20000 g for 20 min at 4 °C. The inter-phase extract was taken in another Eppendorf tube and heated at 95 °C for 10 min. The precipitated protein fraction was removed by centrifuging at 20000 g for 10 min at

4 °C. The supernatant (~270 µL) was taken in an Eppendorf tube and 30 µL of 1 mM pre-annealed DNA ON stock in water (see above) was added to it so that the final concentration of the DNA ON was 100 µM.

3.4.11.3 Egg extract[44]

850–900 oocytes were transferred into a Petri dish containing 1 µM progesterone in Ori Ca^{2+} buffer and incubated for 12 h. Matured eggs were then washed thoroughly with Ori buffer (5 x 15 ml). The eggs in Ori Ca^{2+} buffer were kept on an ice bath for 20 min. The eggs were then transferred to another Petri dish containing ice cold intraoocyte buffer, and finally to an Eppendorf tube containing intraoocyte buffer (200 µL) composed of 20% D_2O. The buffer just above the eggs was removed and again intraoocyte buffer (200 µL) containing 20% D_2O was added. This step was repeated two more times. The eggs were then packed by centrifuging at 400 g for 1 min (4 °C) and the buffer just above the eggs was removed very carefully. Afterwards, the eggs were centrifuged for 5 min at 12000 g (4 °C) and were mechanically crushed on an ice bath. The extract was centrifuged for 30 min at 12000 g (4 °C). Around 280 µL of crude inter-phase extract was obtained and 15 µL of 5 mM pre-annealed DNA ON stock in water was added. The prepared mixture was incubated at 37 °C for 30 min and subjected to NMR analysis.

3.4.11.4 Live oocytes

The chemically synthesized ONs, which were PAGE purified, were again precipitated using butanol (400 µL of butanol for 40 µL of DNA sample). ONs were dissolved in water and filtered two times using a 0.45 µm spin filter tube. A stock solution of DNA ONs (3.6 mM) in autoclaved water was heated at 90 ºC for 3 min and slowly cooled to room temperature. 50 nL of DNA ON stock was microinjected into each oocyte using a NARISHIGE micromanipulator equipped with a manual oocyte microinjection pipette (Drummond Scientific Company). Around 200 oocytes were injected with the labeled DNA ON over a period of 3.5 h. Each injected oocyte will contain ~180 µM of the labeled DNA ON based on the injected volume and size of an oocyte (~ 1 µL). Microinjected oocytes were transferred to a Petri dish containing Ori Ca^{2+} buffer and incubated for 1 h. The oocytes were thoroughly washed with Ori Ca^{2+} buffer (5 x 15 mL) and transferred to a Shigemi NMR tube prefilled with 1 mL of Ori Ca^{2+} buffer containing 20% D_2O and 10% Ficoll 400. Ficoll was used in the extracellular buffer so that the oocytes are not crushed during the NMR acquisition time.[61]

3.4.12 ^{19}F NMR analysis of H-Telo DNA ONs

^{19}F NMR spectra were recorded at a frequency of 564.9 MHz on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe. All spectra were referenced relative to an external standard trifluorotoluene (TFT, -63.72 ppm). In phosphate buffers, NMR studies were performed at 25 °C, and in intraoocyte buffer, lysate, egg extract and live oocytes, spectra were recorded at 18 °C. The following spectral parameters were used[72,48] : [19]F excitation pulse: 11 µs; spectral width: 21.28 ppm; transmitter frequency offset: -121.14 ppm; acquisition time: 1.4 s in phosphate buffer, 80 ms in intraoocyte buffer and lysate, 50 ms in egg extract and live oocytes; relaxation delay: 1.5 s in phosphate buffer, 1 s in intraoocyte buffer, lysate, egg extract and live oocytes; number of scans 28000 or 39000. Using these parameters ^{19}F NMR spectra were obtained in 8–10 h. Each spectrum was processed with an exponential window function using lb = 10 Hz in case of studies in buffer, lysate and egg extract and lb = 50 Hz in case of in-cell study.[72]

3.4.13 ON leakage study by ^{19}F NMR and fluorescence

To check the leakage of microinjected ONs during the course of NMR analysis (~10 h), the extracellular buffer was isolated and ^{19}F NMR was recorded. No signal was observed in ^{19}F NMR spectrum (Figure 25A).

The leakage of ON was also tested by fluorescence analysis (Figure 25B). 3.6 mM stock of ON **9** was injected into two different sets of 30 oocytes. Oocytes were incubated for 1 h in Ori Ca^{2+} buffer and thoroughly washed with the same buffer. One set of 30 injected oocytes were transferred to an Eppendorf tube loaded with 150 µL of Ori Ca^{2+} buffer containing 10% Ficoll. Oocytes were crushed, heat denatured and centrifuged as before to obtain a clear lysate. Final volume of the clear lysate was adjusted to 500 µL with Ori Ca^{2+} buffer containing 10% Ficoll. This lysate sample was considered as a control (zero time) containing 100% of the labeled H-Telo ON **9** (sample A). Another set of 30 injected oocytes was prepared in a similar fashion and incubated in 150 µL of Ori Ca^{2+} buffer containing 10% Ficoll for 10 h. After this period, the extracellular buffer was removed and the volume was adjusted to 500 µL with Ori Ca^{2+} buffer containing 10% Ficoll (sample B). Further, the oocytes from this set were carefully washed with Ori Ca^{2+} buffer containing 10% Ficoll (4 x 200 µL). Oocytes were crushed, heat denatured and centrifuged as before to obtain a clear lysate (sample C). Samples A, B and C were excited at 330 nm with an excitation and

emission slit width of 4 nm and 6 nm, respectively. Appropriate buffer/lysate was used for blank correction. By comparing the area under the fluorescence curve of lysate obtained after 0 h of incubation (sample A) and 10 h incubation (sample C), it was found that only 6% of the ON sample leaked out of the oocytes into the extracellular buffer during the course of NMR acquisition time. Similarly, comparing sample A and B nearly 10% leakage was estimated.

3.4.14 HPLC and mass analysis of egg extract and oocytes after NMR analysis

The egg extract containing H-Telo DNA ON **9** after NMR analysis was heated at 95 °C and centrifuged at 20000 g for 10 min to obtain clear lysate. Oocytes microinjected with ON **9** after NMR experiment were crushed and the lystae was prepared as before. Both the samples were concentrated using Amicon Ultra 3K centrifugal filter and analyzed by HPLC. Appropriate fraction corresponding to ON **9** was isolated and further subjected to mass analysis. See Figure 26 and Table 7.

3.5 References

1. Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 279–284.
2. Bedrat, A.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res.* **2016**, *44*, 1746–1759.
3. Meyne, J.; Ratliff, R. L.; Moyzis, R. *Proc. Natl. Acad. Sci. U. S. A* **1989**, *86*, 7049–7053.
4. Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. *Nucleic Acids Res.* **2008**, *36*, D381–D385.
5. Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627–8637.
6. Halder, K.; Wieland, M.; Hartig, J. S. *Nucleic Acids Res.* **2009**, *37*, 6811–6817.
7. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U. S. A* **2002**, *99*, 11593–11598.
8. Gomez, D.; Guédin, A.; Mergny, J.-L.; Salles, B.; Riou, J.-F.; Teulade-Fichou, M.-P.; Calsou, P. A. *Nucleic Acids Res.* **2010**, *38*, 7187–7198.
9. Alzeer, J.; Vummidi, B. R.; Roth, P. J. C.; Luedtke, N. W. *Angew. Chem. Int. Ed.* **2009**, *48*, 9362–9365.
10. Wang, X.-D.; Ou, T.-M.; Lu, Y.-J.; Li, Z.; Xu, Z.; Xi, C.; Tan, J.-H.; Huang, S.-L.; An, L.-K.; Li, D.; Gu, L.-Q.; Huang, Z.-S. *J. Med. Chem.* **2010**, *53*, 4390–4398.
11. McLuckie, K. I. E.; Waller, Z. A. E.; Sanders, D. A.; Alves, D.; Rodriguez, R.; Dash, J.; McKenzie, G. J.; Venkitaraman, A. R.; Balasubramanian, S. *J. Am. Chem. Soc.* **2011**, *133*, 2658–2663.
12. Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867–5892.
13. Neidle, S. *Nat. Rev. Chem.* **2017**, *1*, 0041.
14. Micco, M.; Collie, G. W.; Dale, A. G.; Ohnmacht, S. A.; Pazitna, I.; Gunaratnam, M.; Reszka, A. P.; Neidle, S. *J. Med. Chem.* **2013**, *56*, 2959–2974.

15. Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.

16. (a) Dhamodharan, V.; Harikrishna, S.; Bhasikuttan, A. C.; Pradeepkumar, P. I. *ACS Chem. Biol.* **2015**, *10*, 821–833. (b) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. *J. Am. Chem. Soc.* **2013**, *135*, 367–376.

17. Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K; Rice, W. G. *Cancer Res.* **2009**, *69*, 7653–7661.

18. Xu, H.; Di Antonio, M.; McKinney, S.; Mathew, V.; Ho, B.; O'Neil, N. J.; Santos, N. D.; Silvester, J.; Wei, V.; Garcia, J.; Kabeer, F.; Lai, D.; Soriano, P.; Banáth, J.; Chiu, D. S.; Yap, D.; Le, D. D.; Ye, F. B.; Zhang, A.; Thu, K.; Soong, J.; Lin, S.; Tsai, A. H. C.; Osako, T.; Algara, T.; Saunders, D. N.; Wong, J.; Xian, J.; Bally, M. B.; Brenton, J. D.; Brown, G. W.; Shah, S. P.; Cescon, D.; Mak, T. W.; Caldas, C.; Stirling, P. C.; Hieter, P.; Balasubramanian, S.; Aparicio, S. *Nat. Commun.* **2017**, *8*, 14432.

19. Chen, Y.; Yang, D. *Curr. Protoc. Nucleic Acid Chem.* **2012**, *50,* 17.5.1–17.5.17.

20. Marušič, M.; Plavec, J. *Angew. Chem. Int. Ed.* **2015**, *54*, 11716–11719.

21. Vialet, B.; Gissot, A.; Delzor, R.; Barthélémy, P. *Chem. Commun.* **2017**, *53*, 11560–11563.

22. Miyoshi, D.; Fujimoto, T.; Sugimoto, N. *Top. Curr. Chem.* **2012**, *330*, 87–110.

23. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Chem. Commun.* **2010**, *46*, 1700–1702.

24. Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. *ChemBioChem* **2013**, *14*, 540–558.

25. Panda, D.; Saha, P.; Das, T.; Dash, J. *Nat. Commun.* **2017**, *8*, 16103.

26. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. *J. Am. Chem. Soc.* **2010**, *132*, 17105–17107.

27. Lannan, F. M.; Mamajanov, I.; Hud, N. V. *J. Am. Chem. Soc.* **2012**, *134*, 15324–15330.

28. Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.

29. Manna, S.; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. *ChemBioChem* **2017**, *18*, 1604–1615.

30. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723–2735.

31. Chen, Z.; Zheng, K.-W.; Hao, Y.-H.; Tan, Z. *J. Am. Chem. Soc.* **2009**, *131*, 10430–10438.

32. Shrestha, P.; Jonchhe, S.; Emura, T.; Hidaka, K.; Endo, M.; Sugiyama, H.; Mao, H. *Nat. Nanotechnol.* **2017**, *12*, 582–588.

33. Artusi, S.; Perrone, R.; Lago, S.; Raffa, P.; Di Iorio, E.; Palu, G.; Richter, S. N. *Nucleic Acids Res.* **2016**, *44*, 10343–10353.

34. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. *Proc. Natl. Acad. Sci. U. S. A* **2001**, *98*, 8572–8577.

35. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186.

36. Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh Jr, R. M.; Sen, D.; Lansdrop, P. M. *Nucleic Acids Res.* **2013**, *42*, 860–869.

37. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 75–80.

38. (a) Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M. Y.; Monchaud, D. *J. Am. Chem. Soc.* **2015**, *137*, 8521–8525. (b) Tseng, T.-Y.; Wang, Z.-F.; Chien, C.-H.; Chang, T.-C. *Nucleic Acids Res.* **2013**, *41*, 10605–10618. (c) Laguerre, A.; Wong, J. M. Y.; Monchaud, D. *Sci. Rep.* **2016**, *6*, 32141. (d) Doria, F.; Nadai, M.; Zuffo, M.; Perrone, R.; Freccero, M.; Richter, S. N. *Chem. Commun.* **2017**, *53*, 2268–2271.

39. Huang, W.-C.; Tseng, T.-Y.; Chen, Y.-T.; Chang, C.-C.; Wang, Z.-F.; Wang, C.-L.; Hsu, T.-N.; Li, P.-T.; Chen, C.-T.; Lin, J.-J.; Lou, P.-J.; Chang, T.-C. *Nucleic Acids Res.* **2015**, *43*, 10102–10113.

40. Chen, S.-B.; Hu, M.-H.; Liu, G.-C.; Wang, J.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Tan, J,-H. *J. Am. Chem. Soc.* **2016**, *138*, 10382–10385.

41. Shivalingam, A.; Izquierdo, M. A.; Marois, A. L.; Vyšniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. *Nat. Commun.* **2015**, *6*, 8178.

42. Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ren, J.; Zhao, Y.; Huang, Z.-S. *Cell Chem. Biol.* **2016**, *23*, 1261–1270.

43. Azarkh, M.; Singh, V.; Okle, O.; Dietrich, D. R.; Hartig, J. S.; Drescher, M. *ChemPhysChem* **2012**, *13*, 1444–1447.

44. Hänsel, R.; Löhr, F.; Foldynoá-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. *Nucleic Acids Res.* **2011**, *39*, 5768–5775.

45. (a) Hänsel, R.; Foldynoá-Trantírková, S.; Löhr, F.; Buck, J.; Bongartz, E.; Bamberg, E.; Schwalbe, H.; Dötsch, V.; Trantírek, L. *J. Am. Chem. Soc.* **2009**, *131*, 15761–15768. (b) Hänsel, R.; Löhr, F.; Trantírek, L.; Dötsch, V. *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.

46. Tanpure, A. A., Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149.

47. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.

48. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. *Nucleic Acids Res.* **2017**, *45*, 5501–5511.

49. Ishizuka, T.; Zhao, P.-Y.; Bao, H.-L.; Xu, Y. *Analyst* **2017**, *142*, 4083–4088.

50. (a) Giam, C. S.; Lyle, J. L. *J. Am. Chem. Soc.* **1973**, *95*, 3235–3239. (b) Sinkeldam, R. W.; Wheat, A. J.; Boyaci, H.; Tor, Y. *ChemPhysChem* **2011**, *12*, 567–570.

51. (a) Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res.* **2007**, *35*, 3064–3075. (b) Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A. *ACS Chem. Biol.* **2015**, *10*, 1311–1318.

52. Rachwal, P. A.; Fox, K. R. *Methods* **2007**, *43*, 291–301.

53. Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. U. S. A* **2001**, *98*, 37–41.

54. Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2009**, *10*, 1389–1398.

55. Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225–255.

56. Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.

57. Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759.

58. Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S. *J. Med. Chem.* **2006**, *49*, 582–599.

59. Le, D. D.; Di Antonio, M.; Chan, L. K. M.; Balasubramanian, S. *Chem. Commun.* **2015**, *51*, 8048–8050.

60. (a) Selenko, P.; Serber, Z.; Gadea, B.; Ruderman, J.; Wagner, G. *Proc. Natl Acad. Sci. U. S. A* **2006**, *103*, 11904–11909. (b) Selenko, P.; Frueh, D. P.; Elsaesser, S. J.; Haas, W.; Gygi, S. P.; Wagner, G. *Nat. Struct. Mol. Biol.* **2008**, *15*, 321–329. (c) Serber, Z.;

Selenko, P.; Hänsel, R.; Reckel, S.; Löhr, F.; Ferrell, J. E., Jr.; Wagner, G.; Dötsch, V. *Nat. Protoc.* **2006**, *1*, 2701–2709.

61. Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. *Chem. Sci.* **2015**, *6*, 3314–3320.

62. Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 14579–14584.

63. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.

64. Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res*. **2006**, *34*, 5715–5719.

65. (a) Selenko, P.; Wagner, G. *J. Struct. Biol.* **2007**, *158*, 244–253. (b) Ye, Y.; Liu, X.; Zhang, Z.; Wu, Q.; Jiang, B.; Jiang, L.; Zhang, Xu.; Liu, M.; Pielak, G. J.; Li, C. *Chem. Eur. J.* **2013**, *19*, 12705–12710.

66. Buer, B. C.; Chugh, J.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2010**, *49*, 5760–5765.

67. Kleckner, I. R; Foster, M. P. *Biochim. Biophys. Acta* **2011**, *1814*, 942–968.

68. Giassa, I.-C.; Rynes, J.; Fessl, T.; Foldynova-Trantirkova, S.; Trantirek, L. *FEBS Letters* **2018**, *592*, 1997–2011.

69. Mewshaw, R. E.; Zhou, D.; Zhou, P.; Shi, X.; Hornby, G.; Spangler, T.; Scerni, R.; Smith, D.; Schechter, L. E.; Andree, T. H. *J. Med. Chem.* **2004**, *47*, 3823–3842.

70. Lavabre, D.; Fery-Forgues, S. *J. Chem. Educ.* **1999**, *76*, 1260–1264.

71. Tanpure, A. A.; Srivatsan, S. G. *ChemBioChem* **2012**, *13*, 2392–2399.

72. Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. *J. Org. Chem.* **2012**, *77*, 1026–1044.

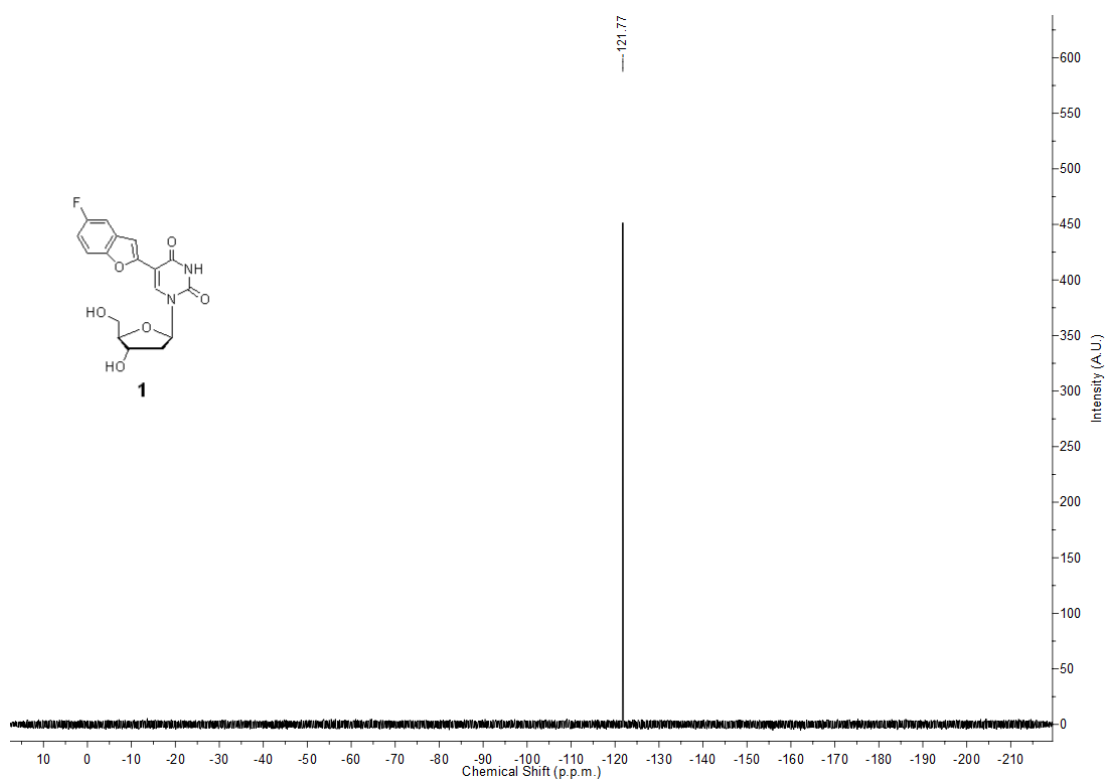# 3.6 Appendix-I: Characterization data of synthesized compounds

^{1}H NMR of **1** (400 MHz, d_6-DMSO)

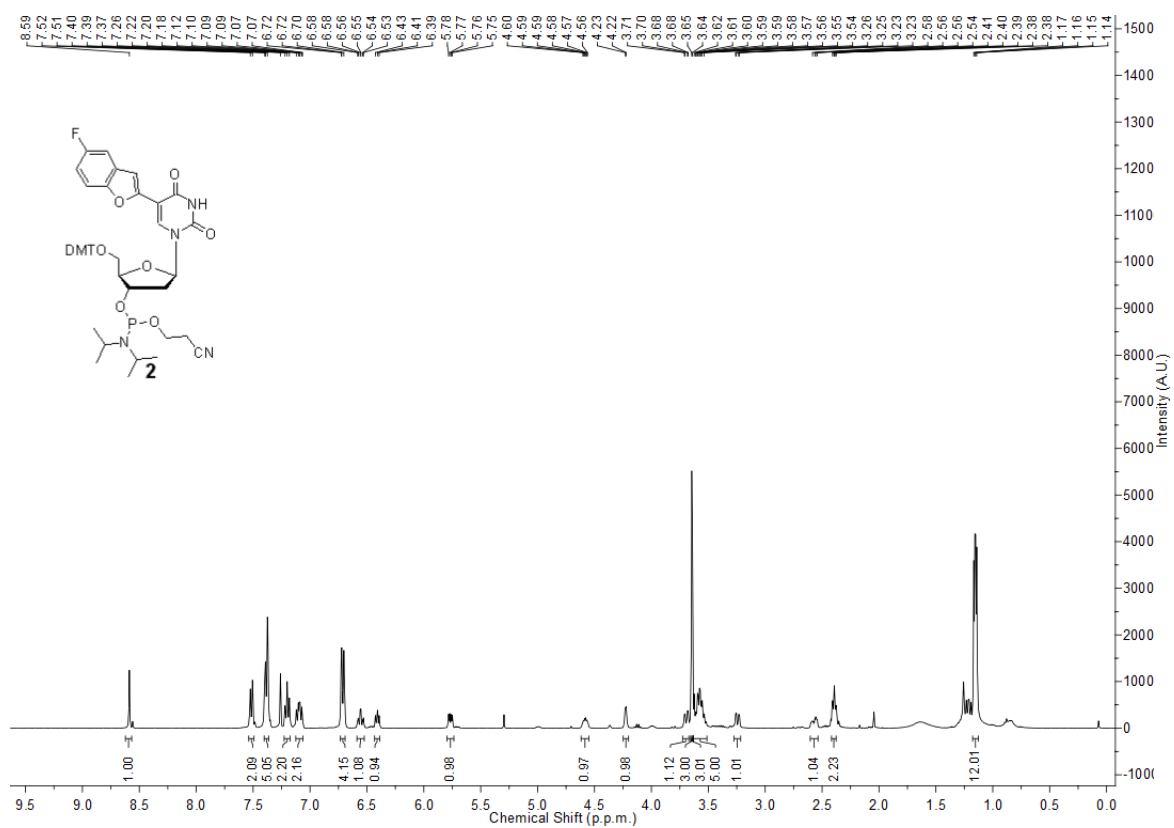^{13}C NMR of **1** (100 MHz, d_6-DMSO)
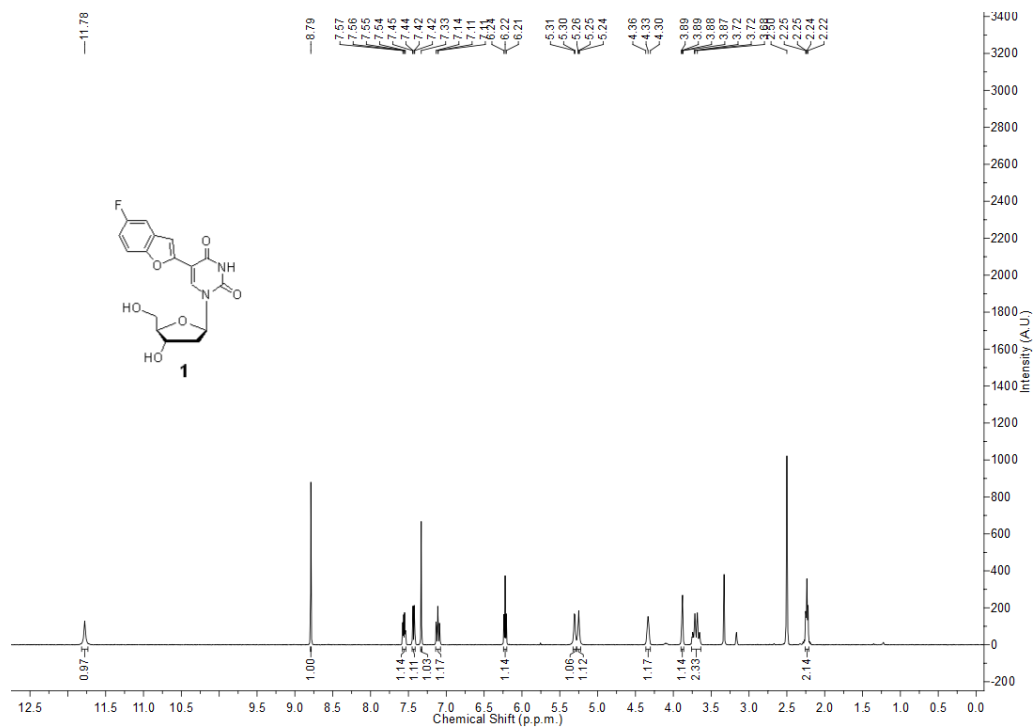
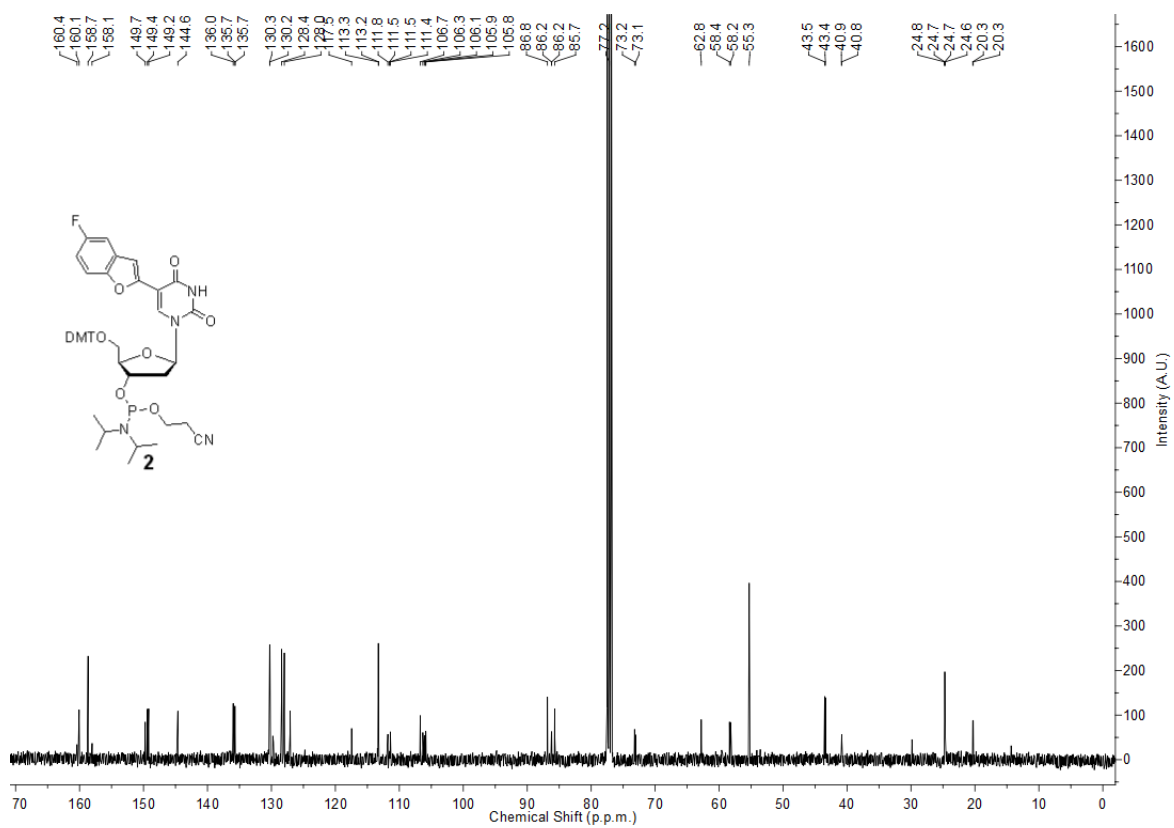¹⁹F NMR of **1** (376.6 MHz, *d₆*-DMSO)
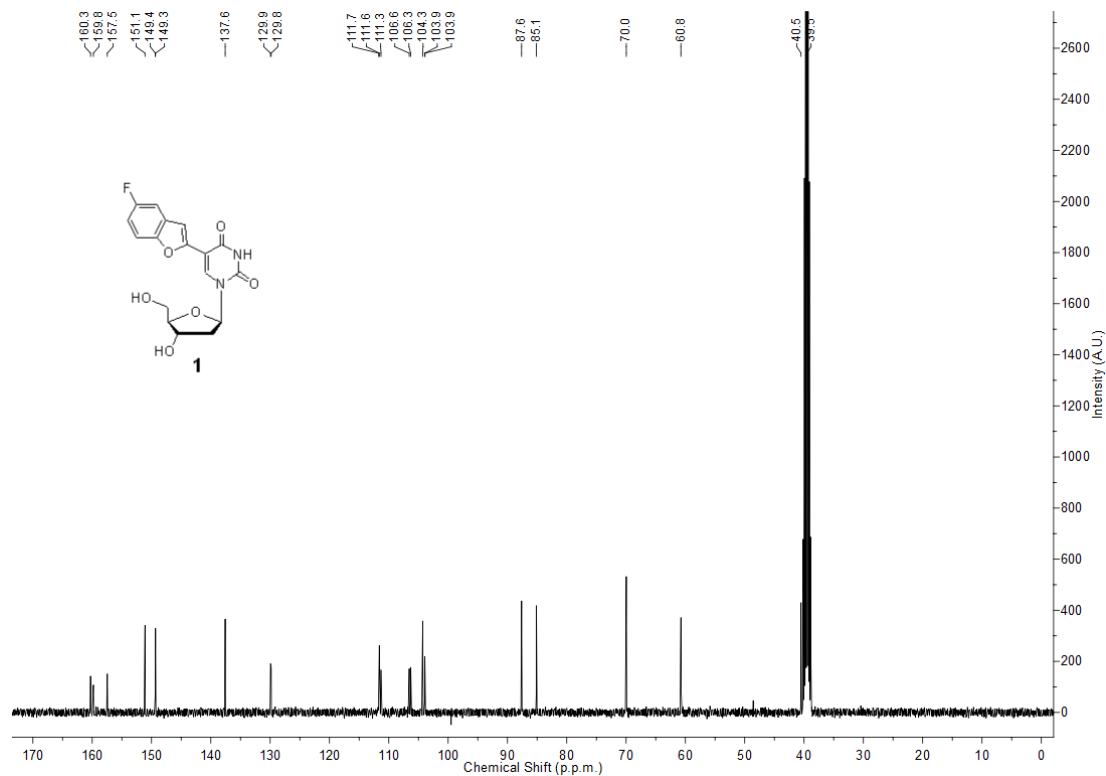
^{19}F NMR of **1** (376.6 MHz, d_6-DMSO)

^{1}H NMR of **2** (400 MHz, CDCl$_3$)

116

¹³C NMR of **2** (100 MHz, CDCl₃)
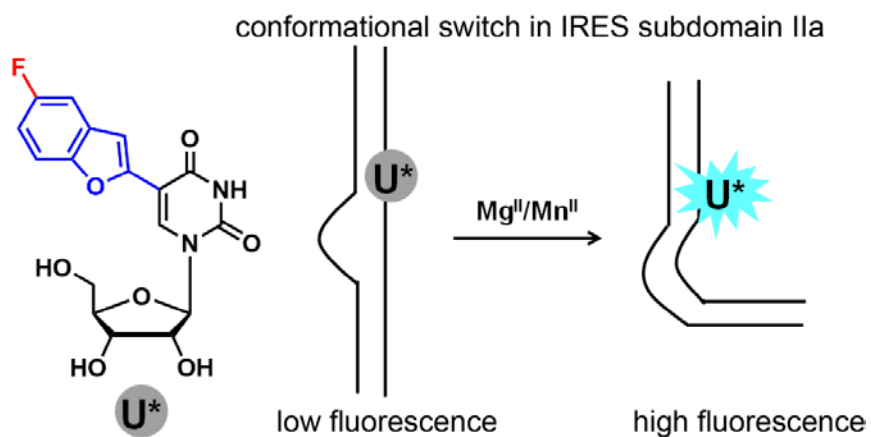

³¹P NMR of **2** (162 MHz, CDCl₃)

117
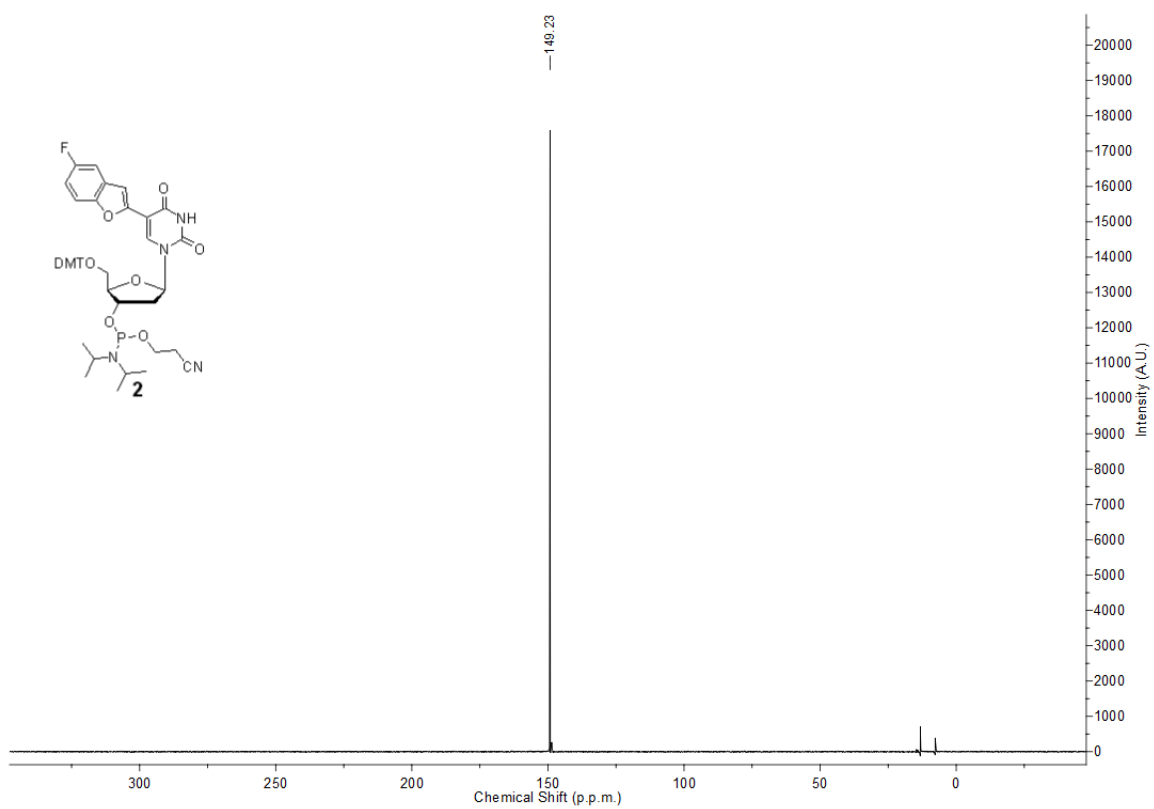
¹⁹F NMR of **2** (376.6 MHz, CDCl₃)
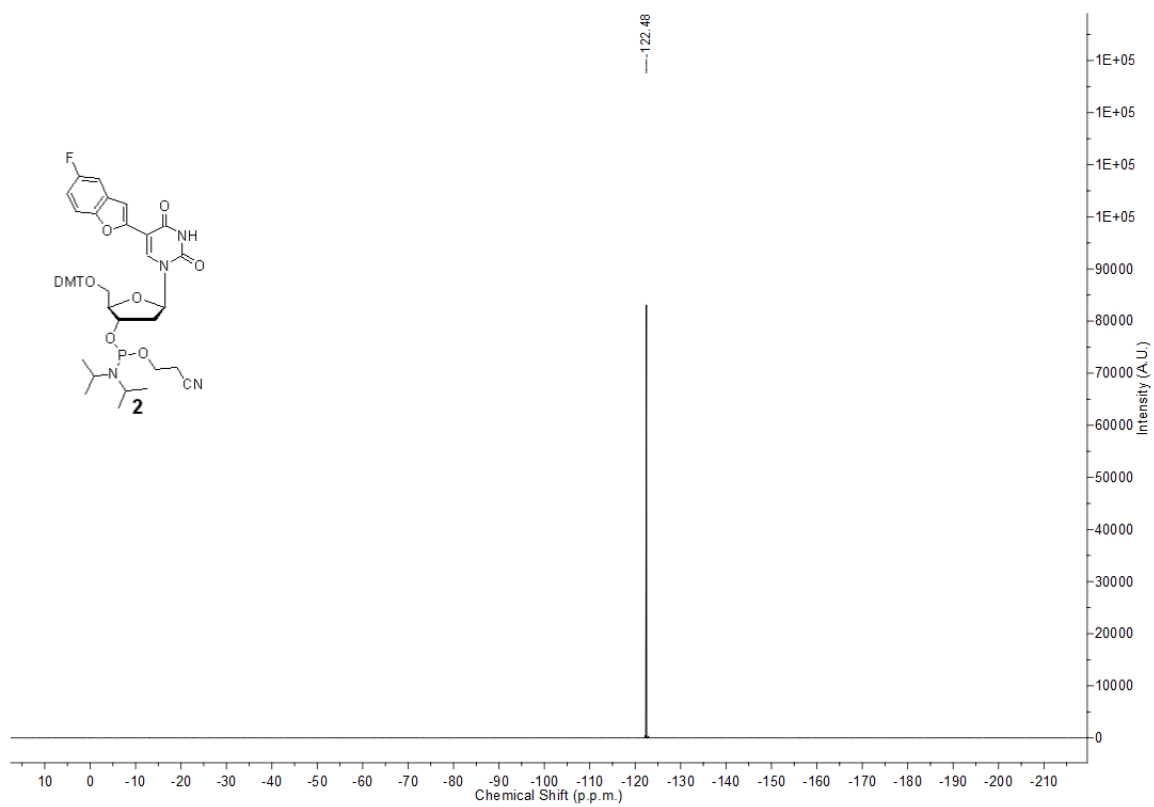
Chapter 4

Enzymatic Incorporation and Utility of Fluorobenzofuran-Modified Uridine in Probing Conformational Change in Viral RNA

4.1 Introduction

The integral role of RNA in key biological processes emanates from its ability to adopt complex structures, which interconvert between different functional conformations.[1] Basic understanding of RNA folding and its interaction with metal ions, proteins, nucleic acids and metabolites has not only provided valuable information on structure-function relationship but has also provided means to target RNA sequences related to disease states. Biophysical techniques like fluorescence, NMR, EPR and X-ray crystallography have been routinely used to study RNA structure, dynamics and functions in real time and atomic level.[2] Notably, fluorescence-based methods employing RNA oligonucleotides (ONs) labeled with fluorescent nucleoside probes have significantly benefited the study of RNA.[3] While, NMR-based approaches have been very useful in the study of RNA, isotope labeling (^{13}C and ^{15}N) of RNA is not straightforward and fairly expensive. ^{19}F NMR has provided alternative avenues to study nucleic acids because of its high sensitivity and 100% natural abundance.[4] ^{19}F NMR has been utilized to study metal ion-induced hammered ribozyme folding,[5] metal ion and peptide binding to HIV TAR RNA,[6,7] RNA secondary structures,[8] RNA-ligand interactions,[9,10] and RNA invasion.[11] Taking cue for these studies, we sought to determine the conformation sensitivity of fluorobenzofuran-modified nucleoside analogue in the context of RNA.

In this Chapter, synthesis, enzymatic incorporation and utility of 5-fluobenzofuran-modified ribonucleoside analogue as a dual-purpose RNA probe is described. The nucleoside analogue shows emission in visible region with a reasonable quantum yield and displays excellent solvatochromism like the 2'-deoxy version.[12] Its ^{19}F chemical shift is also sensitive to changes in solvent polarity and viscosity. The modified ribonucleoside triphosphate substrate was successfully incorporated into RNA transcripts by *in vitro* transcription reaction using T7 RNA polymerase with good to moderate efficiency. The nucleoside phosphoramidite was also compatible for incorporation by conventional solid-phase method. The fluorescence and ^{19}F signal of the probe incorporated into model transcripts were sensitive to flanking bases. The responsiveness of the nucleoside probe was further used in studying metal ion-induced conformational change in internal ribosome entry site (IRES) RNA motif of hepatitis C virus (HCV) by fluorescence and NMR techniques.

4.2 Results and Discussion

### 4.2.1 Synthesis and environment sensitivity of 5-fluorobenzofuran-modified uridine

5-Fluorobenzofuran-modified uridine analogue **1** was synthesized by Stille cross coupling of stanylated 5-fluorobenzofuran with 5-iodo-uridine (Scheme 1).[12] Before incorporation into RNA ONs the environment-sensitivity of the dual-app probe **1** was evaluated by performing photophysical and ^{19}F NMR studies in solvents of different polarity and viscosity. In UV-Vis study, it was observed that the absorption maximum of nucleoside was minimally affected by polarity changes (Figure 1). However, steady-state and time-resolved fluorescence properties of nucleoside **1** were significantly influenced by the solvent polarity. Nucleoside **1** in highly polar solvent water, showed intense emission band (λ_{em} = 437 nm) corresponding to a quantum yield 0.14 and an average lifetime of 0.95 ns (Figure 2, Table 1). In less polar solvent like methanol and dioxane its emission band blue shifted to 420 nm and 400 nm, respectively, and exhibited lower quantum yield and comparatively shorter lifetime (Φ = 0.05 and τ_{av} = 0.81 ns in methanol; Φ = 0.05 and τ_{av} = 0.85 ns in dioxane) (Figure 1A, Figure 2, Table 1). Moreover, the rotatable aryl-aryl bond between uridine and fluorobenzofuran moiety makes the probe conformation-sensitive and the relative orientation of the heterocycle and uracil rings is expected to affect the emission properties of the probe. When the fluorescence studies were performed in solvents with similar polarity but different viscosity it exhibited higher quantum yield and longer lifetime in more viscous solvents (Figure 1B, Figure 2 and Table 1). Furthermore, higher anisotropy value in high viscous solvent proved the rigidification of the fluorophore, which would explain the reason for the enhancement of fluorescence intensity in viscous medium. Additionally, it displayed distinct ^{19}F NMR chemical shift in solvents of different polarity and viscosity (Figure 3). Together these observations reveal that both fluorescence and ^{19}F NMR properties of dual-purpose nucleoside probes **1** are highly environment-sensitive and can be potentially useful in studying RNA conformations and recognition properties.
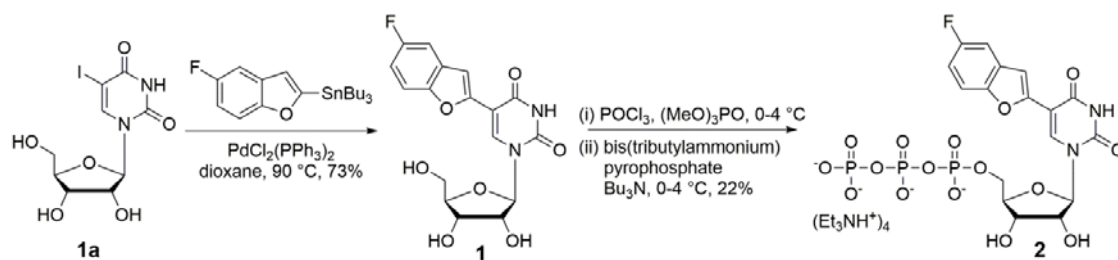
Scheme 1. Synthesis of 5-fluorobenzofuran modified uridine **1** and corresponding triphosphate **2**.
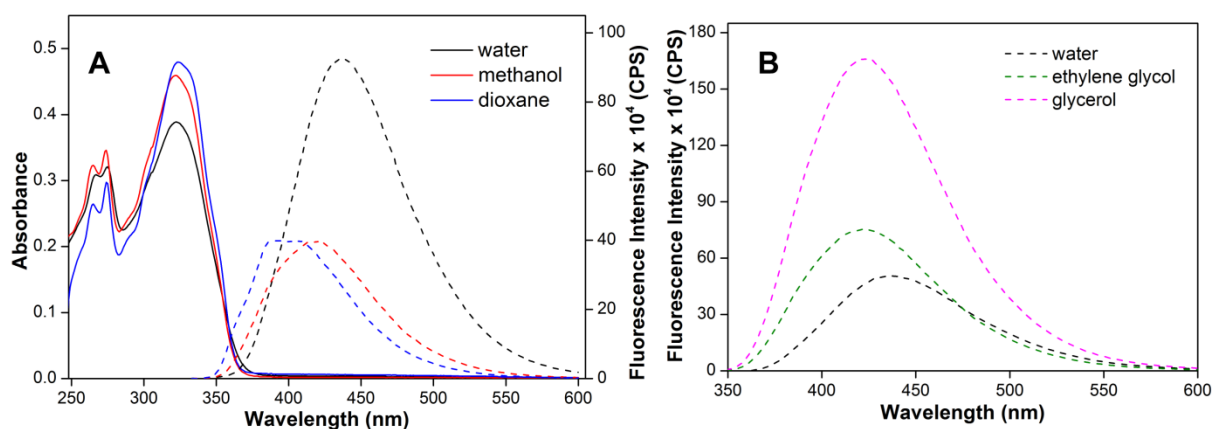
Figure 1. (**A**) UV (25 µM, solid lines) and steady-state fluorescence (5 µM, dashed line) spectra of nucleoside **1** in solvents with different polarity. In fluorescence study, samples were excited at respective lowest energy absorption maxima with excitation and emission slit width 2 nm and 3 nm, respectively. (**B**) Steady-state fluorescence spectra (5 µM, dashed line) of nucleoside **1** in solvents with similar polarity but different viscosity. Samples were excited at respective lowest energy absorption maxima with excitation and emission slit width 2 nm and 2 nm, respectively.
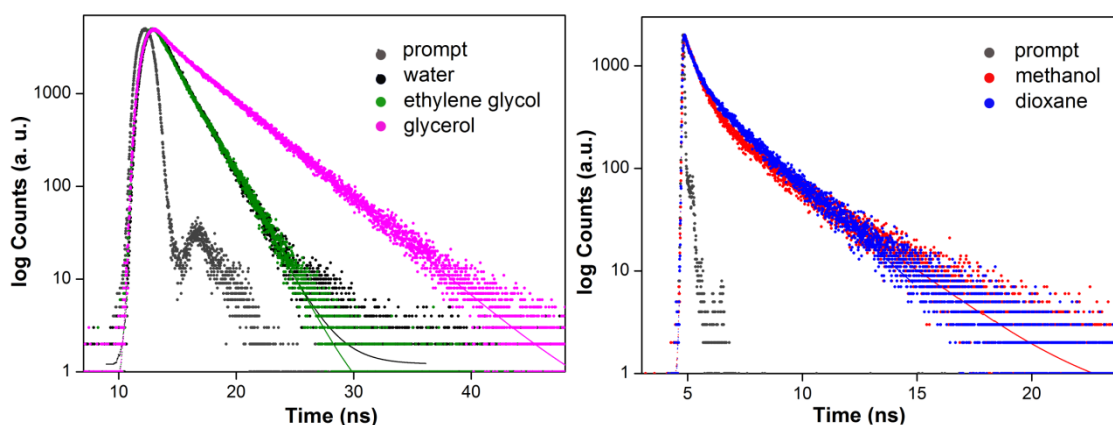

Figure 2. Excited state decay profile of nucleoside **1** in solvents with different polarity and viscosity. Laser response (prompt) has been shown in dark gray line. Curve fits have been shown in solid lines.

122

Table 1. Photophysical properties of nucleoside **1** in different solvents

Solvent	$\lambda_{max}{}^a$ (nm)	λ_{em} (nm)	$I_{rel}{}^b$	Φ^c	$\tau_{av}{}^c$ (ns)	r^c
water	322	437	1.00	0.14	0.95	0.02
methanol	322	420	0.43	0.05	0.81	nd
dioxane	324	400	0.43	0.05	0.85	nd
ethylene glycol	325	423	1.49	0.22	0.99	0.18
glycerol	326	422	3.30	0.54	2.78	0.32

[a]Lowest absorption energy maximum is given. [b]Relative emission intensity at λ_{em} is given with respect to intensity in water. [c]Standard deviation for quantum yield (Φ), average lifetime (τ_{av}) and anisotropy (r) in different solvents is ≤ 0.005, ≤ 0.02 ns and ≤ 0.003 respectively. n.d. = not determined.
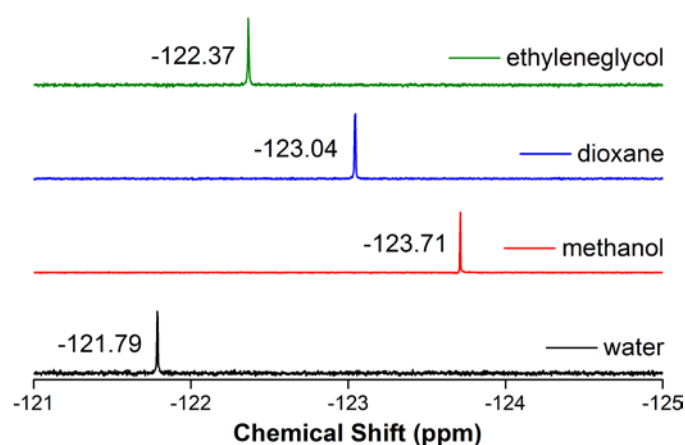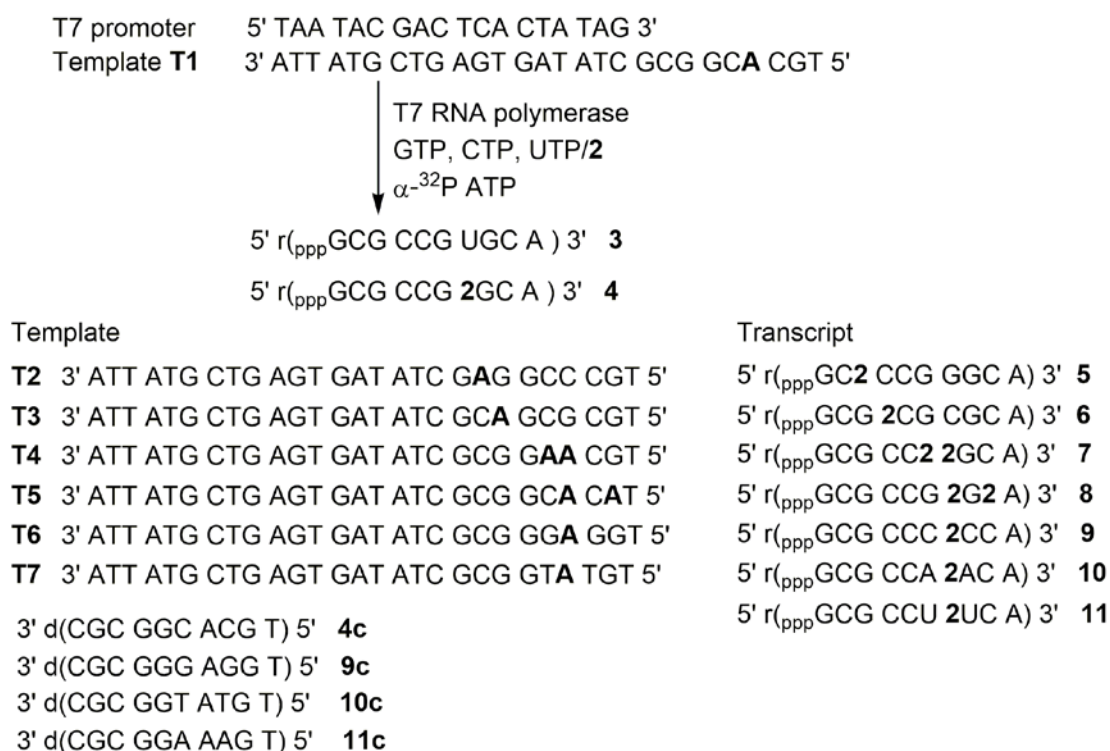

Figure 3. ^{19}F NMR spectra of nucleosides **1** (600 µM) in solvents of different polarity and viscosity. Each NMR sample contained 15% DMSO-*d6* and each spectrum was referenced relative to external standard trifluorotoluene (TFT) (-63.72 ppm).

4.2.2 T7 RNA polymerase efficiently incorporates modified nucleotide into RNA transcripts
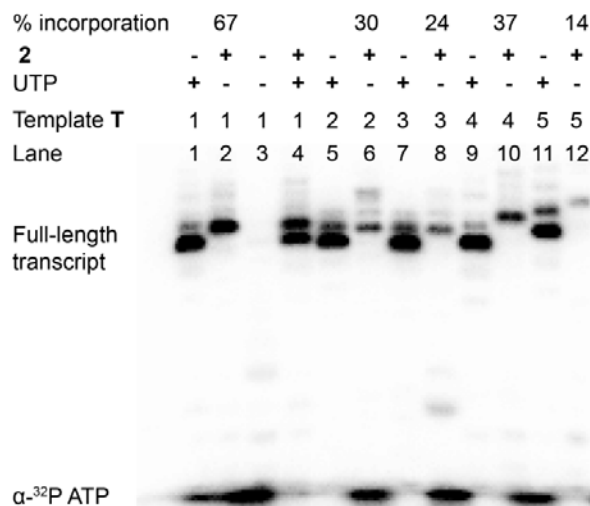
In order to investigate the incorporation efficiency of 5-fluorobenzofuran-modified nucleoside probe **1** by transcription reaction, modified triphosphate **2** was synthesized (Scheme 1).[13] *In vitro* transcription reactions were executed as depicted in Figure 4. A series of templates (**T1**–**T5**), which have one or two deoxy-adenosine residues at different positions to direct the incorporation of uridine (U) or **2** in the transcript were designed and respective duplexes were formed by annealing with a 18 mer T7 RNA polymerase consensus DNA promoter. All transcription reactions were conducted in the presence of GTP, CTP, UTP/ **2** and α-^{32}P ATP. Moreover, in each template, at the 5' end, there was a deoxythymidine

residue, which would direct the incorporation of α-^{32}P- labeled adenosine at the 3' end of each transcript. Hence, each full-length transcript containing α-^{32}P label could be detected by phosphor imaging. Transcripts shorter than the full-length product would not be visible as they would not carry α-^{32}P label.

Percentage incorporation of **2** by transcription reaction in case of each template was calculated relative to the transcription reaction performed in the presence of natural UTP. Transcription in the presence of template **T1** where the dA residue was away from the promoter region yielded the full-length 10 mer ON in good efficiency (67 %, Figure 5, Lane 2) with trace amounts of N+1, N+2 and N+3 bands due to non-templated arbitrary incorporation of ribonucleotides. Additionally, the retardation in the migration of transcript **4** with respect to unmodified transcript **3** clearly represented the incorporation of higher molecular weight ribonucleoside into the transcript. Most importantly, in the absence of both UTP and **2** there was no full-length product, which proved that the formation of the full-length transcript was not due to any misincorporation of ribonucleotides (Figure 5, Lane 3). Interestingly, when the reaction was performed with an equimolar concentration of **2** and natural UTP, T7 RNA polymerase showed similar preference for both the substrates (Figure 5, lane 4). In the presence of template **T2** and **T3** where the dA residue was near the promoter region, +3 and +4 position, respectively, a significant reduction in transcription efficiency was observed (Figure 5, lane 6 and lane 8). Transcription reaction in the presence of template **T4,** which directed two successive incorporations of nucleotide **2** showed moderate efficiency (Figure 5, lane 10). But in the presence of template **T5**, the incorporation of **2** in two alternative positions was quite low (Figure 5, lane 12). Altogether, these reactions pointed out the acceptance of nucleotide **2** by T7 RNA polymerase and usefulness of transcription reaction to synthesize labeled RNA ONs. Further large-scale transcription was performed with the template **T1** to isolate and characterize the labeled RNA transcripts. Gel purified fluorobenzofuran modified transcript was characterized by MALDI-TOF mass analysis and purity of the transcript was examined by HPLC (Figure 6A and 6B). Enzymatic digestion of transcript **4** followed by HPLC analysis of resulting ribonucleoside mixture further confirmed the presence of nucleoside **1** in the transcript (Figure 6C and 6D).

T7 promoter   5' TAA TAC GAC TCA CTA TAG 3'
Template **T1**   3' ATT ATG CTG AGT GAT ATC GCG GC**A** CGT 5'

T7 RNA polymerase
GTP, CTP, UTP/**2**
α-^{32}P ATP

5' r($_{ppp}$GCG CCG UGC A) 3' **3**

5' r($_{ppp}$GCG CCG **2**GC A) 3' **4**

Template

T2 3' ATT ATG CTG AGT GAT ATC G**A**G GCC CGT 5'
T3 3' ATT ATG CTG AGT GAT ATC GC**A** GCG CGT 5'
T4 3' ATT ATG CTG AGT GAT ATC GCG G**AA** CGT 5'
T5 3' ATT ATG CTG AGT GAT ATC GCG GC**A** C**A**T 5'
T6 3' ATT ATG CTG AGT GAT ATC GCG GG**A** GGT 5'
T7 3' ATT ATG CTG AGT GAT ATC GCG GT**A** TGT 5'

3' d(CGC GGC ACG T) 5' **4c**
3' d(CGC GGG AGG T) 5' **9c**
3' d(CGC GGT ATG T) 5' **10c**
3' d(CGC GGA AAG T) 5' **11c**

Transcript

5' r($_{ppp}$GC**2** CCG GGC A) 3' **5**
5' r($_{ppp}$GCG **2**CG CGC A) 3' **6**
5' r($_{ppp}$GCG CC**2** **2**GC A) 3' **7**
5' r($_{ppp}$GCG CCG **2**G**2** A) 3' **8**
5' r($_{ppp}$GCG CCC **2**CC A) 3' **9**
5' r($_{ppp}$GCG CCA **2**AC A) 3' **10**
5' r($_{ppp}$GCG CCU **2**UC A) 3' **11**

Figure 4. Incorporation of triphosphate **2** into RNA ONs by transcription reaction in presence of different DNA templates **T1**–**T7**. ONs **11** was synthesized by solid phase RNA synthesis. Synthesis of ON **9**–**11** by enzymatic and solid phase synthesis was carried out by Dr. Vyankat Sontakke. **4c**, **9c**, **10c** and **11c** are custom DNA ONs used for fluorescence and ^{19}F NMR studies. "r" represent the RNA ONs and "d" represents the DNA ONs.

% incorporation		67				30		24		37		14
2	-	+	-	+	-	+	-	+	-	+	-	+
UTP	+	-	-	+	+	-	+	-	+	-	+	-
Template **T**	1	1	1	1	2	2	3	3	4	4	5	5
Lane	1	2	3	4	5	6	7	8	9	10	11	12

Full-length transcript

α-^{32}P ATP

Figure 5. Polyacrylamide gel electrophoresis of the transcripts obtained from *in vitro* transcription reactions with DNA templates **T1**–**T5** in the presence of UTP and **2** under denaturing conditions. The percentage incorporation of **2** is determined with respect to the amount of full-length product formed in the presence of UTP. All reactions were performed in duplicate and standard deviations in yields were ≤ 4%. See the Experimental Section for details.

125
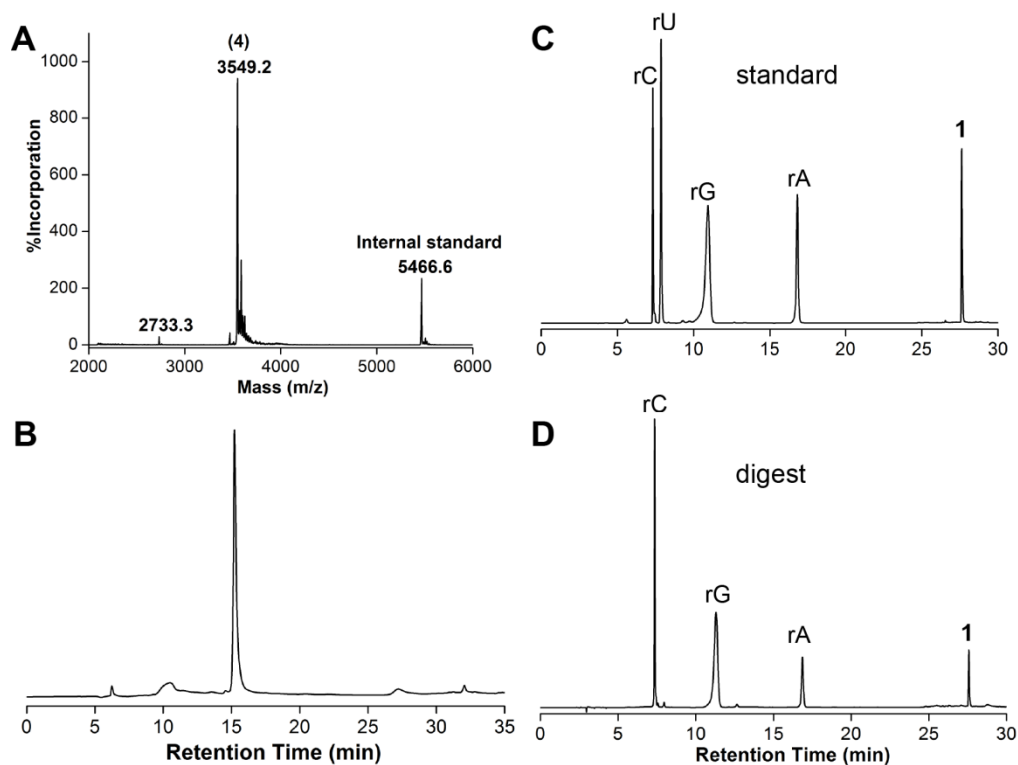
Figure 6. (**A**) MALDI-TOF spectrum of transcript **4** calibrated relative to the +1 and +2 ions of an internal 18-mer DNA ON standard (m/z for +1 and +2 ions are 5466.6 and 2733.3, respectively). m/z calculated for **4**: 3549.0 [M]; found: 3549.2. (**B**). HPLC chromatogram of PAGE purified transcript **4**. Mobile Phase: A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: acetonitrile. Flow rate: 1 mL/min. Gradient: 0-30% B in 25 min and 30-100% B in 10 min. HPLC chromatogram of (**C**) mixture of nucleosides (rC, rU, rG, rA and **1**) and (**D**) ribonucleoside products obtained from digestion of transcript **4** at 260 nm. See experimental section for details.

4.2.3 Nucleoside probe is sensitive to neighbouring base environment

In ONs, the interactions of the probe with different neighbouring base environments can alter its fluorescence properties and ^{19}F NMR chemical shift by different mechanisms. In order to examine the sensitivity of the probe to changes in flanking bases, model RNA ONs **4**, **9**, **10** and **11** were synthesized where the probe **1** was flanked between two rG, rC, rA and rU residues, respectively. ON **4**, **9** and **10** were synthesized by large-scale transcription reaction using the appropriate templates (Figure 4). As in transcription reaction, site-specific incorporation of **2** can't be controlled and T7 RNA polymerase would have a higher preference for natural UTP over **2**, ON **11** with the probe flanked between rU residues, was synthesized by solid-phase RNA synthesis using phosphoramidite **12**. The phosphoramidite **12** was synthesized as depicted in Scheme 2.

Before performing the fluorescence and ^{19}F NMR studies of single-stranded RNA ONs and their respective RNA-DNA duplexes (**4•4c**, **9•9c**, **10•10c** and **11•11c**), the effect of modification on the stability of duplexes was examined by thermal melting analysis. Similar T_m values exhibited by unmodified (**3•4c**) and modified (**4•4c)** duplexes indicated that modification has minimal impact on duplex stability (Figure 7A and Table 2). It was also observed that except **11•11c**, other duplexes were reasonably stable at experimental condition (Table 2). When the fluorescence studies were performed with the single-stranded RNA ONs it was observed that the probe exhibited different fluorescence intensity in different flanking base environment. It showed quenched fluorescence when placed in between rG and rC residues. However, when it was placed in between rA or rU residues there was an enhancement in fluorescence intensity (Figure 7B and 7C). In fluorescence studies of duplexes, it was observed that duplexes of ON **10** and **11** (**10•10c** and **11•11**c, respectively) exhibited significantly higher fluorescence compared to respective single-stranded RNA ONs (Figure 7B and 7C). Further, in ^{19}F NMR studies, the probe showed distinct ^{19}F NMR peak in the different flanking base environment and each duplex showed chemical shift in the shielded region compared to respective single-stranded RNA ON (Figure 7D). The duplex **11•11c** did not provide well-resolved ^{19}F NMR peak, which could be because of its less thermal stability (Table 2). Collectively, these results demonstrated the sensitivity of the probe to neighbouring base environment and prompted us to incorporate it into a therapeutically important RNA ON, IRES element of Hepatitis C virus to study its conformation and recognition properties.
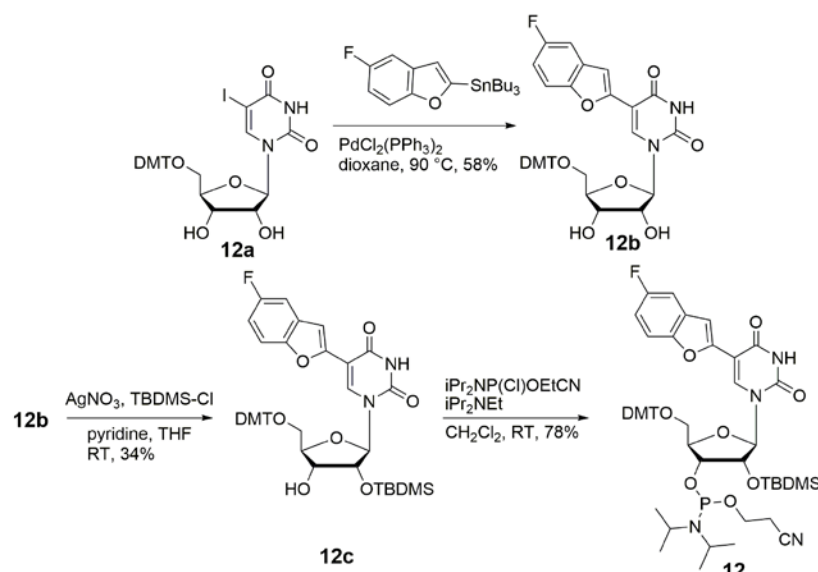
Scheme 2. Synthesis of 5-fluorobenzofuran modified uridine phosphoramidite **12**. DMT = 4,4'-dimethoxytrityl, TBDMS = *tert*-butyldimethylsilyl, THF = Tetrahydrofuran.
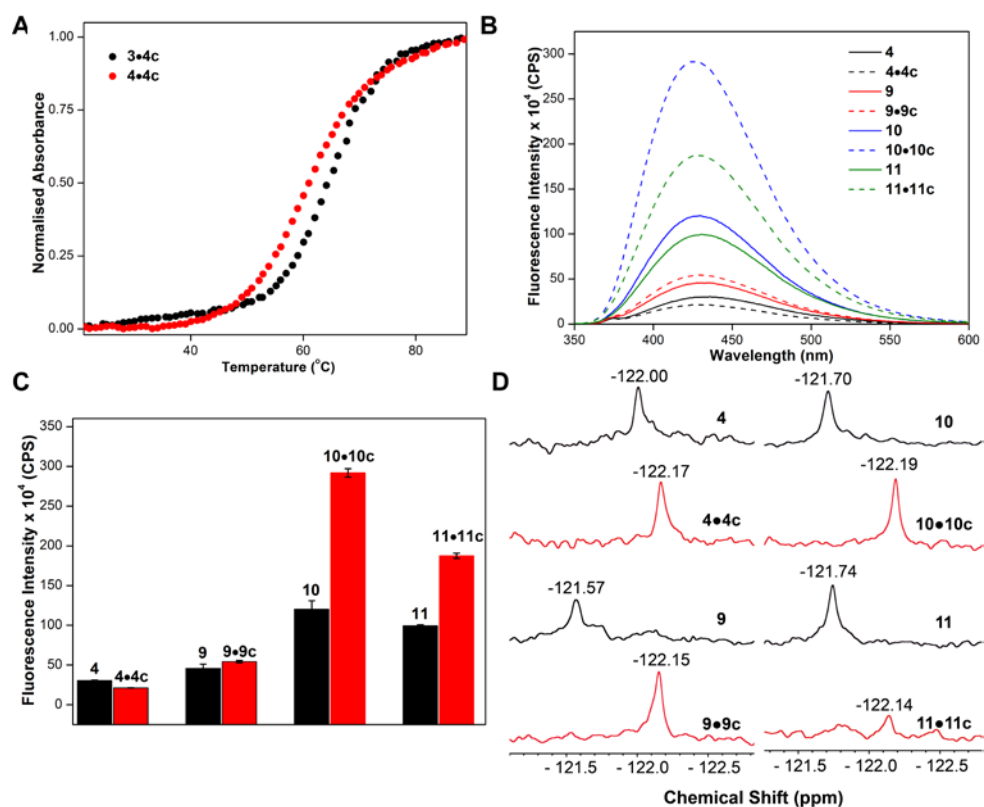
Figure 7. (A) Representative UV-thermal melting profile of control (**3•4c**) (5 µM) and modified (**4•4c**) (5 µM) RNA-DNA duplexes in sodium phosphate buffer (pH 7.1) containing 500 mM NaCl. (**B**) Emission spectra (0.5 µM), (**C**) fluorescence intensity and (**D**) ^{19}F NMR spectra of RNA ONs **4**, **9–11** and their duplexes (50 µM) with complementary DNA sequences in sodium phosphate buffer containing 500 mM NaCl. In emission profile, the spectra corresponding to single-stranded RNA have been represented in solid lines and the spectra corresponding to duplexes have been represented in dotted line. Fluorescence analysis was performed by Dr. Vyankat Sontakke.

Table 2. T_m values of RNA-DNA duplexes

duplex	$T_m(°C)$
3•4c	65.8± 0.7
4•4c	61.8± 0.6
9•9c	33.4± 2.4
10•10c	50.8± 0.5
11•11c	26.6± 1.9

Thermal melting analysis was conducted with duplexes (5 µM) in sodium phosphate buffer (pH 7.1) containing 500 mM NaCl Thermal melting experiments were performed by Dr. Vyankat Sontakke

4.2.4 Probe reports metal-mediated conformational change in HCV IRES element

Hepatitis C virus (HCV), a human pathogen, is an RNA virus and the infection caused by this virus is a serious health threat worldwide.[14] HCV contains highly structured internal ribosome entry site (IRES) at the 5' end of its RNA genome, which binds to the host 40S ribosomal subunit and initiates cap-independent translation process.[15] The IRES structural element is composed of four independently folding domains.[16] Among them domain II is highly conserved in HCV clinical isolates[17] and plays a key role in crucial steps of translation for example positioning of HCV mRNA at the ribosomal decoding site,[18] dissociation of initiation factor from 40S subunit to promote 80S ribosomal assembly[19] and transition from initiation to elongation steps.[20] Particularly subdomain IIa (Figure 8A), a RNA conformational switch, adopts a 90° bent structure, which is stabilized by divalent metal ions like Mg^{II} and Mn^{II} and this conformation is essential for translation initiation.[21] Because of this, subdomain IIa has been identified as an attractive target for small molecule translation inhibitor in IRES-targeted antiviral drug development.[17] Benzimidazoles-based ligands have been shown to inhibit viral translation by locking extended conformation of IIa subdomain.[22] High resolution three dimensional structure of IIa for both metal ion stabilized bent and benzimidazole-locked straightened conformation have been obtained by X-ray and NMR spectroscopy.[23] The L-shaped structure of subdomain IIa has a fold at internal bulge, which is flanked by two stems arranged at a right angle to each other (Figure 8B).[21b] Additionally, a looped-out U106 residue at the upper stem is a characteristic feature of the bent structure. The whole L-shaped structure is stabilized by base stacking, H-bonding and direct interaction with the metal ions. Moreover, in the presence of benzimidazole ligand, the structure becomes straightened where the internal bulge refolds from its bent conformation and gets

flanked by two coaxial helices (Figure 8C).[22b] Subdomain IIa construct labeled with two cyanine dye, a FRET pair, at the two terminals reports the metal-mediated RNA folding by an increase in FRET signal and ligand-induced straightening of the structure by a reduction in FRET signal.[24,22a] Apart from global structural studies, site specifically **2** aminopurine (2AP)-labelled fluorescent RNA has been used to quantify the binding affinity of Mg^{II} and Mn^{II} at different sites of the structural element.[21b] Based on these understanding, we aimed to use the responsiveness of our dual-app probe in monitoring metal ion-induced conformational change in the IRES subdomain IIa.
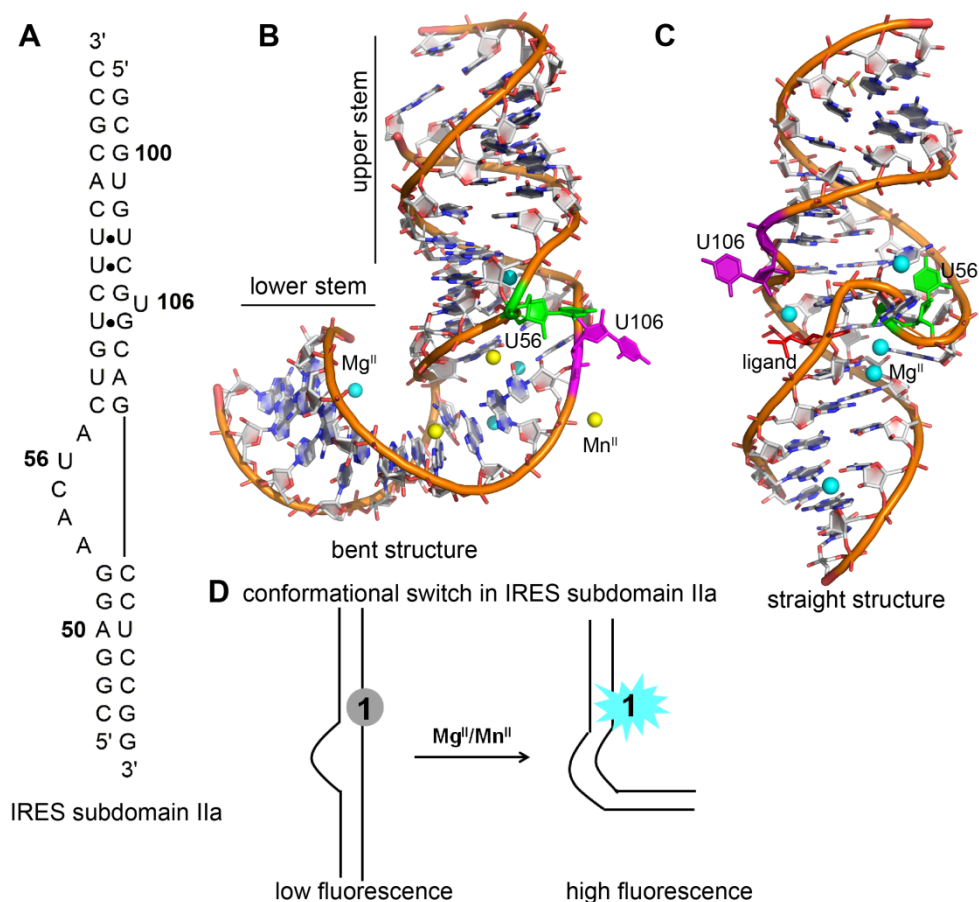
Figure 8. (**A**) Subdomain IIa of HCV IRES element.(**B**) Crystal structure (PDB ID: 2NOK)[21b] of L-shaped bent conformation of IRES subdomain IIa stabilized by Mg^{II} and Mn^{II}. (C) Crystal structure (PDB ID: 3TZR)[24] of ligand-induced straight conformation of IRES subdomain IIa. U106 and U56 which are replaced with nucleoside **1**, have been represented in magenta and green color, respectively. Mg^{II} and Mn^{II} are represented as cyan and yellow color sphere, respectively. Benzimidazole ligand is represented in red color. (**D**) Nucleoside probe **1** reports the metal-ion stabilized bent conformation of IRES subdomain IIa by showing turn-on fluorescence.

4.2.4. 1 Incorporation of Nucleoside 1 into IRES subdomain IIa

ONs **13** and **14** were synthesized by solid-phase method using phosphoramidite **12** so that after annealing with complementary RNA sequences (**13c** and **14c**, respectively) they can form IRES subdomain IIa duplexes **13•13c** and **14•14c** having the modification at U56 and U106 positions, respectively (Figure 9). We purposely incorporated the probe in these positions (U106 in the stem region and U56 in the bulge region) as crystal structures revealed substantial conformational change in these residues during metal ion-mediated bending and ligand-induced straightening of IRES element (Figure 8B and 8C). The labeled ONs were purified by PAGE and characterized by MALDI-TOF mass analysis (Figure 10A and 10B).
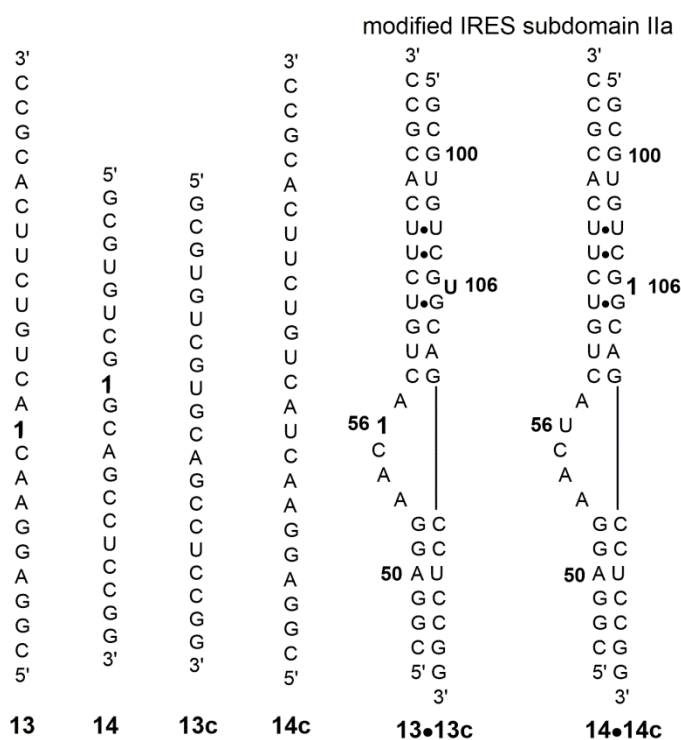
Figure 9. Fluorobenzofuran modified RNA ON **13** and **14** and their complementary RNA sequence **13c** and **14c**. They form respective modified IRES domain IIa duplexes, **13•13c** and **14•14c** where U56 and U106 residues were replaced with nucleoside **1**, respectively.
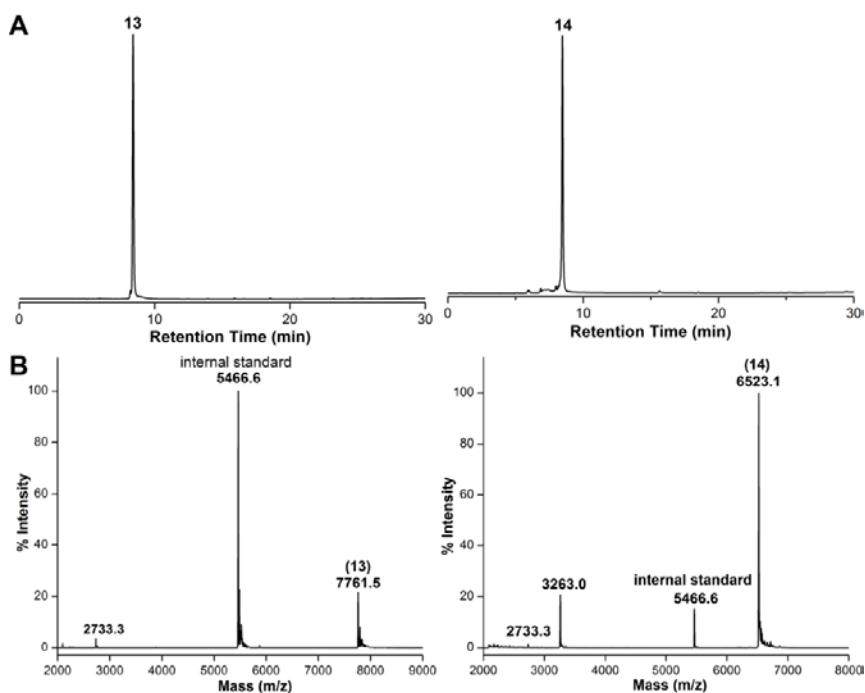
Figure 10. (**A**) HPLC chromatogram of PAGE purified RNA **13** and **14** at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0–100% B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron). (**B**) MALDI-TOF spectrum of modified RNA ONs calibrated relative to the +1 and +2 ions of an internal 18-mer DNA ON standard (m/z for +1 and +2 ions are 5466.6 and 2733.3, respectively). m/z calculated for **13**: 7761.7 [M]; found: 7761.5. m/z calculated for **14**: 6524.0 [M]; found: 6523.1.

4.2.4.2 Quantification of metal ion affinity to different site of IRES subdomain IIa

IRES subdomain IIa duplexes (**13•13c** and **14•14c**), preannealed in sodium cacodylate buffer (10 mM, pH 6.5), were titrated with increasing concentration of metal ions (Mg^{II} and Mn^{II}) and fluorescence spectrum was recorded at each titration point. Fluorescence enhancement in a dose-dependent manner was observed upon titrating with Mg^{II} and Mn^{II} ion (Figure 11 and 12). The extent of enhancement was larger in case of duplex **14•14c** than the duplex **13•13c**. A plot of normalized fluorescence intensity *versus* metal ion concentration provided apparent dissociation constant K_d for the binding of metal ions to IRES. The K_d values determined for **13•13c** were lower than that determined for the duplex **14•14c,** which suggests the difference in the binding affinity of metal ions to different sites of IRES subdomain IIa (Table 5). Different affinity of metal ions to different binding site of IRES subdomain IIa has been reported earlier in the literature.[21b] The crystal structure of metal stabilized IRES subdomain IIa bent structure reveals that conformation of U56 in bulge region and conformation of U106

in stem region is different which could be the reason of different response at different site. The significant fluorescence enhancement in case of modification at U106 can be attributed to the expulsion of this uridine residue from double helix in bent structure which results exposure of the probe to bulk water and destacking of it. Additionally, it was observed that in each binding site, there was not much of a difference in the binding affinity for two metal ions (Mg^{II} and Mn^{II}) (Table 5). Moreover, control binding experiments with nucleoside **1** showed only minor changes in fluorescence intensity which indicated that the nucleoside signalled only the metal-mediated conformational change in HCV IRES subdomain IIa (Figure 13). Taken together, the probe was successful in reporting the Mg^{II} and Mn^{II} stabilized L-shaped bend structure of IRES subdomain IIa and also enabled the quantification of metal ion binding by means of changes in fluorescence intensity.
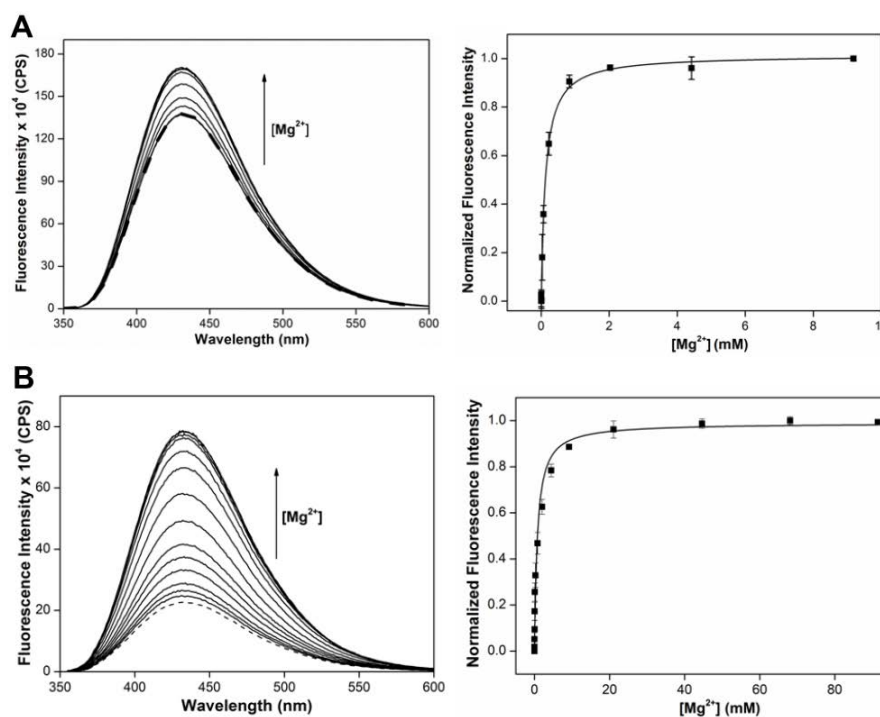
Figure 11. Emission spectra and corresponding curve fit for titration of IRES subdomain IIa duplexes (**A**) **13•13c** and (**B**) **14•14c** (0.5 µM) in sodium cacodylate buffer (10 mM, pH 6.5) with increasing concentration of Mg^{II} ion. The dashed line represents the spectrum of ON duplexes in absence of Mg^{II}.
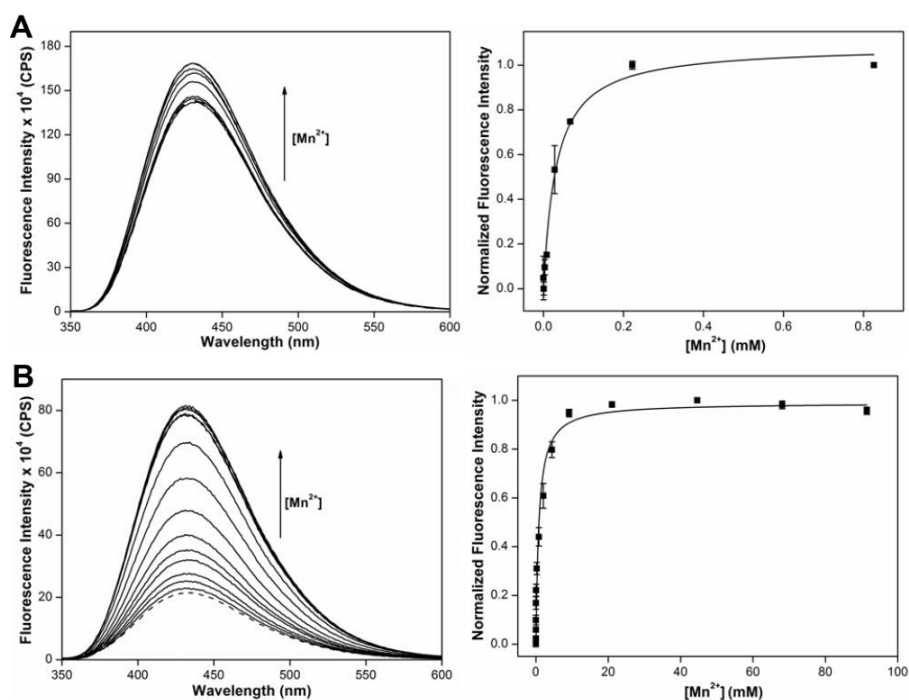
Figure 12. Emission spectra and corresponding curve fit for titration of IRES subdomain IIa duplexes (0.5 µM) (**A**) **13•13c** and (**B**) **14•14c** (0.5 µM) in sodium cacodylate buffer (10 mM, pH 6.5) with increasing concentration of Mn^{II} ion. The dashed line represents the spectrum of ON duplexes in absence of Mn^{II} ion.

**Table 5.** Dissociation constant (K_d) for Mg^{II} and Mn^{II} binding to duplexes (**13•13c** and **14•14c**)

metal ion	K_d (mM)	
	13•13c	**14•14c**
Mg^{II}	0.130 ± 0.020	0.773 ± 0.153
Mn^{II}	0.032 ± 0.005	0.833 ± 0.157
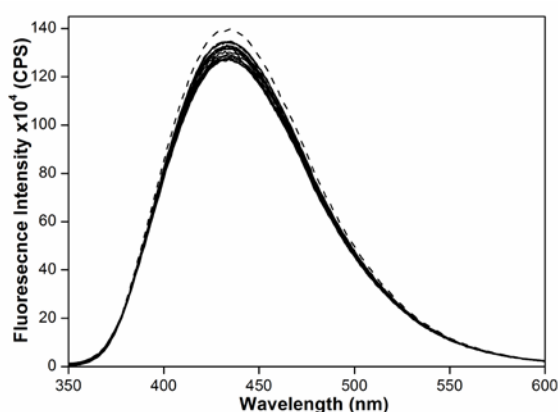
Figure 13. Emission spectra of nucleoside **1** (0.5 µM) with increasing concentration of Mg^{II} ion. The dashed line represents the spectrum of nucleoside **1** in absence of Mn^{II} ion.

4.2.4.3 ^{19}F NMR studies of modified IRES subdomain IIa

As duplex **14•14c** displayed significant fluorescence response upon metal ion binding, we performed ^{19}F NMR study of this duplex in the presence and absence of Mg^{II} ion. Slight change in chemical shift and discernible sharpening of the ^{19}F NMR peak in the presence of Mg^{II} ion suggest that the nucleoside probe at U106 position should have undergone conformational change as a result of bending of the domain (Figure 14). Due to the paramagnetic nature of Mn^{II}, IRES-Mn^{II} complex did not give any ^{19}F NMR signal.
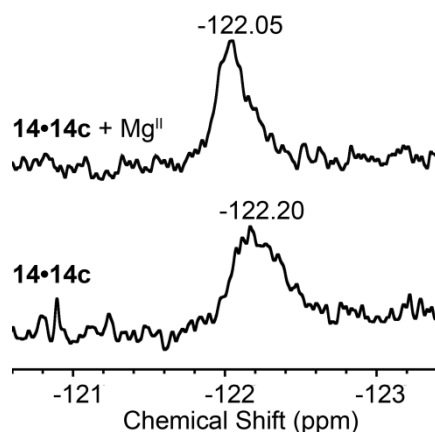
Figure 14. ^{19}F NMR study of fluorobenzofuran-modified IRES subdomain IIa duplex **14•14c** (50 µM) in 10 mM sodium cacodylate buffer (pH 6.5) in absence and presence of 5 mM Mg^{II}.

Based on this preliminary data, we have decided to perform ^{19}F NMR relaxation experiment in future using the ultrasensitive ^{19}F cryoprobe, which will help in understanding the dynamics of this RNA structural element at nucleoside level. We also plan to examine whether the dual-purpose probe can sense ligand-induced straightening of the IRES subdomain so that one can use the nucleoside probe to screen for small molecules binding to IRES.

4.3 Conclusions

We have introduced a dual-app ribonucleoside analogue, which acts as an environment-sensitive fluorescence and ^{19}F NMR probe. Fluorescence quantum yield, emission maximum, excited-state lifetime and ^{19}F NMR chemical shift of the probe were sensitive to polarity and viscosity changes. We have been successful in incorporating the dual-app nucleoside probe into RNA ONs by both transcription reaction and solid-phase methods. Interestingly, after incorporation into RNA ONs, the fluorescence and ^{19}F NMR properties of the nucleoside

probe were responsive to its flanking bases. Further, using the responsiveness of the probe we have developed a method to detect an important metal ion-mediated conformational change in HCV viral RNA, an attractive target in antiviral drug discovery. Taken together, the probe provides a potential biophysical platform to investigate RNA structure and recognition. Efforts to study RNA dynamics by ^{19}F NMR relaxation experiments using the nucleoside probe are currently in progress. Moreover, the scope of this probe in establishing a bio-physical platform to screen small-molecule based antiviral drugs will be explored in future.

4.4 Experimental Section

4.4.1 Materials

N,N,N',N'-tetramethylethylenediamine, *n*-butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) dichloride, *tert*-butyldimethylsilyl chloride, silver nitrate, *N,N*-diisopropylethylamine, and all reagents (Bio-Ultra grade) used for buffer preparation were purchased from Sigma-Aldrich. 5-iodo uridine (**1a**),[25] 5'-DMT-protected 5-iodouridine (**12a**)[25] and Stanylated 5-fluorobenzofuran (**1b**)[12] were synthesized following a previously reported procedure. PoCl$_3$ was purchased from Acros Organic and was freshly distilled before use. T7 RNA polymerase, ribonuclease inhibitor (RiboLock), NTPs, RNase A and RNase T1 were acquired from Fermentas, Thermo Fisher Scientific. Calf intestinal alkaline phosphatase (CIP) and snake venom phosphodiesterase I were acquired from Invitrogen and Sigma-Aldrich, respectively. Radiolabeled α-^{32}P ATP (2000 Ci/mmol) was obtained from the Board of Radiation and Isotope Technology, Government of India. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite was purchased from Alfa Aesar. TBDMS-protected ribonucleoside phosphoramidites and solid support for RNA synthesis were obtained from Glen Research. All other reagents, required for solid-phase oligonucleotide (ON) synthesis were purchased either from ChemGenes Corporation or from Sigma-Aldrich. RNA ONs **9**, **10**, and **11** were synthesized by Dr. Vyankat Sontakke either by enzymatic or solid phase RNA synthesis. DNA ONs **4c**, **9c**, **10c** and **11c** were purchased from Integrated DNA Technology. RNA oligonucleotide **13c** and **14c** were purchased from Dharmacon RNAi Technologies and deprotected according to the provider's procedure. All custome DNA and RNA oligonucleotides were purified by denaturing polyacrylamide gel and desalted on Sep-Pak Classic C18 cartridges (Waters Corporation). Autoclaved Millipore water was used in all biophysical analysis.

4.4.2 Instrumentation

Modified RNA ONs were synthesized on an ABI applied Biosystems 392 DNA/RNA synthesizer. Mass measurements were conducted either on an Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer or on a Water Synapt G2 High Definition mass spectrometers. HPLC analysis was done using Agilent Technologies 1260 Infinity HPLC. Reverse-phase flash chromatographic (C18 RediSepRf column) purifications were carried out using Teledyne ISCO, Combi Flash Rf. Absorption spectra were recorded on a Shimadzu UV-2600 spectrophotometer. Steady-state and time-resolved fluorescence spectra were recorded on a Fluoromax-4 spectrofluorometer. UV-thermal melting analysis of ONs was carried out on Cary 300 Bio UV-Vis spectrophotometer. NMR spectra of small molecules were recorded on a Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. NMR spectra of ONs were recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe and processed using Bruker TopSpin Software.

4.4.3. Synthesis of 5-fluorobenzofuran modified uridine 1, corresponding triphosphate 2 and phosphoramidite 12

4.4.3.1 5-fluorobenzofuran modified uridine (1)

A mixture of 5-iodouridine (0.6 g, 1.5 mmol, 1 equiv) and *bis*(triphenylphosphine)-palladium(II) dichloride (0.05 g, 0.075 mmol, 0.05 equiv) was dissolved in degassed anhydrous dioxane (30 ml) and tributyl(5-fluorobenzofuran-2-yl)stannane (1.9 g, 4.5 mmol, 3 equiv) was added to it. The mixture was heated at 90 °C for 2 h under N_2 atmosphere and filtered through celite pad. The celite pad was washed with methanol (2 × 15 ml). The filtrate was evaporated and resulting residue was purified by reverse phase column chromatography (50 % methanol in water) to afford product **1** as white solid compound (0.45 g, 73%). TLC (15% methanol in DCM) R_f = 0.46; ^{1}H NMR (400 MHz, d_6-DMSO): δ (ppm) 11.70 (br s, 1H), 8.93 (s, 1H), 7.56 (dd, J_1= 9 Hz, J_2 = 2.8 Hz, 1H), 7.43 (dd, J_1 = 8.8 Hz, J_2 = 2.8 Hz, 1H), 7.33 (br, 1H), 7.11 (td, J_1 = 9.2 Hz, J_2 = 2.4 Hz, 1H), 5.87 (d, J = 4 Hz, 1H), 5.51 (br, 1H), 5.41 (br, 1H), 5.14 (br, 1H), 4.16–4.14 (m, 1H), 4.11–4.08 (m, 1H), 3.97–3.96 (m, 1H), 3.83–3.80 (m, 1H), 3.70–3.67 (m, 1H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 160.4, 159.9, 157.5, 151.2, 149.7, 149.4, 137.9, 130.0, 129.9, 111.8, 111.7, 111.7, 111.4, 106.6,

106.4, 104.4, 104.0, 104.0, 88.9, 84.6, 74.5, 69.3, 60.0; ^{19}F NMR (376.6 MHz, d_6-DMSO): δ (ppm) -121.81; HRMS: m/z Calculated for $C_{17}H_{16}FN_2O_7[M+H]^+$ = 379.0942, found = 379.0919; λ_{max} (H$_2$O) = 267, 275 and 322 nm, ε_{267} = 12360 M^{-1} cm^{-1}, ε_{275} = 12840 M^{-1} cm^{-1}, ε_{322} = 15546 M^{-1} cm^{-1}, ε_{260} = 10360 M^{-1} cm^{-1}

4.4.3.2 5-fluorobenzofuran modified uridine triphosphate (2)

Freshly distilled POCl$_3$ (60 µL, 0.64 mmol, 2.5 equiv) was added to a ice-cold solution of 5-fluorobenzofuran modified uridine (96 mg, 0.25 mmol, 1.0 equiv) in trimethylphosphate (1.5 ml). The solution was allowed to stir for 24 h at ~ 4 °C. After 24 h, a solution of bis(tributylammonium) pyrophosphate (0.5 M in DMF, 2.5 ml, 1.27 mmol, 5.0 equiv) and tributylamine (0.67 ml, 2.79 mmol, 11 equiv) were added simultaneously to the solution in ice-cold condition. The reaction was stirred for 30 min at ~ 4 °C. The reaction was quenched with 1 M triethylammonium bicarbonate buffer (TEAB, pH 7.5, 15 ml) and washed with ethylacetate (2 x 15 ml). The aqueous layer was evaporated to dryness and purified using an DEAE sephadex-A25 anion exchange column (10 mM–1M TEAB buffer, pH 7.5) followed by reverse phase flash chromatography (C18 Redi*SepRf*, 0–30% acetonitrile in 50 mM triethylammonium acetate buffer, pH 7.2, flow rate 7 ml/ min, 40 min). Evaporation of the appropriate fraction resulted into the desired triphosphate **2** as tetraethyl ammonium salt (58 mg, 22 %). ^{1}H NMR (400 MHz, D$_2$O): 8.03 (s, 1H), 7.15–7.13 (m, 1H), 7.02–6.98 (m, 2H), 5.82 (s, 1H), 4.44 (br, 2H), 4.29 (br, 3H); ^{13}C NMR (100 MHz, D$_2$O): δ (ppm) 160.97, 160.14, 157.80, 150.07, 149.58, 149.10, 136.91, 129.43, 129.32, 112.28, 112.03, 106.51, 106.26, 105.84, 105.07, 89.25, 83.42, 73.44, 69.83, 65.43, 58.84; ^{31}P NMR (162 MHz, D$_2$O): -6.00 (br, P$_\gamma$), -11.12 (d, J = 18.95 Hz, P$_\alpha$), -21.94 (br, P$_\beta$); ^{19}F NMR (376.6 MHz, D$_2$O): δ (ppm) -122.32; HRMS: m/z Calculated for $C_{17}H_{17}FN_2O_{16}P_3$ [M-H] = 616.9775, found = 616.9771.

4.4.3.3 5-fluobenzofuran modified 5'- *O*- DMT-uridine (12b)

A mixture of 5'-DMT-protected 5-iodouridine **12a**[25] (1.5 g, 2.2 mmol, 1 equiv) and *bis*(triphenylphosphine)-palladium(II) dichloride (0.08 g, 0.11 mmol, 0.05 equiv) was dissolved in degassed anhydrous dioxane (40 ml) and tributyl(5-fluorobenzofuran-2-yl)stannane (2.8 g, 6.7 mmol, 3 equiv) was added to it. The mixture was heated at 90 °C for 2 h under N$_2$ atmosphere and filtered through celite pad. The celite pad was washed with methanol (2 × 20 ml). The filtrate was evaporated and resulting residue was purified by

column chromatography (2 % methanol in DCM containing 0.5 % Et$_3$N) to afford product **12b** as off white foam (0.6 g, 58%). TLC (15% methanol in DCM) R_f = 0.42; ^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 8.53 (s,1H), 7.47–7.45 (m, 2H), 7.34–7.32 (m, 4H), 7.23 (s, 1H), 7.18 (t, J = 7.6 Hz, 1H), 7.07 (t, J = 7.2 Hz, 1H), 6.91–6.89 (m, 1H), 6.69 (d, J = 8.4 Hz, 4 H), 6.61–6.56 (m, 1H), 6.06 (d, J = 4.8, 1H), 5.97 (dd, J_1 = 8.8 Hz, J_2 = 3.6 Hz, 1H), 4.58 (t, J = 5.0 Hz, 1H), 4.37 (t, J = 4.4 Hz, 1H), 4.28 (br, 1H), 3.66 (br, 1H), 3.62 (s, 6H), 3.27–3.25 (m, 1H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 160.6, 160.3, 158.6, 157.9, 150.7, 149.6, 148.9, 144.5, 136.0, 135.7, 135.6, 130.2, 130.1, 129.5, 129.4, 128.3, 128.0, 127.0, 113.3, 113.3, 111.8, 111.6, 111.4, 111.3, 107.0, 106.2, 106.0, 106.0, 89.9, 86.9, 84.8, 75.7, 70.7, 62.8, 55.2, 55.2; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -122.47; HRMS: m/z Calculaed. for C$_{38}$H$_{33}$FN$_2$NaO$_9$ [M+Na]$^+$ = 703.2068, found = 703.2072.

4.4.3.4 5-(5-fluobenzofuran) modified 2'- *O*- TBDMS protected uridine (12c)

Compound **12b** (0.59 g, 0.866 mmol, 1.0 equiv) and silver nitrate (0.32 g, 1.9 mmol, 2.2 equiv) were dissolved in 3 ml anhydrous pyridine and was added 8.4 ml dry THF. The mixture was stirred for 10 min and *tert*-butyldimethylsilyl chloride (0.29 g, 1.9 mmol, 2.25 equiv) was added to it. The reaction mixture was allowed to stir for 0.5 h and filtered through celite pad. Celite pad was washed with ethyl acetate (2 x 10 ml) and organic layer was washed with 5% sodium bicarbonate (25 ml) and brine solution (25 ml) successively. Solvent was evaporated and crude was purified 3 times successively by column chromatography (30% EtOAc in hexane) to afford compound **12c** as white solid (0.25 g, 34%). TLC(40% EtOAc in hexane containing few drops of Et$_3$N) R_f = 0.64; ^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 8.60 (s, 1H), 8.39 (s, 1H), 7.55–7.52 (m, 2H), 7.41–7.37 (m, 5H), 7.20–7.17 (m, 2H), 7.10–7.05 (m, 2H), 6.72–6.69 (m, 4H), 6.50 (td, J_1 = 9.0 Hz, J_2= 2.7 Hz, 1H), 6.17 (d, J = 6 Hz, 1H), 5.53 (dd, J_1 = 8.8 Hz, J_2 = 4.0 Hz), 4.57 (t, J = 5.4 Hz), 4.21–4.20 (m, 1H), 4.14–4.11 (m, 1H), 3.73 (dd, J_1 = 11 Hz, J_2= 1.8 Hz, 1H), 3.65 (s, 3H), 3.63 (s, 3H), 3.20 (dd, J_1 = 11.2 Hz, J_2= 2.4 Hz, 1H), 2.64 (d, J = 3.6 Hz, 1H), 0.90 (s, 9H), 0.13 (s, 3H), 0.11 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 159.6, 158.7, 158.7, 149.7, 149.2, 149.0, 144.7, 135.9, 135.5, 135.4, 130.2, 128.3, 128.0, 127.1, 113.4, 113.3, 111.9, 111.6, 111.5, 111.5, 107.2, 106.3, 106.2, 106.2, 106.2, 106.0, 87.9, 87.1, 84.3, 76.2, 71.3, 63.2, 55.3, 55.3, 25.8, 18.1, -4.6, -5.0; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -122.42; HRMS: m/z Calculaed. for C$_{44}$H$_{47}$FN$_2$NaO$_9$Si [M+Na]$^+$ = 817.2933, found = 817.2930.

4.4.3.5 5-(5-fluobenzofuran) modified uridine phosphoramidite substrate 12

To a solution of **12c** (0.27 g, 0.34 mmol, 1.0 equiv) in anhydrous DCM (2.7 ml) was added DIPEA (0.15 ml, 0.85 mmol, 2.5 equiv) and stirred for 10 min. To this solution 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.11 ml, 0.51 mmol, 1.5 equiv) was added and mixture was stirred for 12 h. The solvent was evaporated to dryness and the residue was redissolved in ethyl acetate (20 ml). The organic layer was washed with 5% sodium bicarbonate (20 ml) and brine solution (20 ml), dried over sodium sulphate and evaporated to dryness. The crude product was purified by silica gel column chromatography (40% EtOAc in hexane containing 0.5% Et$_3$N) to afford the product **12** (0.27 g, 78%) as white foam. TLC (40% EtOAc in hexane containing few drops of Et$_3$N) R_f = 0.64, 0.74 for two diastereomer;^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 8.66 (s, 1H), 7.57–7.55 (m, 2H), 7.43–7.41 (m, 4H), 7.37 (br, 1H), 7.22–7.18 (m, 2H), 7.11–7.08 (m, 1H), 7.05 (dd, J_1 = 8.6 Hz, J_2 = 2.6 Hz, 1H), 6.73–6.69 (m, 4H), 6.44 (td, J_1 = 9.2 Hz, J_2 = 2.8 Hz, 1H), 6.14 (d, J = 6.4 Hz, 1H), 5.36 (dd, J_1 = 8.8 Hz, J_2 = 4.0 Hz, 1H), 4.58–4.56 (m, 1H), 4.36 (br, 1H), 4.17–4.13 (m, 1H), 3.80 (dd, J_1 = 11.2, J_2 = 2 Hz, 1H), 3.64 (s, 3H), 3.63 (s, 3H), 3.57–3.47 (m, 4H), 3.13 (dd, J_1 = 11.2 Hz, J_2 = 2.4 Hz, 1H), 2.32–2.19 (m, 1H), 1.16–1.11 (m, 12H), 0.87 (s, 9H), 0.07 (s, 3H), 0.07 (s, 3H); ^{13}C NMR (100 MHz, CDCl$_3$):δ (ppm) 160.3, 159.8, 158.7, 158.0, 149.7, 149.3, 149.2, 144.7, 136.0, 135.6, 135.5, 130.4, 130.3, 129.7, 129.6, 128.5, 128.0 127.1, 117.4, 113.3, 113.2, 111.7, 111.5, 111.5, 111.4, 106.9, 106.2, 105.9, 105.9, 87.9, 87.1, 84.3, 75.3, 75.3, 73.0, 72.9, 62.7, 57.6, 57.4, 55.3, 43.6, 43.5, 25.8, 24.9, 24.8, 24.7, 20.3, 20.2, 18.1, -4.8, -4.8, -4.7; ^{31}P NMR (162 MHz, CDCl$_3$): δ (ppm) 150.93; ^{19}F NMR (376.6 MHz, CDCl$_3$): δ (ppm) -122.61; HRMS: m/z Calculated. for C$_{53}$H$_{65}$FN$_4$O$_{10}$Si [M+H]$^+$ = 995.4192, found = 995.4218.

4.4.4. Photophysical analysis of 5-fluorobenzofuran-modified uridine analogue 1 in different solvents.

4.4.4.1. UV-absorption and steady-state fluorescence studies

UV-absorption spectra of nucleoside **1** (25 µM) was recorded on a Shimadzu UV-2600 spectrophotometer in quartz cuvette (Hellma, path length 1 cm). Each UV sample contained 2.5% DMSO. Steady-state fluorescence study of nucleoside **1** (5 µM) in different solvent was

performed in micro fluorescence cuvette (Hellma, path length 1 cm) on Fluoromax-4 spectrofluorometer. In this study samples were excited at their lowest energy absorption maximum (Table 1). Each fluorescence sample contained 0.5% DMSO. Anisotropy values (r) of nucleoside **1** in different solvents were determined by analysing the data using software provided with the instrument Fluoromax-4 spectrofluorometer. Each anisotropy value was an average of 10 successive measurements. All UV-absorption, steady-state fluorescence and anisotropy measurements were performed in triplicate.

4.4.4.2 Time-resolved fluorescence study

Time resolved fluorescence study of nucleoside **1** was performed using time correlated single photon counting (TCSPC) setup (Horiba Jobin Yvon, U.S.A.). Nucleoside **1** (5 µM) in water, ethylene glycol and glycerol was excited by using 339 nm LED source (IBH, UK, NanoLED-339L). In dioxane and methanol nucleoside **1** (400 µM) was excited by using 375 nm diode laser source (IBH, UK, NanoLED-375L). Fluorescence signal at respective emission maxima was collected. All studies were done in triplicate and lifetimes were calculated by fitting the decay profile using IBH DAS6 software. The χ^2 value for all the curve fits was found to be nearly one.

4.4.4.3 Quantum yield

Quantum yield of nucleoside **1** in different solvents was determined relative to 2-aminopurine as the standard. Following equation was used to calculate the quantum yield.[26]

$$\Phi_{F(x)} = \left(A_s/A_x\right)\left(F_x/F_s\right)\left(n_x/n_s\right)^2 \Phi_{F(s)}$$

Where s is the standard, x is the modified nucleoside, A is the absorbance at excitation wavelength, F is the area under the emission curve, n is the refractive index of the solvent, and Φ_F is the quantum yield. Quantum yield of 2-aminopurine in water is 0.68.

4.4.5 Transcription reaction with α-^{32}P ATP

A 1:1 solution of DNA template **T1**–**T5** (5 µM) and RNA polymerase consensus 18 mer promoter DNA sequence was annealed in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) at 90 °C for 3 min and allowed to attain room temperature. The solution was kept on an ice bath for 30 min and stored at -40 °C. Transcription was performed at 37 °C in 40 mM Tris-HCl buffer (pH 7.8) using 250 nM annealed promoter-template duplexes, 10 mM NaCl, 10 mM MgCl$_2$, 10 mM of dithiothreitol (DTT), 2 mM spermidine, 1 U/µL,

RNase inhibitor (Riboblock), 1 mM GTP, CTP, UTP and or modified UTP **2**, 20 μM ATP, 5 μCi α-^{32}P ATP and 3 U/μL (total 60 units) T7 RNA polymerase in a total 20 μL reaction volume. After 3.5 h, the reaction was quenched using 20 μL of loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8). The samples were heated at 75 °C for 3 min and then cooled on an ice bath. The samples (4 μL) were loaded on a sequencing 18% denaturing polyacrylamide gel and were electrophoresed at a constant power (11 W) for nearly 4 h. The bands corresponding to the radioactive products were imaged using an X-ray film. The relative transcription efficiency was calculated using GeneTools software from Syngene. The % of incorporation of modified UTP **2** into RNA oligonucleotide by T7 RNA polymerase was determined considering the transcription efficiency in presensce of all natural NTPs as 100 %. All reactions were performed in duplicate and the errors in yields were ≤ 4%. See Figure 5.

4.4.6 Large scale transcription

Large scale transcription reaction with a total volume of 250 μL was performed with the template **T1** for isolation and further characterization of transcript **4**. The transctiption reaction contained 2 mM GTP, ATP, CTP and modified UTP **2**, 20 mM MgCl$_2$, 0.4 U/μL of RNase inhibitor (RiboLock), 800 units T7 RNA polymerase and 300 nM promoter template duplex. The transcription reaction was performed at 37 °C for 12 h and the white precipitation of pyrophosphate was observed. The white precipitate of pyrophosphate appeared and the volume was reduced to almost one third of the total volume using speed vac. 40 μL loading buffer (10 mM Tris HCl, 7 M urea, 100 mM EDTA, pH 8.0) was added to the residual mixture and loaded onto to the 20% polyacrylamide gel electrophoresis and run at 25 W constant power for 6 h. Gel was UV shadowed, the appropriate band corresponding to transcript was cut and transferred to the Poly-Prep column (Bio-Rad). The gel pieces were crushed, the transcript was extracted with 0.5 M ammonium acetate for 12 h and desalted using Sep-Pak classic C18 cartridges (Waters). Around 12-14 nmole of transcript 4 (ε_{260} = 91560) was isolated from each reaction. The purity of the transcript was examined by HPLC and the transcript was characterized by MALDI TOF mass analysis.

4.4.7 Enzymatic digestion of transcript 4

Transcript **4** (4 nmole) was incubated with snake venom phosphodiesterase I (0.01 U), calf intestinal alkaline phosphatase (1 U/μL) and RNase A (0.25 μg), 50 mM Tris-HCl buffer (pH

142

8.5, 40 mM MgCl$_2$, 0.1 mM EDTA) in a total volume of 100 µL for 12 h at 37 °C. Afterwards, RNase T1 (0.2 U/µL) was added and the sample was incubated for additional 4 h at 37 °C. The resulting mixture obtained from the digestion was analyzed by RP-HPLC using Phenomenex-Luna C18 column (250 × 4.6 mm, 5 micron) at 260 and 322 nm. Mobile phase A: 50mM TEAA buffer (pH 7.3), mobile phase B: acetonitrile. Flow rate: 1 ml/min. Gradient: 0−10% B in 20 min, 10−100% B in 10 min.

4.4.8 Solid phase RNA synthesis

5-fluorobenzofuran-modified RNA ONs **11, 13** and **14** were synthesized on 1.0 µmol scale (1000 Å CPG solid support) by a standard ON synthesis protocol using phosphoramidite **12**.[27] After trityl deprotection on the synthesizer, the solid support was treated with 1.5 ml 1:1 solution of 10 M methylamine in ethanol and water for 12 h. The resulting mixture was then centrifuged and the supernatant was evaporated to dryness on a Speed Vac. After evaporation, the residue was redissolved in DMSO (100 µL) and TEA.3HF (150 µL). The resulting solution was heated at 65 °C for 2.5 h and slowly cooled to RT. The deprotected ON solution was lyophilized to dryness. The RNA ON residues was purified by 20% denaturing PAGE. The band corresponding to the full-length modified ON product was identified by UV shadowing, cut and transferred to a Poly-Prep column (Bio-Rad). The gel pieces were crushed with a sterile glass rod, and the ON was extracted using ammonium acetate buffer (0.5 M, 3 ml) for 12 h and desalted using Sep-Pak classic C18 cartridges (Waters).

4.4.9 MALDI-TOF mass analysis of ONs

A mixture of 2 µL of the modified ON (~200 µM), 4 µL of a solution of 9:1 3-hydroxypicolinic acid and ammonium citrate buffer (100 mM, pH 9) and 2 µL of an internal DNA standard (100 µM) was desalted using an ion-exchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form). The sample was then spotted on a MALDI plate, air dried and subjected to mass analysis.

4.4.10 Thermal melting analysis of RNA-DNA duplexes

The 1:1 mixture of RNA ON (**3**, **4**, **9**, **10** or **11**) (5 µM) and its respective complementary DNA sequence (5 µM) was heated at 90 ºC for 3 min in sodium phosphate buffer (10 mM, pH 7.1) containing 500 mM NaCl. Samples were allowed to come to room temperature over a period of 2 h and kept at 4 ºC for 1 h. Thermal melting analysis of the duplexes were

performed using Cary 300Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 260 nm. The thermal melting analysis was performed by Dr. Vyankat Sontakke.

4.4.11 Steady-state fluorescence studies of modified model RNA ONs and their duplexes

RNA ONs (**4**, **9**, **10** and **11**) (0.5 µM) and the 1:1 mixture of RNA ONs (0.5 µM) and their respective complementary DNA sequences were heated at 90 ºC for 3 min in sodium phosphate buffer (10 mM, pH 7.1) containing 500 mM NaCl. Samples were allowed to come to room temperature over a period of 2 h and kept at 4 ºC for 1 h. Samples were excited at 330 nm with excitation and emission slit width 5 nm and 5 nm, respectively on a Fluoromax-4 spectrofluorometer. The fluorescence studies were performed by Dr. Vyankt Sontakke.

4.4.12 Fluorescence Binding assay

Modified IRES subdomain IIa constructs (0.5 µM) were assembled by heating the 1:1 mixture of either **13** and **13c** or **14** and **14c** in sodium cacodylate buffer (10 mM, pH 6.5) at 75 ºC for 4 min and allowing the mixtures to come to room temperature over a period of 1 h. 200 µL of IRES subdomain IIa construct (**13•13c** or **14•14c**) (0.5 µM) was taken in a micro fluorescence cuvette (Hellma, path length 1.0 cm) and titrated by addition of 1 µL of metal ion stock solutions of increasing concentration. After each addition the samples were incubated for 2 min and were excited at 330 nm with excitation and emission slit widths of 3 nm and 4 nm, respectively. The changes in fluorescence intensity at respective emission maximum were calculated. The final concentration range of metal ions for these titrations was 0.6 µM to 91 mM. The total volume change upon addition of metal ions during titration was ≤ 7%. The spectrum corresponding to a blank without any ON but containing the particular metal ion concentration was subtracted from each spectrum. A control titration where instead of metal ion, water was added to IRES constructs, did not show any detectable change in fluorescence intensity. From the dose-dependent quenching curves, the apparent dissociation constants (K_d) for the binding of metal ion (Mg^{II} or Mn^{II}) to respective IRES subdomain IIa were determined by fitting normalized fluorescence intensity (F_N) *versus* ligand concentration plot to equation 1, where one-site binding model was used (Origin 8.5).[21b]

$F_N = (F_i - F_0)/(F_s - F_0)$

F_N = normalized florescence intensity, F_i = florescence intensity at each titration point, F_0 = fluorescence intensity in absence of metal ion, F_s = fluorescence intensity at saturation

Equation 1

$F_N = F_s * ([M]/(K_d + [M]))$

[M] = metal ion concentration

4.4.13 ^{19}F NMR studies of modified model short RNA ONs and IRES subdomain IIa

Model RNA ONs (**4**, **9**, **10** and **11**) (50 µM) and their respective duplexes (**4•4c**, **9•9c**, **10•10c** and **11•11c**) were annealed in sodium phosphate buffer (10 mM, pH 7.1) containing 500 mM NaCl similarly like steady state fluorescence samples and taken for ^{19}F NMR analysis. ^{19}F NMR study of modified IRES subdomain IIa construct **14•14c** (50 µM), pre-annealed in 10 mM sodium cacodylate buffer (pH 6.5) was performed either in absence or presence of 5 mM Mg^{II}. ^{19}F NMR spectra of ONs were recorded at a frequency of 564.9 MHz on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe at 298 K. The following spectral parameters were used[8b]: [19]F excitation pulse: 11 µs; spectral width: 21.28 ppm; transmitter frequency offset: -121.14 ppm; acquisition time: 80 ms; relaxation delay: 1.5 s and number of scans 28000. Each spectrum was processed with an exponential window function using lb = 10 Hz.

4.5.References

1. (a) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1013–1023. (b) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234. (c) Serganov, A.; Patel, D. J. *Curr Opin Struct Biol.* **2012**, *22*, 279–286. (d) Doudna, J. A.; Charpentier, E. *Science* **2014**, *346*, 1258096. (e) Dethoff, E. A.; Chugh, J.; Mustoe, A. M.; Al-Hashimi, H. M. *Nature* **2012**, *482*, 322–330.
2. (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579–2619. (b) Okamoto, A.; Saito, Y.; Saito, I. *J. Photochem. Photobiol C* **2005**, *6*, 108–122. (c) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struct. Biol.* **2004**, *14*, 350. (d) Al-Hashimi, H. M.; Walter, N. G. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321–329. (d) Bardaro, M. F.; Varani, Jr, G. *WIREs RNA* **2012**, *3*, 122–132. (e) Mooers, B. H. *Methods* **2009**, *47*, 168. (f) Ennifar, E.; Carpentier, P.; Ferrer, J. L.; Walter, P.; Dumas, P. *Acta Crystallogr., Sect. D* **2002**, *58*, 1262–1268.

3. (a) Wilhelmsson, L. M. Q. *Rev. Biophys.* **2010**, *43*, 159–183. (b) Srivatsan, S. G.; Sawant, A. A. *Pure Appl. Chem.* **2011**, *83*, 213–232.
4. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.
5. Hammann, C.; Norman, D. G.; Lilley, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5503–5508.
6. Olejniczak, M.; Gdaniec, Z.; Fischer, A.; Grabarkiewicz, T.; Bielecki, L.; Adamiak, R. W. *Nucleic Acids Res.* **2002**, *30*, 4241–4249.
7. Olsen, G. L.; Edwards, T. E.; Deka, P.; Varani, G.; Sigurdsson, S. T.; Drobny, G. P. *Nucleic Acids Res.* **2005**, *33*, 3447–3454.
8. (a) Granqvist, L.; Virta, P. *J. Org. Chem.* **2015**, *80*, 7961–7970. (b) Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. *J. Org. Chem.* **2012**, *77*, 1026–1044.
9. Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. *Angew. Chem. Int. Ed.* **2006**, *45*, 3450–3453.
10. Lombés, T.; Moumné, R.; Larue, V.; Prost, E.; Catala, M.; Lecourt, T.; Dardel, F.; Micouin, L.; Tisné, C. *Angew. Chem. Int. Ed.* **2012**, *51*, 9530–9534.
11. Kiviniemi, A.; Virta, P. *J. Am. Chem. Soc.* **2010**, *132*, 8560–8562.
12. Manna, S.; Sarkar, D.; Srivatsan, S. G. *J. Am. Chem. Soc.* **2018**, *140*, 12622–12633.
13. Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J.* **2011**, *17*, 12820–12827.
14. (a) Jang, J. Y; Chung, R. T. *Gut Liver* **2011**, *5*, 117–132. (b) Alter, H. J.; Purcell, R. H.; Shih, J. W.; Melpolder, J. C.; Houghton, M.; Choo, Q.-L.; Kuo, G. *N Engl J Med.* **1989**, *321*, 1494–1500.
15. (a) Tsukiyama-Kohara, K.; Iizuka, N.; Kohara, M.; Nomoto, A. *J. Virol.* **1992**, *66*, 1476–1483. (b) Kieft, J. S.; Grech, A.; Adams, P.; Doudna, J. A. *Symp. Quant. Biol.* **2001**, *66*, 277–283. (c) Otto, G. A.; Puglisi, J. D. *Cell* **2004**, *119*, 369–380.
16. Kieft, J. S.; Zhou, K.; Jubin, R.; Murray, M. G.; Lau, J. Y.; Doudna, J. A. *J. Mol. Biol.* **1999**, *292*, 513–529.
17. Dibrov, S. M.; Parsons, J.; Carnevali, M.; Zhou, S.; Rynearson, K. D.; Ding, K.; Sega, E. G.; Brunn, N. D.; Boerneke, M. A.; Castaldi, M. P.; Hermann, T. *J. Med. Chem.* **2014**, *57*, 1694–1707.
18. Filbin, M. E.; Kieft, J. S. *RNA* **2011**, *17*, 1258–1273.
19. Locker, N.; Easton, L. E.; Lukavsky, P. J. *EMBO J.* **2007**, *26*, 795–805.
20. Pestova, T. V.; de Breyne, S.; Pisarev, A. V.; Abaeva, I. S.; Hellen, C. U. *EMBO J.* **2008**, *27*, 1060–1072.
21. (a) Spahn, C. M.; Kieft, J. S.; Grassucci, R. A.; Penczek, P. A.; Zhou, K.; Doudna, J. A.; Frank, J. *Science* **2001**, *291*,1959. (b) Dibrov, S. M.; Johnston-Cox, H.; Weng, Y. H.; Hermann, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 226–229.
22. (a) Parsons, J.; Castaldi, M. P.; Dutta, S.; Dibrov, S. M.; Wyles, D. L.; Hermann, T. *Nat. Chem. Biol.* **2009**, *5*, 823–825. (b) Dibrov, S. M.; Ding, K.; Brunn, N. D.; Parker, M. A.; Bergdahl, B. M.; Wyles, D. L.; Hermann, T. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5223–5228.
23. (a) Zhao, Q.; Han, Q.; Kissinger, C. R.; Hermann, T.; Thompson, P. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 436–443. (b) Lukavsky, P. J; Kim, I; Otto, G. A.; Puglisi, J. D. *Nat Struct Biol* **2003**, *10*, 1033–1038. (c) Paulsen, R. B.; Seth, P. P.; Swayze, E. E.; Griffey, R. H.; Skalicky, J. J.; Cheatham, T. E.; Davis, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7263–7268.
24. Boerneke, M. A.; Dibrov, S. M.; Gu, J.; Wyles, D. L.; Hermann, T. *Proc. Natl. Acad. Sci.U. S. A.* **2014**, *111*, 15952–15957.
25. Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem.* **1994**, *5*, 508–512.
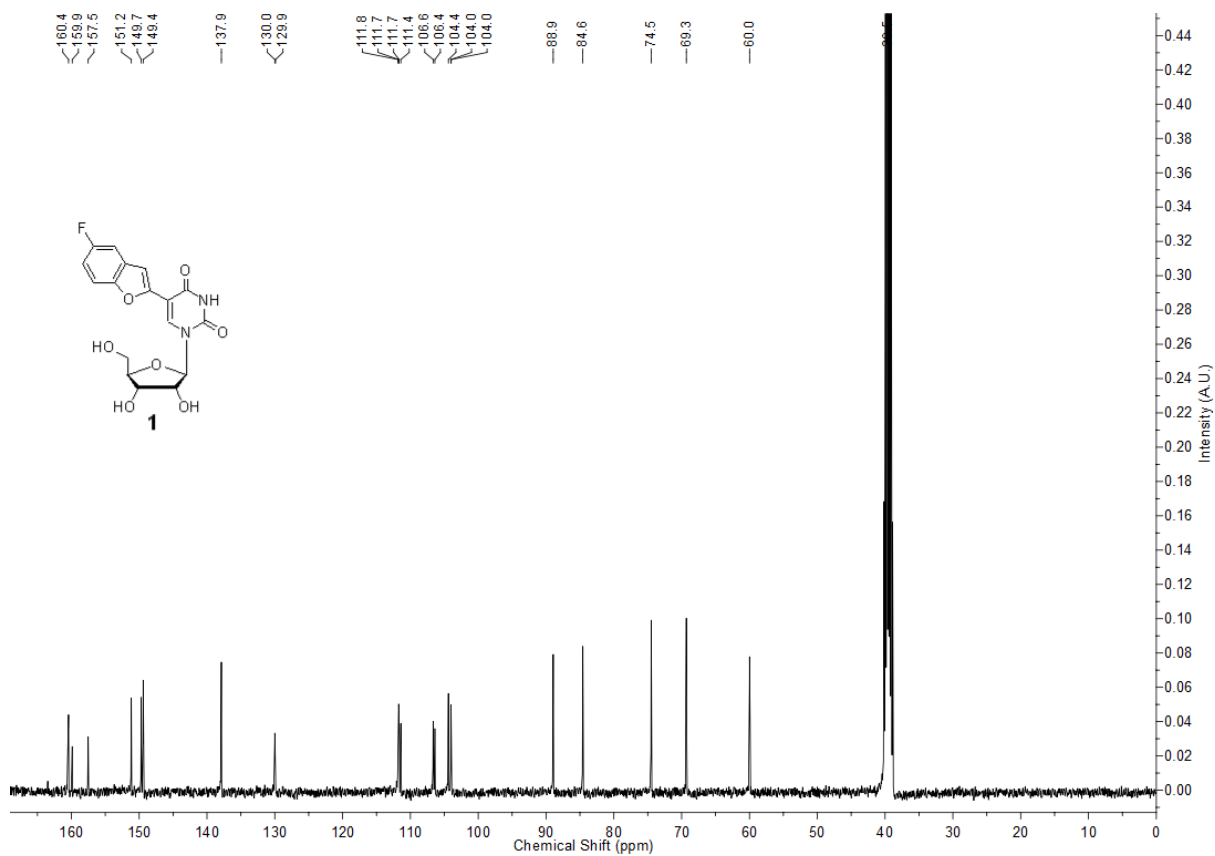26. Lavabre, D.; Fery-Forgues, S. *J. Chem. Educ.* **1999**, *76*, 1260–1264.

27. Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149.

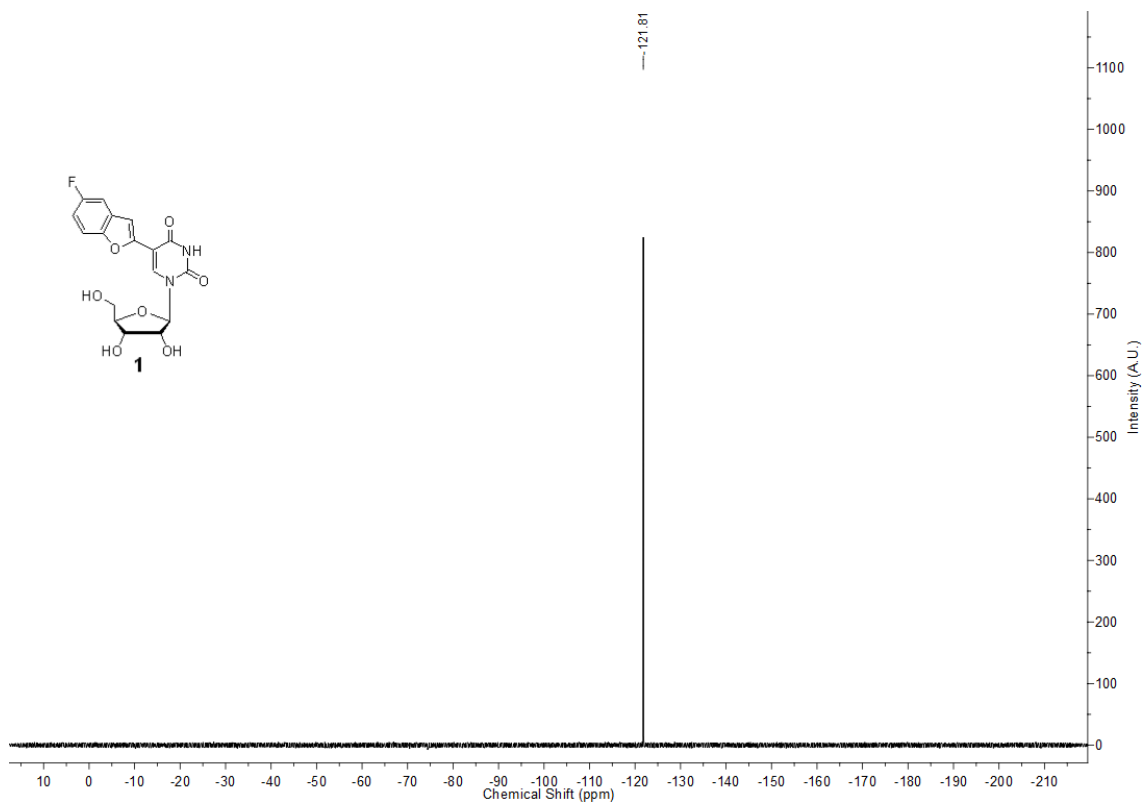## 4.6 Appendix-II: Characterization data of synthesized compounds

^{1}H NMR of **1** (400 MHz, d_6-DMSO)

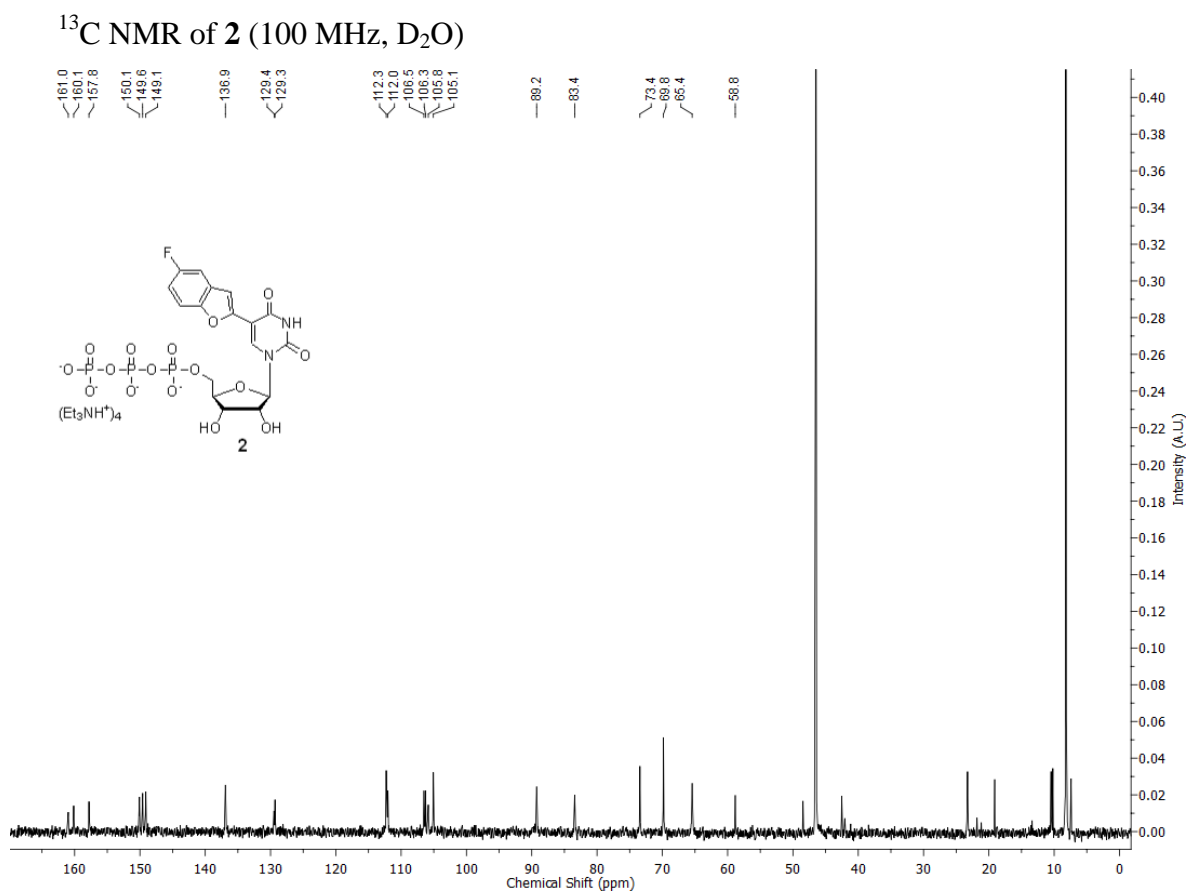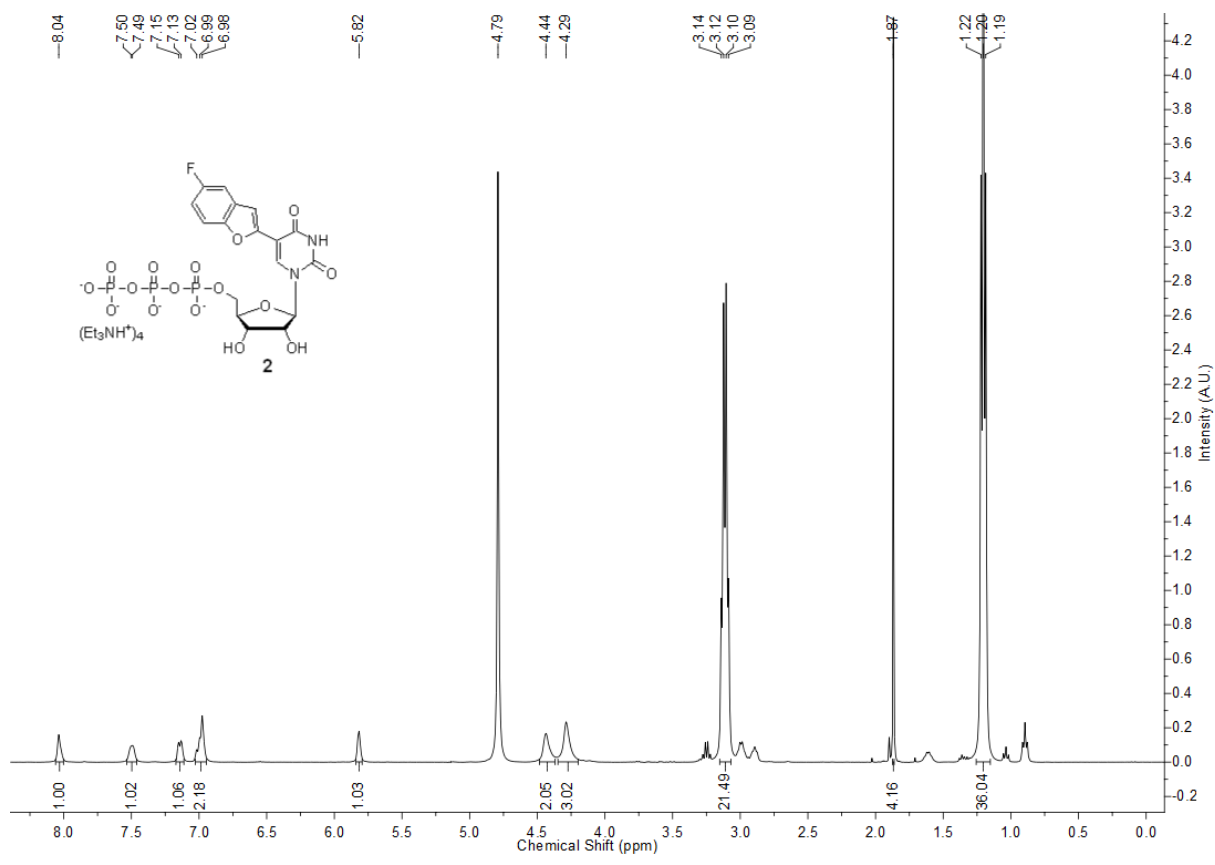^{13}C NMR of **1** (100 MHz, d_6-DMSO)
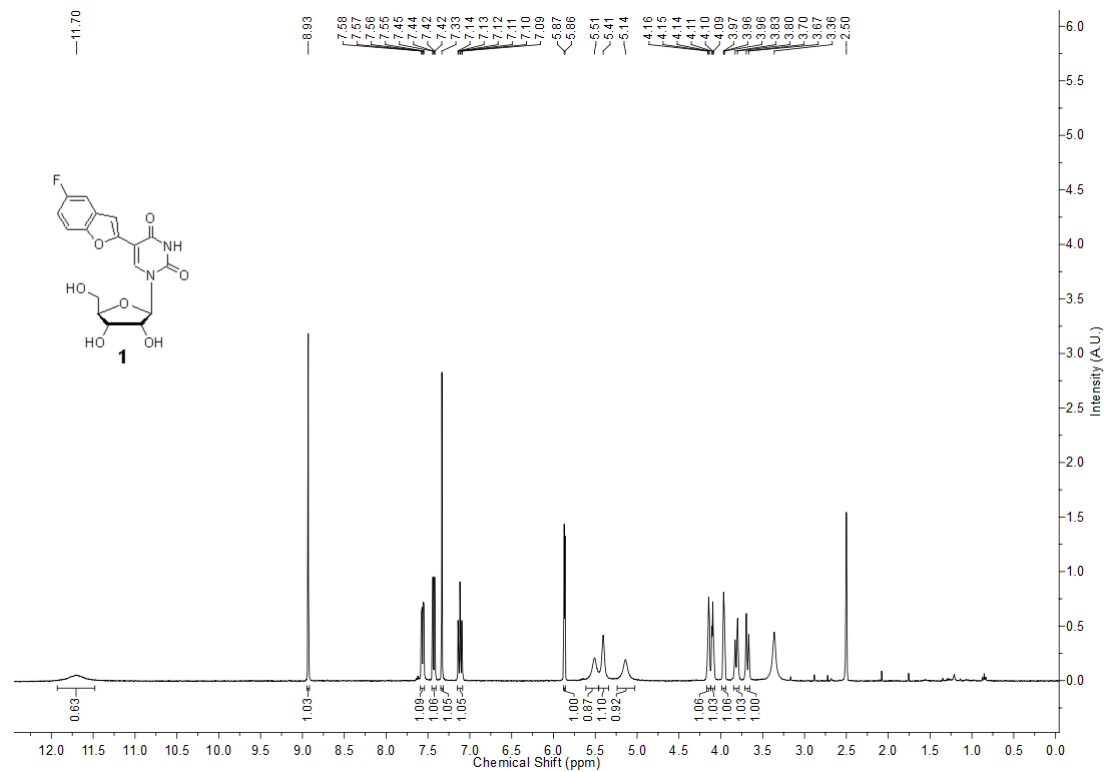
^{19}F NMR of **1** (376.6 MHz, d_6-DMSO)

^{1}H NMR of **2** (400 MHz, D$_2$O)

^{13}C NMR of **2** (100 MHz, D$_2$O)

^{31}P NMR of **2** (162 MHz, D$_2$O)
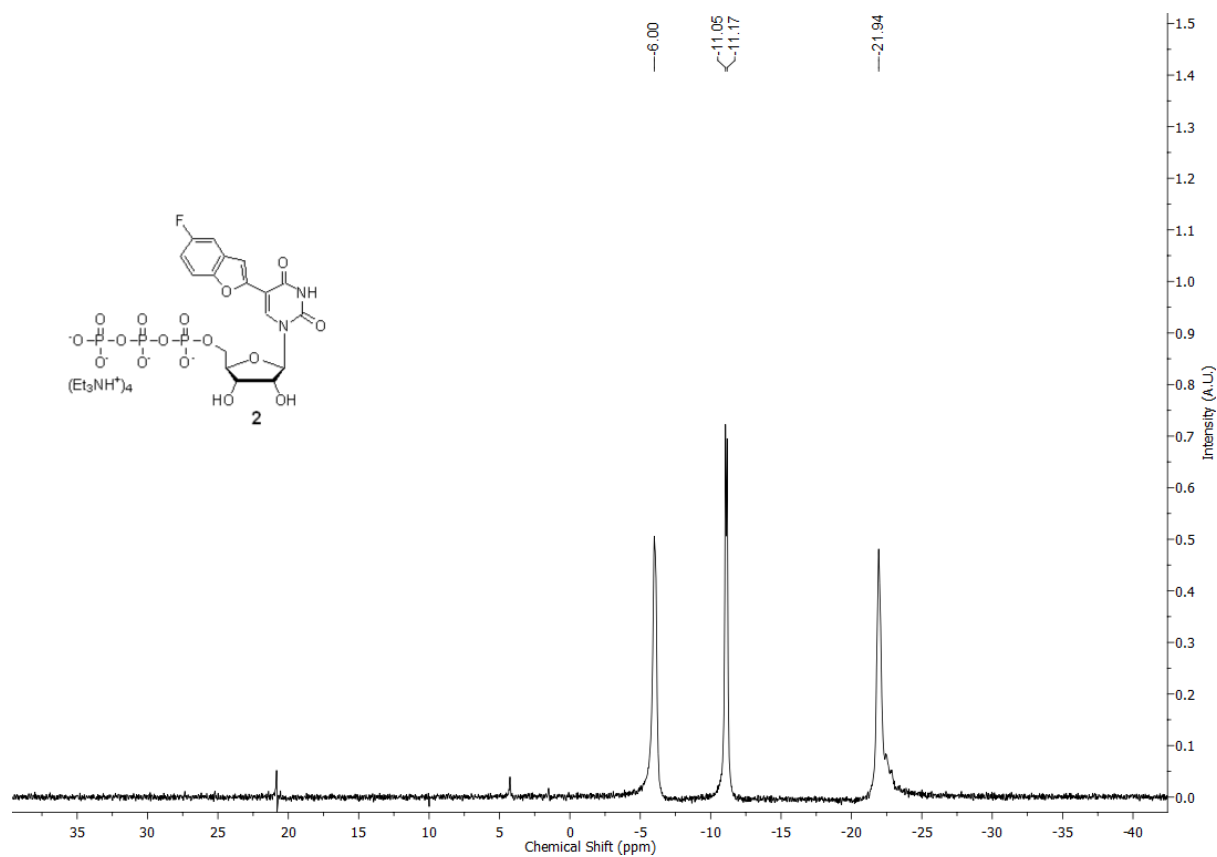
^{19}F NMR of **2** (376.6 MHz, D$_2$O)
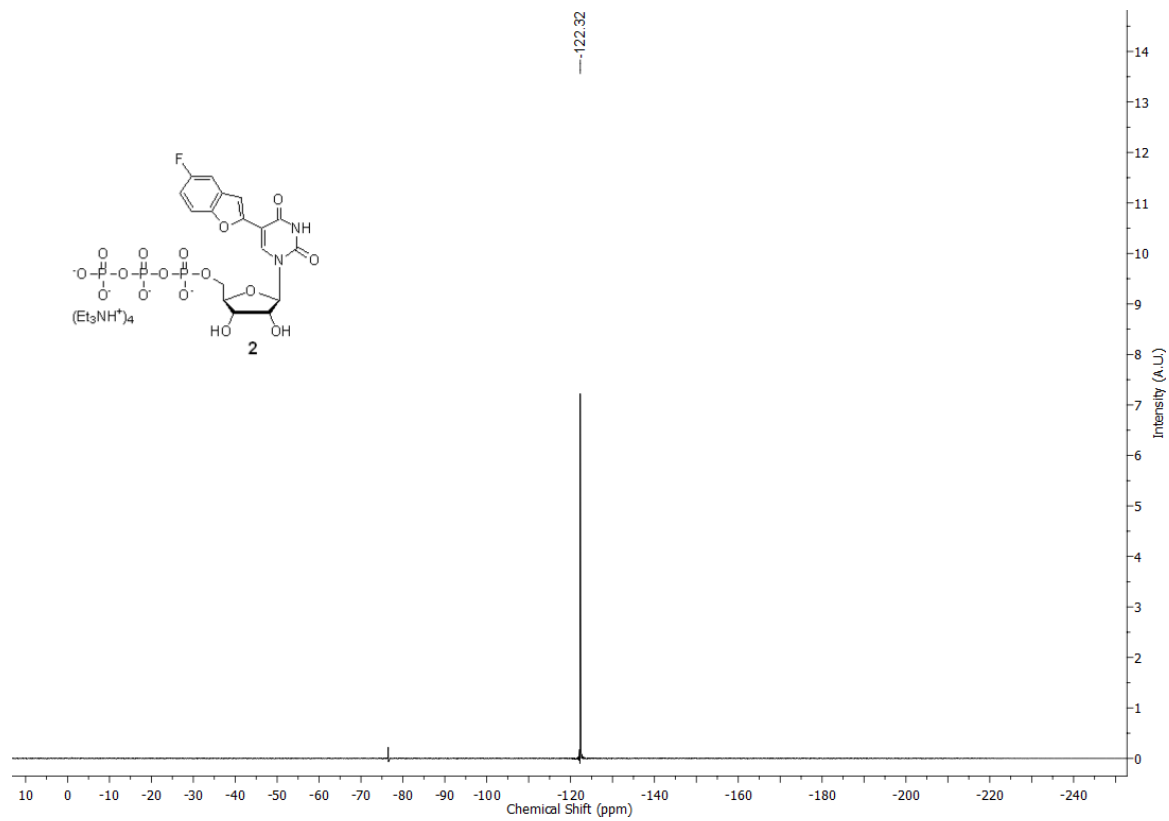
^{1}H NMR of **12** (400 MHz, CDCl$_3$)
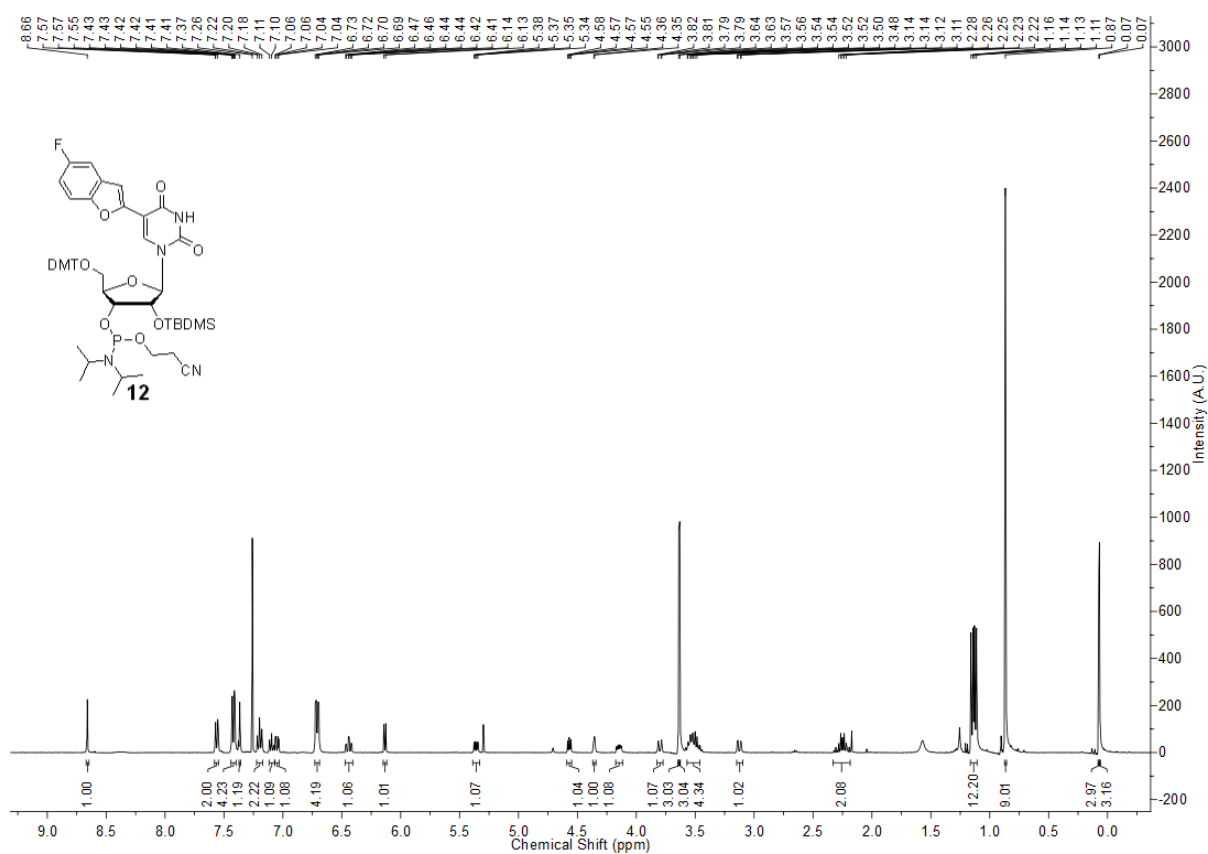
^{13}C NMR of **12** (100 MHz, CDCl$_3$)

³¹P NMR of **12** (162 MHz, CDCl₃)
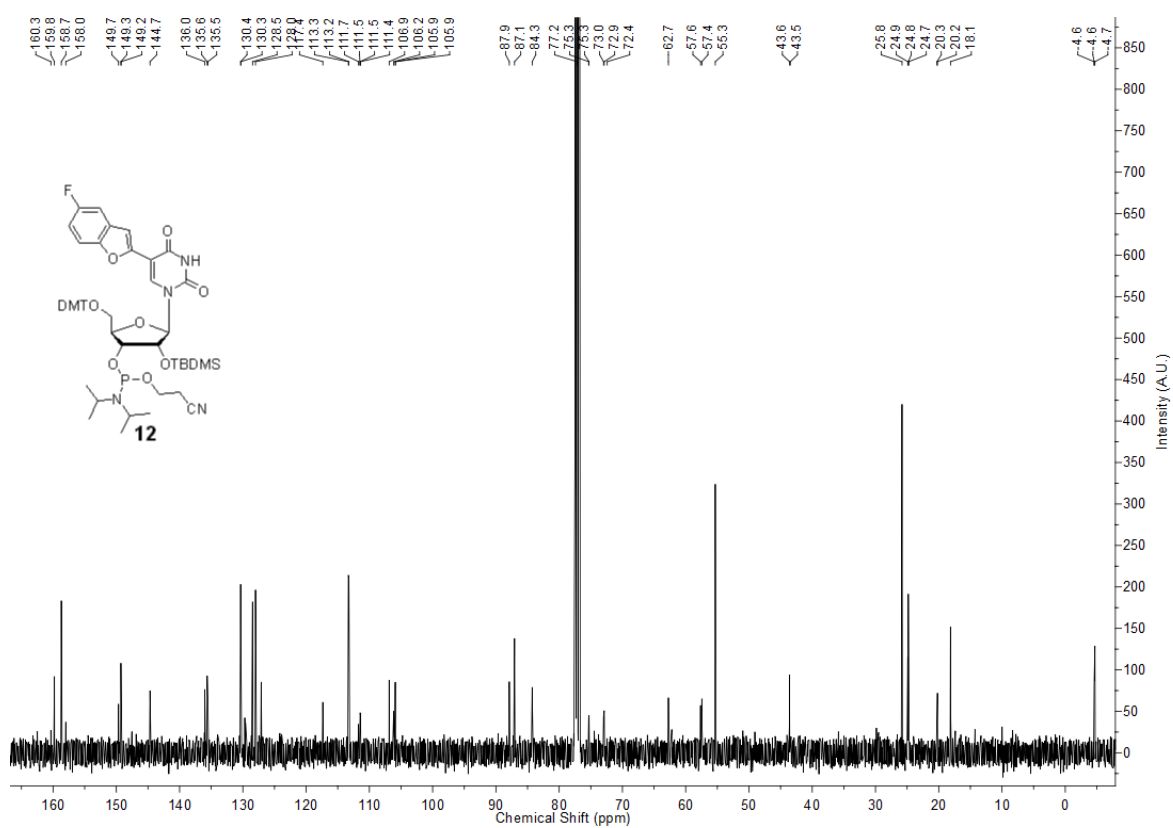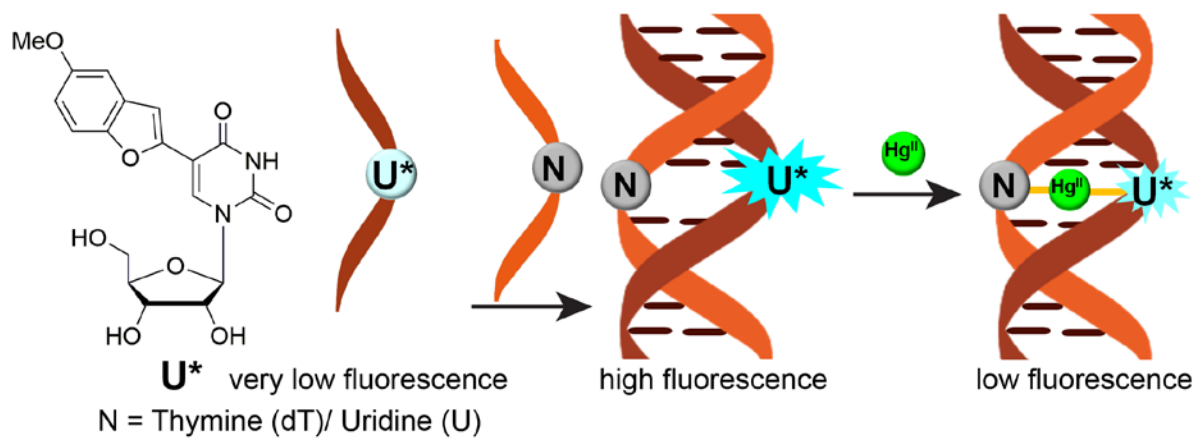
^{31}P NMR of **12** (162 MHz, CDCl$_3$)
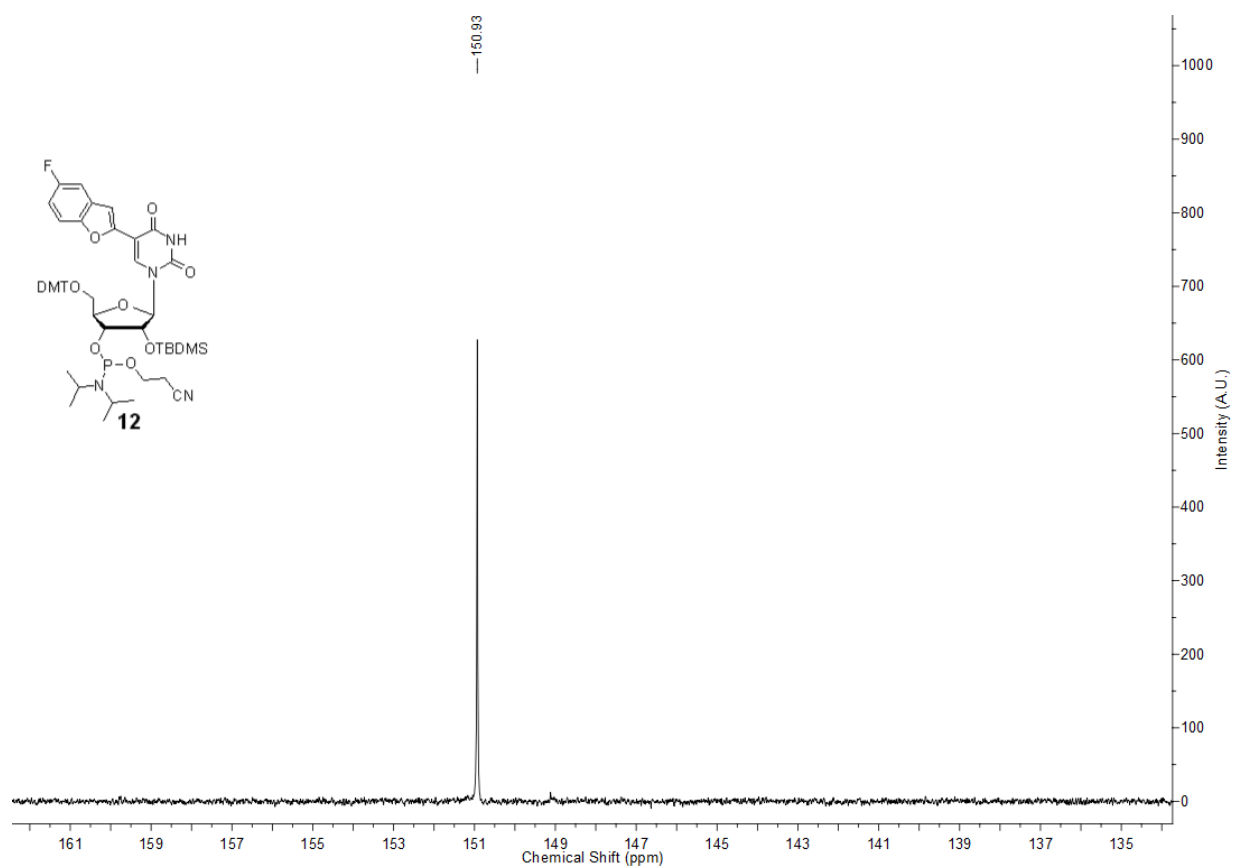
^{19}F NMR of **12** (376.6 MHz, CDCl$_3$)
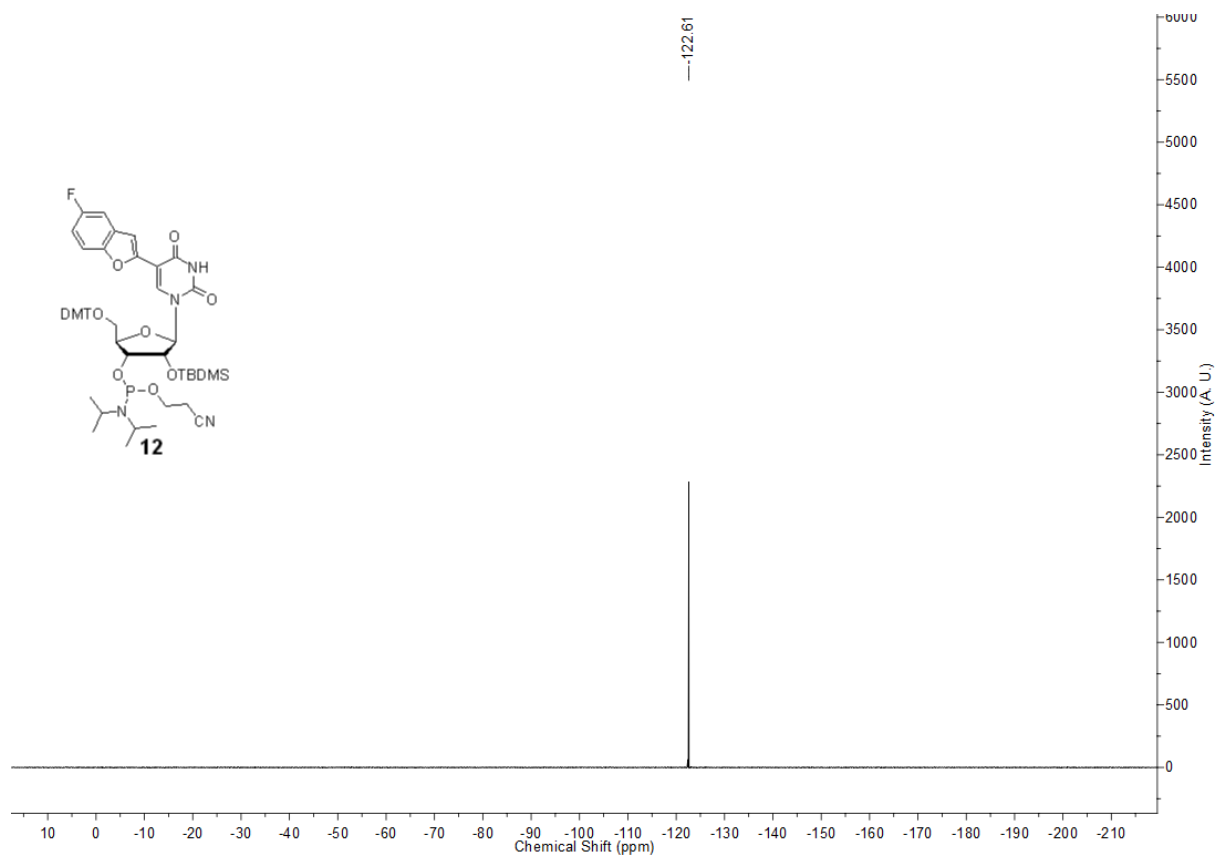
Chapter 5

An Environment-Sensitive Fluorescent Nucleoside Probe Detects Metal-Mediated Base Pairing in Duplexes

U* very low fluorescence
N = Thymine (dT)/ Uridine (U)

high fluorescence

low fluorescence

5.1 Introduction

Microenvironment-sensitive fluorescent nucleoside probes, which report structural or conformational changes at a particular site of the nucleic acid, have served as useful tools in understanding nucleic acid structure, function and recognition.[1] For example, fluorescent nucleoside probes have been used to detect base-pair mismatches,[2] abasic sites,[3] mutations,[4] and nucleic acid-protein[5] and nucleic acid-small molecule interactions.[6] However, some of these probes have limitations like very low quantum yield after incorporation into oligonucleotides, absorption and or emission maxima in UV region, and low sensitivity to changes in their surrounding environment, which have restricted their broad applications.[1] In this context, there is a significant demand for the development of nucleoside probes, which show high quantum yields when incorporated into oligonucleotides and high sensitivity to subtle changes in surrounding environment.

As a part of the ongoing research endeavours in our laboratory, we have developed a series of microenvironment-sensitive fluorescent nucleoside probes by attaching heterocyclic units on to purine and pyrimidine nucleosides.[7] Many of these analogs are structurally non-perturbing when incorporated into the target ONs. The responsiveness of the probes has been empirically used in setting up fluorescence assays to detect DNA and RNA abasic sites,[7b,7c] nucleic acid-ligand interactions[7d,7f,7h] and i-motifs/G-quadruplex topologies[7f-7h] (Figure 1). In a continued effort to expand the repertoire of responsive nucleoside probes, here we describe the development of 5-methoxybenzofuran-modified uridine nucleoside, which is highly sensitive to changes in its microenvironment as compared to many other nucleoside probes. Notably, its fluorescence properties are highly sensitive to slight changes in micro-polarity. This probe was successfully incorporated into RNA ONs by transcription reaction with good efficiency and it showed higher fluorescence intensity in RNA ON as compared to free nucleoside. Importantly, it selectively reported thymine-uracil (dT-U) and uracil-uracil (U-U) mismatches in model DNA-RNA and RNA-RNA duplexes, respectively, with significant enhancement in fluorescence (Figure 2). This property of the nucleoside analog was aptly utilized in establishing a diagnostic method to detect single mercury (Hg) mediate base paring in DNA-RNA and RNA-RNA duplexes. Further, by monitoring the fluorescence profile of 5-methoxybenzofuran-modified duplexes, containing a particular mismatch, as a function of Hg^{II} concentration we were able to compare the binding affinity of Hg^{II} to dT-U and U-U mismatches.
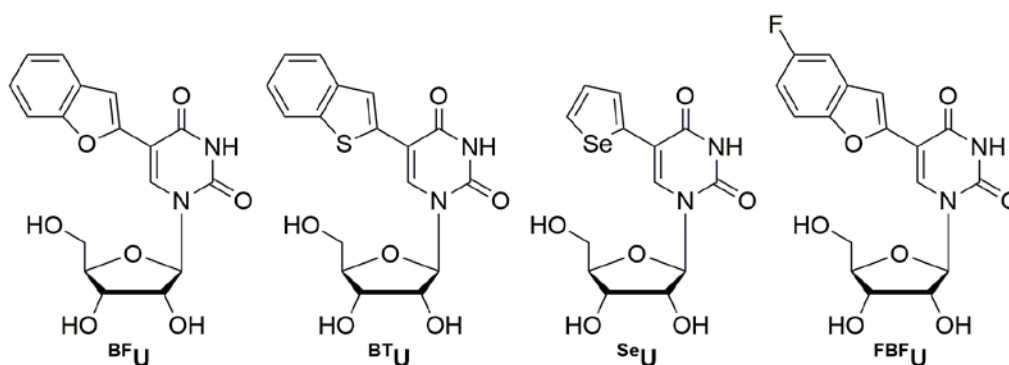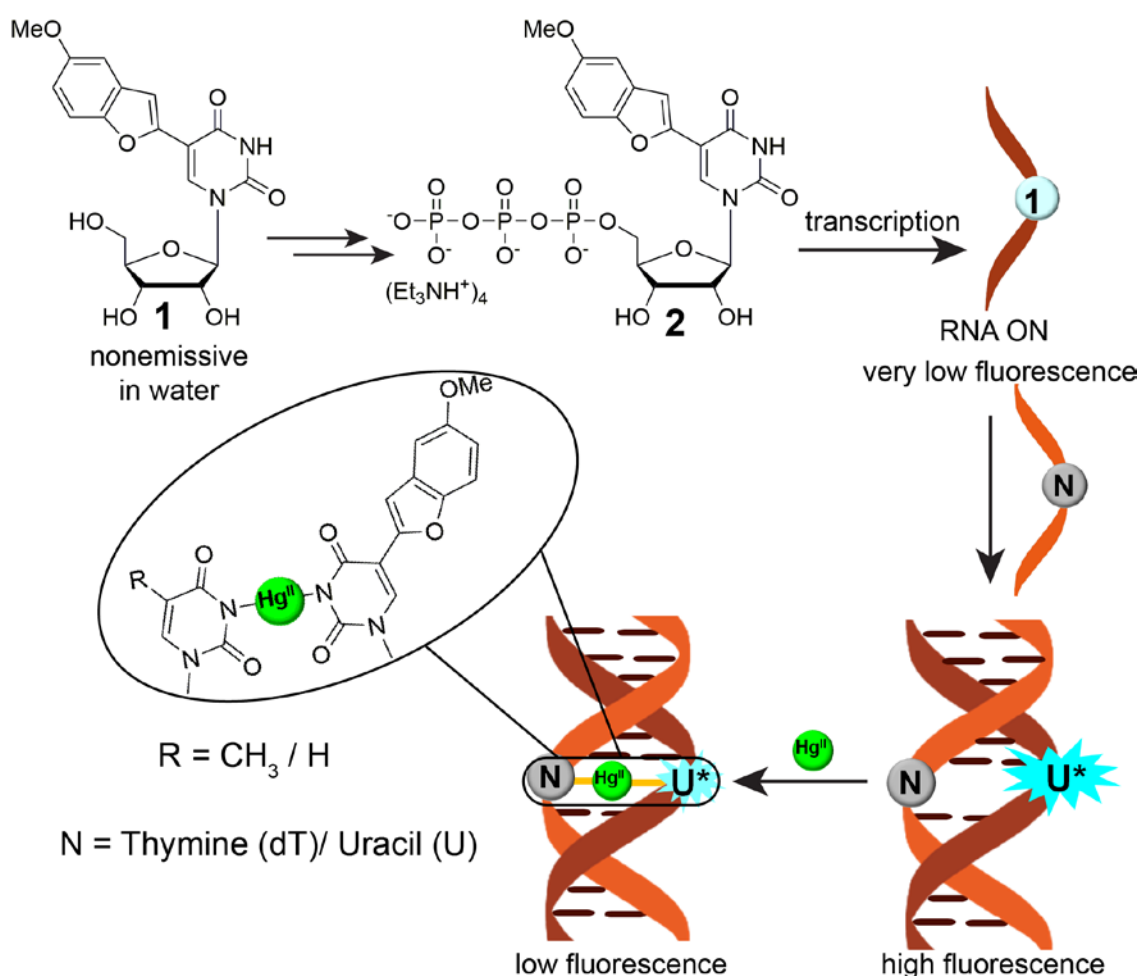
Figure 1. Chemical structure of fluorescent nucleoside probes.

Significance of mercury medicated base pairing

Hg[II] binds to thymine-thymine (dT-dT) and uracil-uracil (U-U) mispairs via deprotonation of imino protons.[8] As Hg[II] is a highly toxic metal, its interaction with nucleic acid has received more attention and has been studied substantially.[9] Moreover, Hg[II]-mediated metallo base-pairs have been observed to be utilized by enzymes like Klenow fragment (KF) and Taq DNA polymerases for primer extension and PCR amplification, respectively.[10] This aspect has also been employed to develop Hg[II] ion sensor,[11] single nucleotide polymorphism (SNP) detector,[12] Hg[II] ion trapper,[13] conducting material[14] and DNA based logic gates.[10c] The stability, structure and thermodynamic properties of these metal-mediated base pairs have been studied by thermal melting,[15] NMR[16] and isothermal calorimetry[17] techniques, which require higher concentration of the ONs.

In contrast, fluorescence-based methods are highly beneficial as they can be used to detect metallo base-pairs at very low concentrations (~nM range). T-rich ON sequences,[11a,18a] hairpin-forming motifs[18b] and thrombin binding aptamer[18c] containing FRET pairs have been used in detecting T-Hg-T base-pair mediated structural changes, thereby serving as good Hg[II] sensors. Fluorescently labelled T-rich DNA sequences in combination with carbon nanotubes or gold nanoparticles or graphene oxide have enabled the fluorescence detection of Hg[II] with high sensitivity.[19] Alternatively, label-free detection of Hg[II] ions have been established by using intercalating organic dyes and metal complexes, which show turn-on fluorescence upon dT-Hg-dT base pair-induced duplex or hairpin structure formation.[20] However, these methods rely on the structural switch in ON sequences containing multiple dT residues or poly dT sequences. Development of fluorescence-based methods to detect the formation of single Hg[II] mediated base-pair at a specific site of a particular nucleic acid structure has not been well explored. Recently, Luedtke and co-worker used fluorescent nucleoside analogues to detect

155

site-specific Hg[II] mediated base-pair formation in DNA duplex and study their kinetic and thermodynamic properties.[21] However, to best of our knowledge, the majority of studies have focused on dT-Hg-dT base pair formation in a DNA-DNA duplex and very few studies have discussed about dT-Hg-U and U-Hg-U pairs in DNA-RNA and RNA-RNA duplexes. Given their ability to interfere with cellular processes and their application in materials, it is important to detect as well as estimate the relative binding affinity of metals to mismatches present in different nucleic acid duplexes.
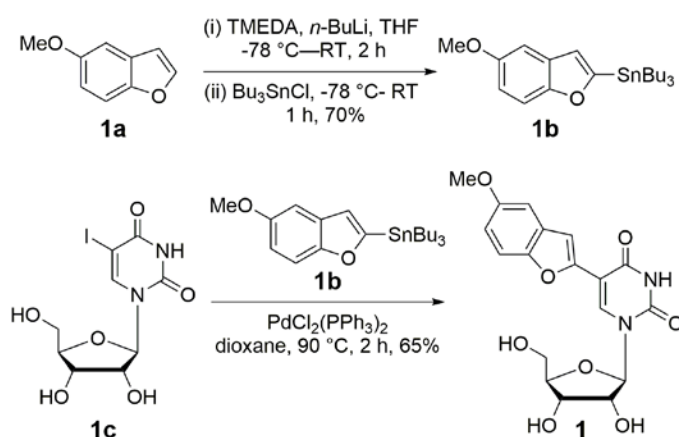

Figure 2. 5-methoxybenzofuran modified probe to detect mismatches and metallo base-pair formation in nucleic acid duplexes. Chemical structure of the 5-methoxybenzofuran modified fluorescent uridine analogue **1** and corresponding triphosphate **2**. Incorporation of nucleoside **1** into RNA ON enables the detection of dT-U and U-U mismatch in a DNA-RNA and RNA-RNA duplex, respectively. The nucleoside **1** also reports the dT-Hg-U and U-Hg-U metallo base-pair formation in respective duplexes.

156

5.2 Results and Discussion

5.2.1 Design and synthesis of 5-methoxybenzofuran modified Nucleoside

In the previous chapters, we have elaborated the use of benzofuran- and 5-fluorobenzofuran-modified nucleoside analogues in studying nucleic acid structure and recognition.[7b,7g] Although these analogs where highly conformation-sensitive, their absorption maximum in UV region (λ_{max} = 323 nm) limits their applications. In this context, we decided to attach an auxochrome (OMe group) to benzofuran so as to push the absorption and emission maxima to a longer wavelength without compromising on the responsiveness of the analogue (Figure 2). Synthesis of nucleoside probe **1** is described in Scheme 1. 5-methoxybenzofuran (**1a**) was synthesized following a previously reported procedure.[22] Stannylation of **1a** followed by palladium-catalyzed cross-coupling with 5-iodouridine afforded the modified nucleoside **1** in good yields (Scheme 1).


Scheme 1. Synthesis of 5-(5-methoxybenzofuran)-uridine. TMEDA: *N,N,N',N'*-tetramethylethylenediamine; THF: tetrahydrofuran.

5.2.2 Methoxybenzofuran-modified uridine is highly sensitive to microenvironment.

Photophysical properties of nucleoside **1** were examined by recording the absorption and steady-state fluorescence spectra of nucleoside **1** in solvents of different polarity and viscosity. It was observed that ground state electronic property like absorption maximum was slightly affected by changes in solvent polarity and viscosity. It was also noticed that depending on solvent polarity, methoxy substitution at 5 position of benzofuran shifted the absorption maximum to higher wavelengths (to 6–11 nm) compared to benzofuran-modified uridine (Table 1 and ref. 7b). Notably, the nucleoside exhibited excellent fluorescence solvatochromism (Table 1). In water, it was practically non-emissive. However, with a slight

change in solvent polarity, for instances from water to methanol, the quantum yield increased remarkably by 137-fold (Φ = 0.0008 in water; Φ = 0.11 in methanol) (Figure 3 and Table 1). The increase in quantum yield was accompanied by a blue-shift in emission maximum. In a more non-polar solvent (dioxane), the nucleoside displayed higher quantum yield (Φ = 0.2) with emission maximum further blue-shifted. Relative quantum yield (Φ_x/Φ_{water}) of **1** and other fluorescent nucleoside probes (SeU[7d], BTU[7a] and BFU[7b]) in different solvents was plotted against $E_T(30)$ (Reichardt's microscopic solvent polarity parameter)[23] to determine the relative sensitivity of probe **1** (Figure 4). Nucleoside **1** was profoundly sensitive to subtle changes in micropolarity as compared to other tested nucleoside probes, which showed only small differences in quantum yield as compared to huge change in quantum yield exhibited by nucleoside **1**. Furthermore, a rotatable aryl-aryl bond between uridine ring and 5-methoxybenzofuran made the emissive nucleoside a molecular rotor.[24] As a result, when the fluorescence studies were performed in solvents with similar polarity but different viscosity it showed higher quantum yield in more viscous medium like glycerol. This enhancement in quantum yield in high viscous solvents is due to rigidification of the fluorophore, which reduces the non-radiative decay processes.[25] The excited state lifetime of the nucleoside probe was also significantly affected upon varying the solvent polarity and viscosity. In methanol, the excited-state lifetime was shorter compared to in dioxane (Figure 5 and Table 1). In high viscous solvent like glycerol the lifetime was discernibly higher than in methanol. Moreover, higher anisotropy in viscous medium confirmed the rigidification of the probe. Taken together, these results indicate that the emission properties of the probe are highly sensitive to subtle changes in its microenvironment.
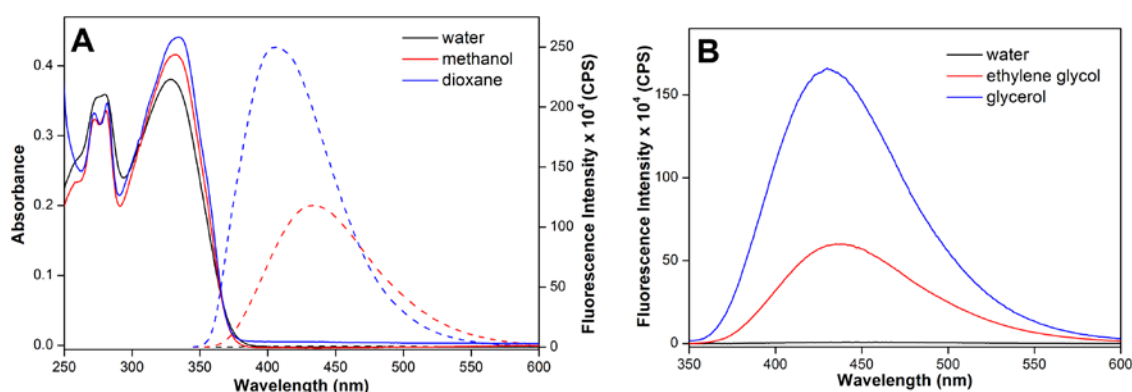
Figure 3. (**A**) Absorption (25 µM, solid lines) and emission (5.0 µM, dashed lines) spectra of nucleoside **1** in various solvents with different polarity. (**B**) Emission spectra of nucleoside **1** (5 µm) in various solvents with different viscosity. Samples were excited at respective lowest energy absorption maximum with excitation and emission slit width of **2** and **2** nm, respectively. All solutions, used for absorption and emission studies contained 2.5% and 0.5% DMSO respectively.
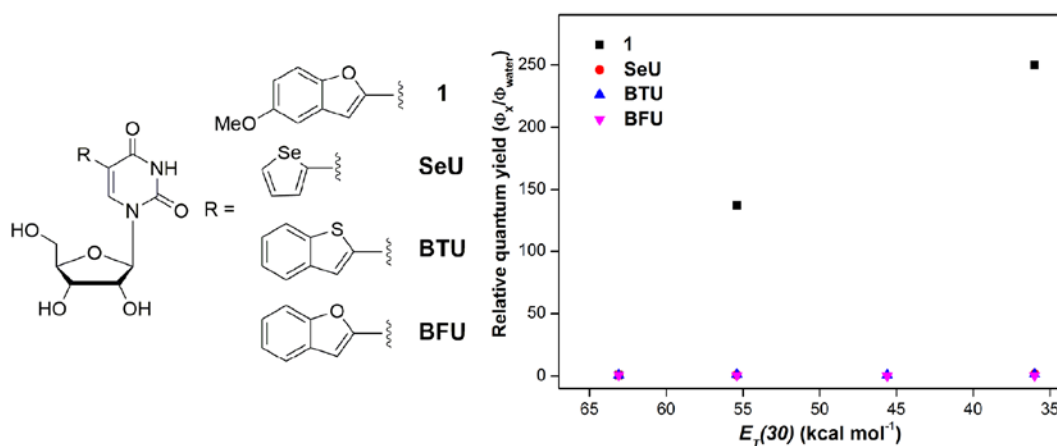
Figure 4. Relative quantum yield ((Φ_x/Φ_{water}), X = particular solvent) *vs* $E_T(30)$ plot for nucleoside **1** and other fluorescent nucleoside probes (SeU[7d], BTU[7a] and BFU[7b]) in various solvents.
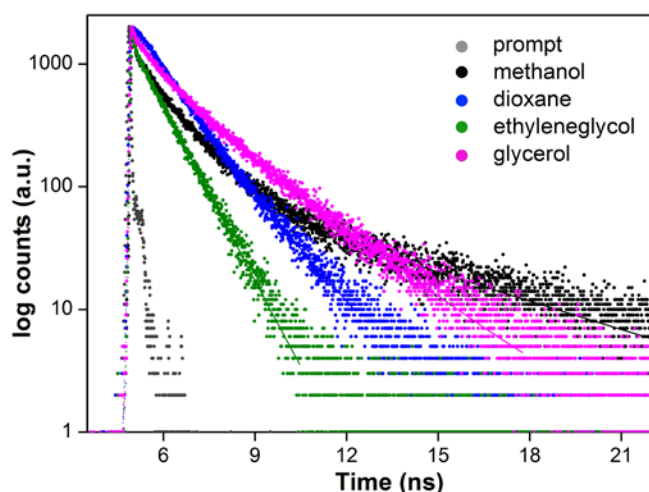
Figure 5. Excited state decay profile of nucleoside **1** in various solvents. Instrument response (prompt) is shown in gray and curve fits are shown in solid lines.

Table 1. Fluorescence properties of nucleoside analogues **1** in solvents of different polarity and viscosity.
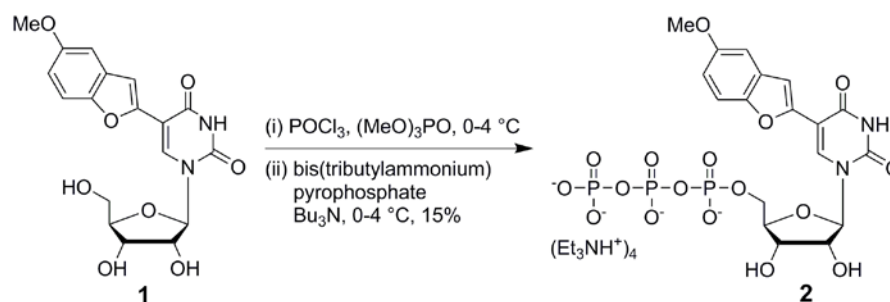
Nucleoside	Solvent	λ_{max}[a]	λ_{em}	$E_T(30)$	I_{rel}[b]	Φ	Φ_x/Φ_{water}[c]	τ_{av}[d]	r[d]
		(nm)	(nm)					(ns)	
	water	329	-	63.1	1	0.0008	1	nd	nd
	methanol	331	433	55.5	165	0.11	137	0.57	nd
1	dioxane	335	406	36.0	349	0.20	250	1.14	0.03
	ethylene glycol	332	435	56.3	84	0.07	88	0.31	0.26
	glycerol	335	430	57.2	232	0.21	263	0.80	0.33

[a]lowest energy absorption maximum is given. [b]Relative emission intensity is given with respect to intensity in water. [c]Relative quantum yield ((Φ_x/Φ_{water}), X= particular solvent) with respect to quantum yield in water (Φ_{water}). [d]Standard deviations for Φ (quantum yield), τ_{av} (average lifetime) and r (anisotropy) in different solvents are ≤ 0.004, ≤ 0.06 ns and ≤ 0.004 respectively. nd = not determined.
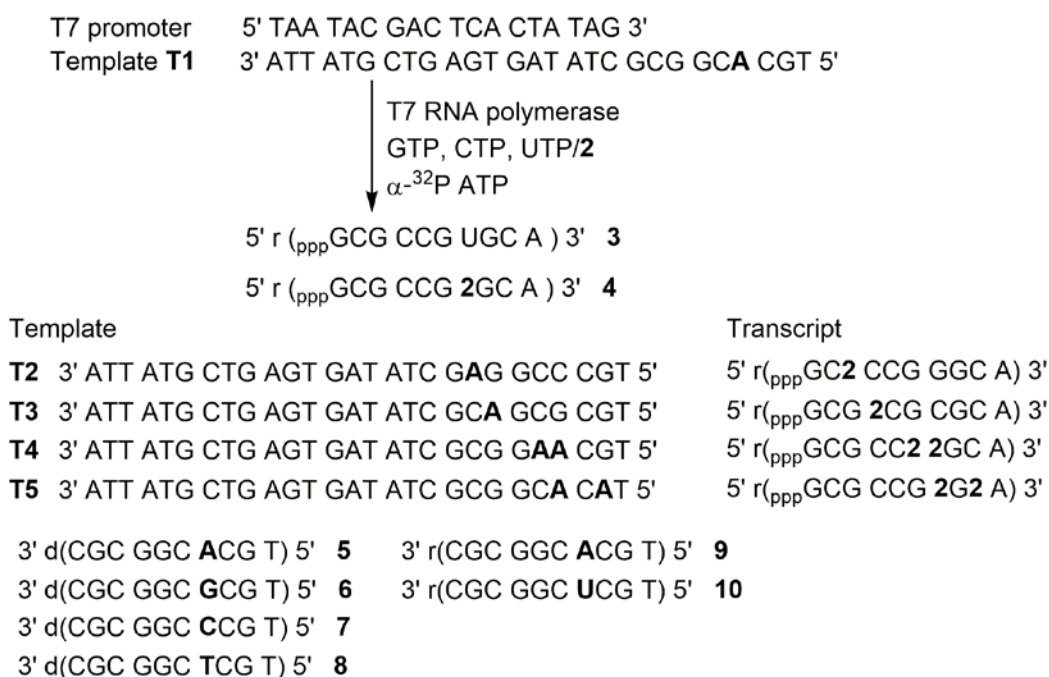
5.2.3 Enzymatic incorporation of nucleoside 1 into RNA ON

As the preliminary photophysical properties revealed the microenvironment sensitivity of the probe **1**, we sought to incorporate it into RNA oligonucleotide for further use. Although different methods have been used to incorporate modified nucleoside into RNA, enzymatic incorporation is one of the principal methods to synthesize modified RNA. To examine whether this probe can be incorporated into RNA ONs by T7 RNA polymerase-catalyzed transcription reaction, the triphosphate of the nucleoside **1** was synthesized (Scheme 2). *In vitro* transcription reactions were performed as depicted in figure 6, where templates **T1**–**T5** were designed such that one or two dA residues at different positions (near the promoter region and away from the promoter region) would direct a single or multiple modifications into RNA transcripts. Promoter template duplexes were formed by annealing templates **T1**–**T5** with T7 RNA polymerase consensus 18 mer promoter sequence and transcription reactions were performed in presence of GTP, CTP, UTP/**2** and α-^{32}P ATP (Figure 6). Purposefully, dT residue was placed at the 5' end of all templates so that each full-length transcript will have an α-^{32}P label at the 3' end, which can be resolved by denaturing polyacrylamide gel electrophoresis and visualized by phosphor imaging. Unsuccessful transcription reactions will produce shorter transcripts with no α-^{32}P label, which won't appear in the image.


Scheme 2. Synthesis of 5-methoxybenzofuran modified uridine triphosphate **2**

```
T7 promoter      5' TAA TAC GAC TCA CTA TAG 3'
Template T1      3' ATT ATG CTG AGT GAT ATC GCG GCA CGT 5'
                                |
                                |  T7 RNA polymerase
                                |  GTP, CTP, UTP/2
                                ↓  α-³²P ATP

                         5' r (ppp GCG CCG UGC A ) 3'   3
                         5' r (ppp GCG CCG 2GC A ) 3'   4
```

Template		Transcript
T2	3' ATT ATG CTG AGT GAT ATC G**A**G GCC CGT 5'	5' r(ppp GC**2** CCG GGC A) 3'
T3	3' ATT ATG CTG AGT GAT ATC GC**A** GCG CGT 5'	5' r(ppp GCG **2**CG CGC A) 3'
T4	3' ATT ATG CTG AGT GAT ATC GCG G**AA** CGT 5'	5' r(ppp GCG CC**2** **2**GC A) 3'
T5	3' ATT ATG CTG AGT GAT ATC GCG GC**A** C**A**T 5'	5' r(ppp GCG CCG **2**G**2** A) 3'

```
3' d(CGC GGC ACG T) 5'  5      3' r(CGC GGC ACG T) 5'  9
3' d(CGC GGC GCG T) 5'  6      3' r(CGC GGC UCG T) 5'  10
3' d(CGC GGC CCG T) 5'  7
3' d(CGC GGC TCG T) 5'  8
```

Figure 6. Incorporation of triphosphate **2** into RNA ONs by transcription reaction in presence of different DNA templates **T1**–**T5**. ONs **3**–**10** have been used for this study. "d" represents the DNA ONs and "r" represent the RNA ONs. See experimental section for details.

Incorporation efficiency of **2** into RNA oligonucleotide in presence of different templates was determined with respect to the transcription reaction in presence of natural UTP for a particular template. In presence of template **T1**, which would direct the incorporation of **2** at the position away from the promoter region, the full-length transcript **4** was formed with very good efficiency (81%) with small amounts of N+1 and N+2 bands, which are due to non-templated incorporation of ribonucleotides (Figure 7, lane 2). Moreover, retardation in the band migration of modified transcript compared to natural one indicated the incorporation of higher molecular weight nucleoside (Figure 7, compare lane 1 and 2). Transcription reaction in absence of both UTP and **2** yielded no full-length product (Figure 7, lane 3). This observation confirmed that production of the full-length transcript was not from any random misincorporation. Interestingly, when the reaction was performed in presence of an equimolar concentration of modified triphosphate **2** and UTP, both the triphosphate were incorporated by T7 polymerase with a slightly higher preference for natural UTP (Figure 7, lane 4). In presence of template **T2** and **T3** where the modification was incorporated near the promoter region, the yield of full-length product was poor (Figure 7 lane 6 and 8). Interestingly, transcription reaction performed with template **T4** produced doubly-labeled full-length transcript in high yields (82%). But incorporation of nucleoside **2** at two alternative positions in presence of template **T5** was quite less efficient (14%).

Together, depending on the positions of incorporation, transcription reaction using T7 RNA polymerase can be used to prepare 5-methoxybenzofuran-modified RNA ONs. Large-scale transcription reaction was performed with template **T1** to isolate the modified RNA ON. The transcript was purified by polyacrylamide gel electrophoresis, and its purity and integrity were confirmed by HPLC and MALDI TOFF mass analysis, respectively (Figure 8A and 8B). Moreover, HPLC and mass analysis of the ribonucleosides obtained from enzymatic digestion of transcript **4** established the presence of nucleoside **1** in the transcript (Figure 9 and Table 2).
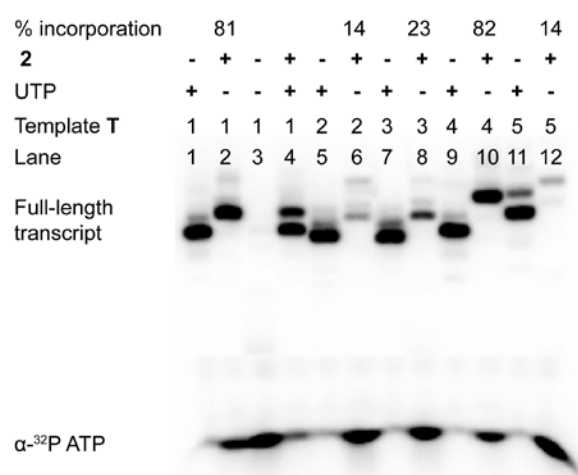
Figure 7. Denaturing polyacrylamide gel electrophoresis of the transcripts obtained from *in vitro* transcription reactions with the DNA templates **T1**–**T5** in the presence of UTP and **2**. The incorporation efficiency of **2** into RNA oligonucleotide by T7 RNA polymerase is determined with respect to amount of full length product formed in the presence of UTP. All reactions were performed in duplicate and standard deviations in the yields were ≤ 2%. See Experimental Section for details.

.
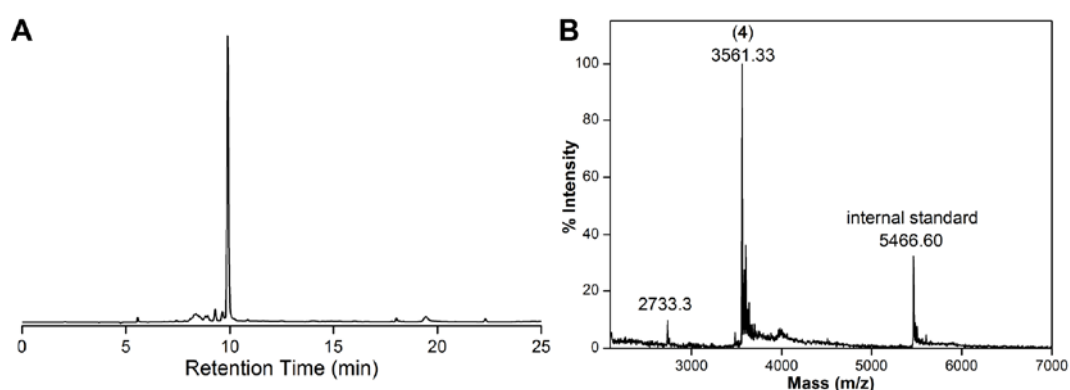
Figure 8. (A) HPLC chromatogram of PAGE purified transcript **4**. Mobile Phase: A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B: acetonitrile. Flow rate: 1 mL/min. Gradient: 0–50% B in 20 min and 50–100% B in 5 min. (**B**) MALDI-TOF spectrum of transcript **4** calibrated relative to the +1 and +2 ions of an internal 18-mer DNA ON standard (m/z for +1 and +2 ions are 5466.6 and 2733.3, respectively). m/z calculated for **4**: 3561.1 [M]; found: 3561.3
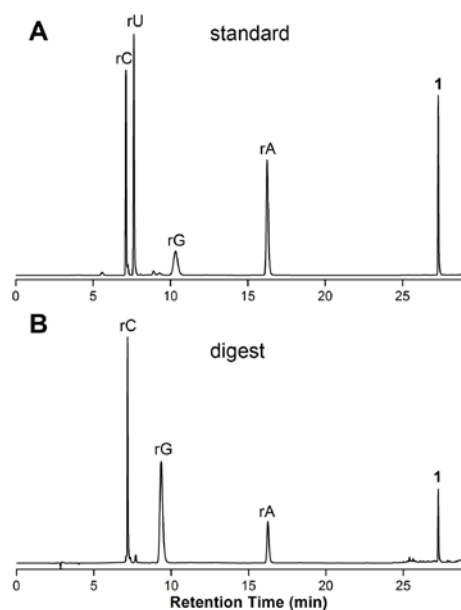
Figure 9. HPLC chromatogram of ribonucleosides obtained from enzymatic digestion of transcript **4**. (**A**) Natural ribonucleosides rC, rU, rG, rA and nucleoside **1** mix. (**B**) Digest of transcript **4**. See experimental section more details.

Table 2. MALDI-TOF mass analysis of HPLC fractions of transcript **4** digest

HPLC fraction of the digest	Calculated mass	Observed mass
rC	$C_9H_{13}KN_3O_5$: 282.0 $[M+K]^+$	282.0
rG	$C_{10}H_{13}N_5NaO_5$: 306.1 $[M+Na]^+$	306.0
rA	$C_{10}H_{13}N_5NaO_4$: 290.1 $[M+Na]^+$	290.0
1	$C_{18}H_{18}KN_2O_8$: 429.1 $[M+K]^+$	429.0

5.2.4 Fluorescence detection of dT-U and U-U mismatch in nucleic acid duplexes

Emissive nucleoside probes after incorporation into ONs can experience several interactions and effects like stacking and H-bonding interactions, solvation-desolvation and rigidification-derigidification and electron transfer process, which influence the fluorescence outcome. In order to understand the behaviour of the nucleoside analog in different neighboring base environment, a model RNA transcript **4** labeled with **1** was synthesized and hybridized to complementary DNA and RNA ONs. The duplexes were designed such that the emissive base was placed opposite to complementary and mismatched bases. It was observed that nucleoside **1** after incorporation into RNA ON **4** exhibited a significant enhancement in fluorescence intensity compared to the free nucleoside (Figure 10A). This observation is noteworthy because most of the fluorescent nucleoside probes, in particular, when placed in

163

the vicinity of guanosine exhibit very low fluorescence.[26] RNA-DNA duplexes, wherein the emissive analogue was placed opposite to mismatched bases (dG and dC), dispayed similar or slightly higher fluorescence intensity as compared to the perfectly matched duplex. Notably, when placed opposite to dT residue, nucleoside **1** showed a significant enhancement in fluorescence intensity (~5-fold) with respect to perfect duplex (Figure 10A). Thus the emissive analogue **1** was able to selectively detect the dT-U mismatch in a DNA-RNA duplex. Rewardingly, the nucleoside probe also reported the presence of a U-U mismatch in an RNA-RNA duplex with a similar enhancement in fluorescence intensity as above (Figure 10B). The enhancement in fluorescence intensity without any change in emission maximum in case of the dT-U and U-U mismatch could be because of very small change in polarity of the surrounding environment, rigidification or destacking of the probe.
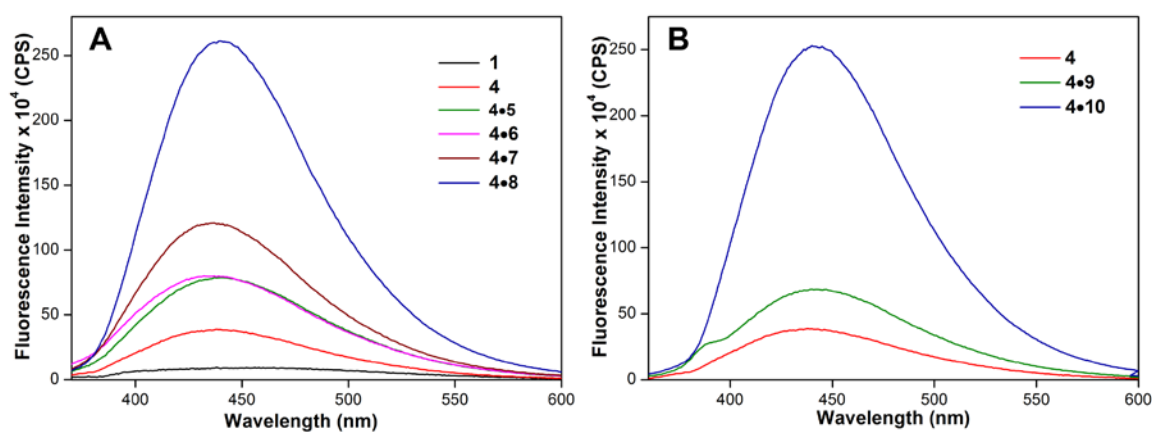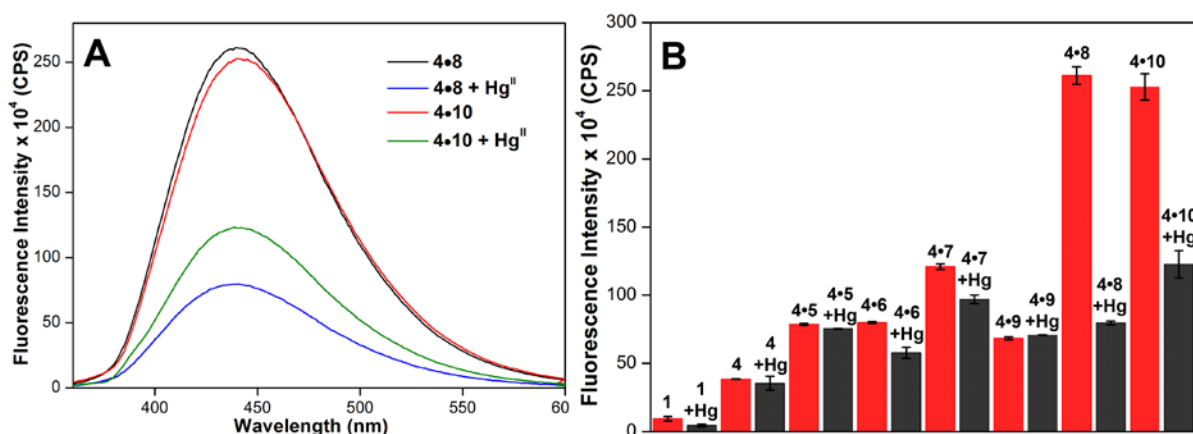

Figure 10. (**A**) Emission spectra of nucleoside **1**, transcript **4** and its duplexes with complementary DNA sequences, where nucleoside dA and mismatch bases dG, dC and dT are placed opposite to nucleoside **1**. (**B**) Emission spectra of transcript **4** and its duplexes with complementary RNA sequences where rA and rU are placed opposite to nucleoside **1**. Samples (1 µM) were excited at 340 nm with excitation and emission slit width 7 nm and 8 nm, respectively.

## 5.2.5 Fluorescence detection of Hg[II] mediated base pairing in DNA-RNA and RNA-RNA duplexes.

Hg[II] ion forms dT-Hg-dT and U-Hg-U noncanonical base pairs by coordinating with N3 of two thymine or uracil residues, and these interactions have received growing attention due to their biological and material applications.[8] As nucleoside **1** photophysically reports mispairing, we envision that upon formation of Hg[II] mediated base-pair, the fluorescence properties of the probe could vary due to changes in the environment of the probe. Hence, the

responsiveness of the probe can be used to device a simple fluorescence method to sense metallo base-pair formation in nucleic acid duplexes.

To examine the feasibility of this approach steady-state fluorescence studies of nucleoside **1**, RNA ON **4**, perfect duplexes (**4•5** and **4•9**) and duplexes with mismatches (**4•6**, **4•7**, **4•8** and **4•10**) were performed in presence or absence of 1 equivalent Hg^{II} ions. A significant reduction in fluorescence intensity was observed in presence of Hg^{II}, only in the case of duplexes **4•8** with dT-**1** and **4•10** with U-**1** mispair (Figure 11A and 11B). However, only minor changes in fluorescence intensity were observed for nucleoside **1**, RNA ON **4** and other perfect and mismatch duplexes (Figure 11B). These results confirm that Hg^{II} ion selectively binds to dT-U and U-U mismatches and the emission property of the probe was highly sensitive to Hg^{II} mediated base-pair formation. To estimate the metal ion specificity of this detection assay, fluorescence studies of duplexes **4•8** and **4•10** were performed in the presence of 1 equivalent different metal ions (Ag^{I}, Hg^{II}, Ca^{II}, Cu^{II}, Cd^{II}, Pb^{II}, Mn^{II}, Mg^{II}, Co^{II}, Zn^{II}, Fe^{II} and, Ni^{II}). Other than Hg^{II} ion, none of the metal ions tested affected the fluorescence of the probe-labeled mismatched pair in the duplexes (Figure 12). These findings confirm that the nucleoside probe reliably and selectively report the Hg^{II} prompted metallo base-pair formation in duplexes. The reduction in fluorescence intensity in the presence of Hg^{II} is more likely due to the base-paired state of the nucleoside analog, which experiences more stacking interaction and quenching effect from adjacent guanosine bases.[26]


Figure 11. (**A**) Fluorescence spectra of duplexes **4•8** and **4•10** (1 µM) in presence and absence of of 1 µM Hg^{II}. (**B**) Bar diagram for fluorescence studies of nucleoside **1**, transcript **4** and its duplexes (1 µM) with complementary DNA and RNA sequences in presence and absence of 1 µM Hg^{II} ion. Samples were excited at 340 nm in 10 mM sodium cacodylate buffer pH 7.0 containing 500 mM $NaNO_3$ with excitation and emission slit width 7 nm and 8 nm, respectively.
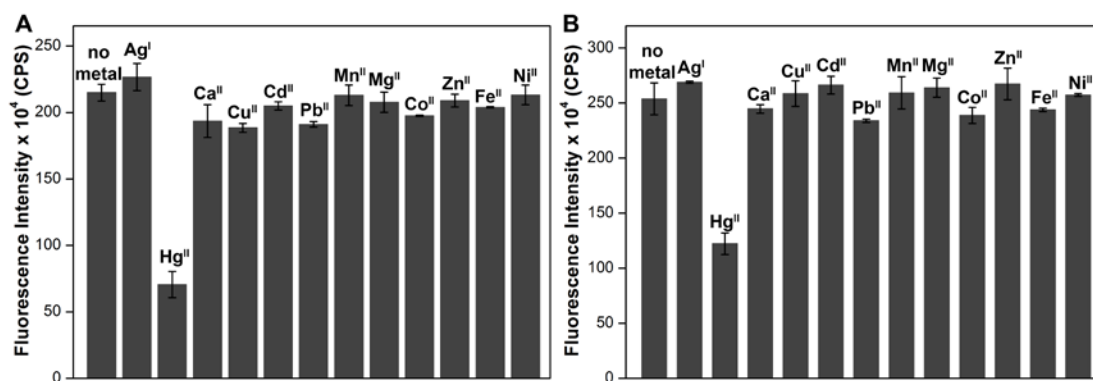
Figure 12. Fluorescence spectra of duplexes (1 µM) (**A**) **4•8** containing dT-**1** mismatch and (**B**) **4•10** containing U-**1** mismatch in absence and presence of different metal ions (1 µM, 1 equivalent) in 10 mM sodium cacodylate buffer pH 7.0 containing 500 mM NaNO₃. Samples were excited at 340 nm with excitation and emission slit width 7 nm and 8 nm, respectively.

5.2.6 Thermal denaturation studies of duplexes

Formation of Hg^{II} mediated base pair was confirmed by performing thermal denaturation studies of both control and unmodified duplexes in presence and absence of equimolar concentration of Hg^{II} ions. In the presence of Hg^{II} ions, both control duplex (**3•8**) with dT-U mismatch and modified duplex (**4•8**) with dT-**1** mismatch showed 18 ℃ and 17 ℃ increase in thermal melting temperature, respectively, due to stabilization of the duplexes by metal-mediated base pairing (Figure 13 and Table 3). Similarly Hg^{II} caused a noticeable increase in melting temperature (ΔT_m = 18–21 ℃) for RNA-RNA duplex **3•10** with U-U mismatch and **4•10** with U-**1** mismatch in the presence of Hg^{II} ions. As expected, perfect DNA-RNA and RNA-RNA duplexes (**3•5**, **4•5**, **3•9** and **4•9**) showed only minor differences in T_m after addition of Hg^{II} ion (Figure 14 and Table 3). These studies confirmed the formation of dT-Hg-**1** and U-Hg-**1** formation and ascertained that the fluorescence response is a true reflection of respective canonical or non-canonical metallo base-pair formation.
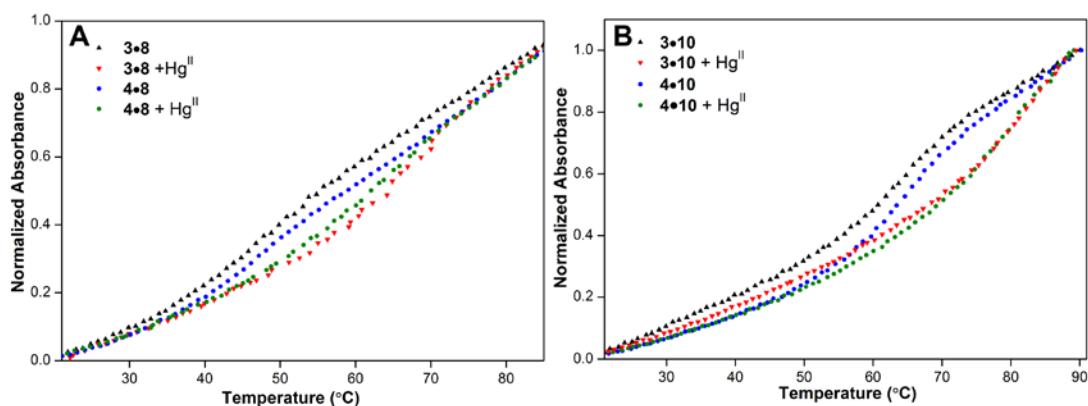
Figure 13. UV-thermal melting profile of duplexes (5 µM) (**A**) **3•8**, **4•8** with dT-U or dT-**1** mismatch and (**B**) **3•10**, **4•10** with U-U or U-**1** mismatch in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM NaNO₃ either in presence or absence of Hg^{II} (5 µM).

166
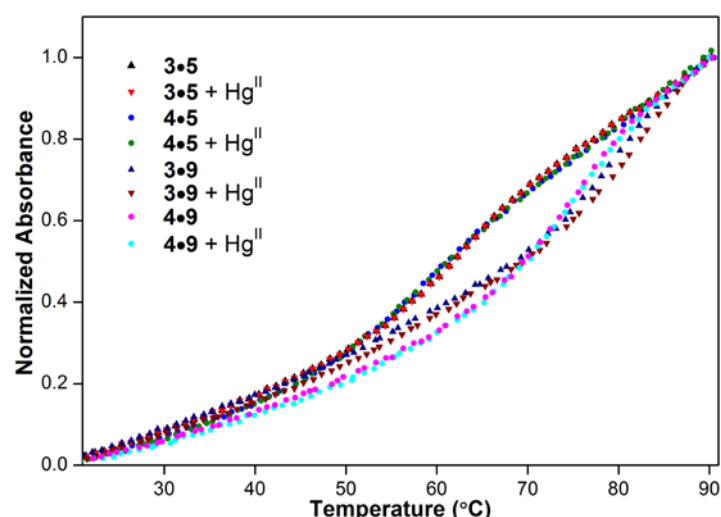
Figure 14. UV-thermal melting profile of perfect duplexes (5 µM) **3•5**, **4•5, 3•9,** and **4•9** in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM NaNO$_3$ either in presence or absence of Hg^{II} (5 µM).

**Table 3.** T_m values of control and modified duplexes.

	ON duplexes	T_m (°C) in absence of Hg^{II}	T_m (°C) in presence of 1 equivalent Hg^{II}
control	**3•5**	59.7 ± 0.9	61.0 ± 0.4
	3•8	46.1 ± 0.8	63.7 ± 0.6
	3•9	77.8 ± 0.4	80.3 ± 0.9
	3•10	61.5 ± 0.5	82.6 ± 0.5
modified	**4•5**	57.5 ± 0.1	57.7 ± 0.8
	4•8	44.0 ± 0.8	60.7 ± 0.9
	4•9	74.8 ± 0.6	73.5 ± 0.5
	4•10	63.3 ± 0.2	80.9 ± 0.2

5.2.7 ^{1}H NMR analysis of duplexes with dT-1 and U-1 mismatches in presence of Hg^{II} ion

The formation of dT-Hg-**1** and U-Hg-**1** were further confirmed by ^{1}H NMR analysis of duplexes **4•8** and **4•10** with increasing concentration of Hg^{II} ions. In the absence of Hg^{II} ions, the duplex **4•8** with dT-**1** mismatch and duplex **4•10** with U-**1** mismatch exhibited imino proton resonance at 10.6 ppm and 10.4 ppm, respectively (Figure 15A and 15B). These signals correspond to thymidine and uridine mismatched N3-H protons.[15c,21c] Upon increasing the Hg^{II} ion concentration, the intensity of the peaks gradually decreased and in presence of 1 equivalent Hg^{II} ion, the signals disappeared completely. Other imino proton signals of Watson-Crick pairs also shifted due to stabilization of the duplex by Hg^{II}

coordination (Figure 15A and 15B). Taken together, these results further confirmed the selective binding of Hg^{II} to dT-**1** and U-**1** mispairs.
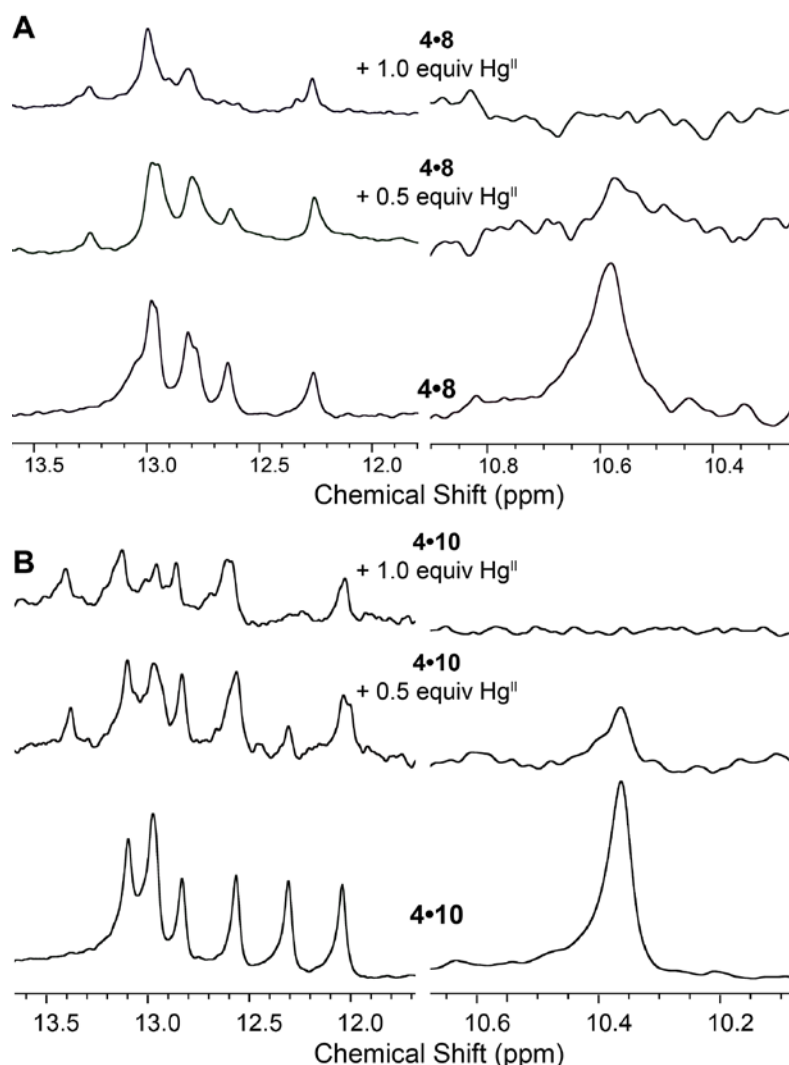

Figure 15. Imino ^{1}H NMR analysis of modified duplexes (**A**) **4•8** with dT-**1** mismatch (50 µM) and (**B**) **4•10** with U-**1** mismatch in absence and presence of 0.5 equivalent (25 µM) and 1.0 equivalent (50 µM) Hg^{II} ion. Spectra were recorder at 20 ℃ in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM NaNO$_3$.

## 5.2.8 Estimation of Hg^{II} ion binding to mismatches by fluorescence

The ability of the probe to detect the Hg-mediated base pairing was used in determining the binding affinity of mercury to mismatches. Addition of increasing concentrations of Hg^{II} ions to duplexes **4•8** and **4•10** resulted in a dose-dependent quenching in fluorescence intensity (Figure 16). The apparent dissociation constant K_d obtained for **4•8** (57 ± 7 nM) was lower than that for **4•10** (104 ± 2 nM) suggesting that Hg^{II} ion has a higher binding affinity for dT-

U mismatch in a DNA-RNA duplex than U-U mismatch in an RNA-RNA duplex (Figure 17). Importantly, the K_d values for dT-U and U-U mismatches at low nM range matches well with the reported K_d or detection limit of other DNA duplexes with T-T mismatches.[11a,21] These observations are important because it showed that Hg^{II} can bind with very high affinity to dT-U and U-U mismatches and can be considered as a potential non-canonical base-pair in enzymatic reactions like transcription and reverse-transcription as well as in developing metallo-RNA based materials.
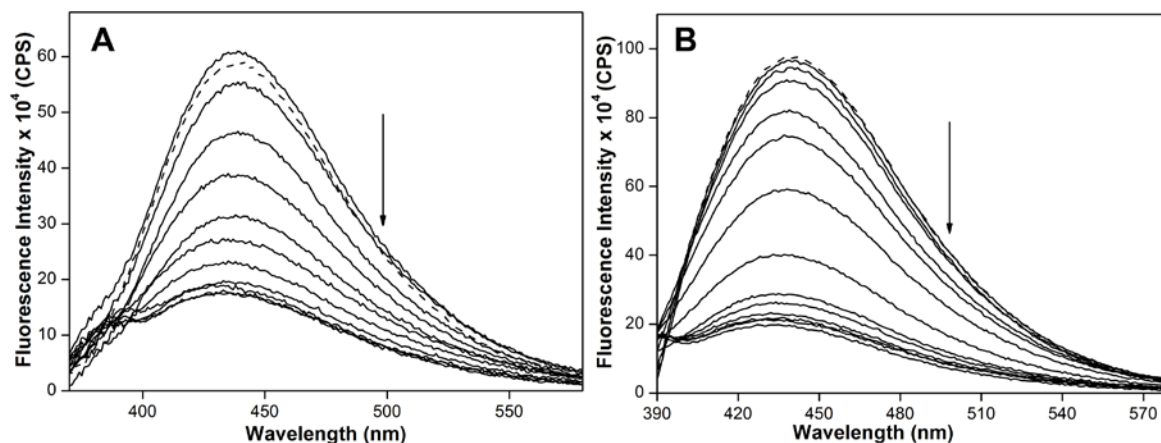
Figure 16. Emission spectra for the titration of fluorescently labelled duplex (**A**) **4•8** and (**B**) **4•10** (0.18 µM) with increasing concentration of Hg^{II} in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM NaNO$_3$. Samples were excited at 340 nm with excitation and emission slit width 10 nm and 12 nm, respectively. (See experimental section for details).
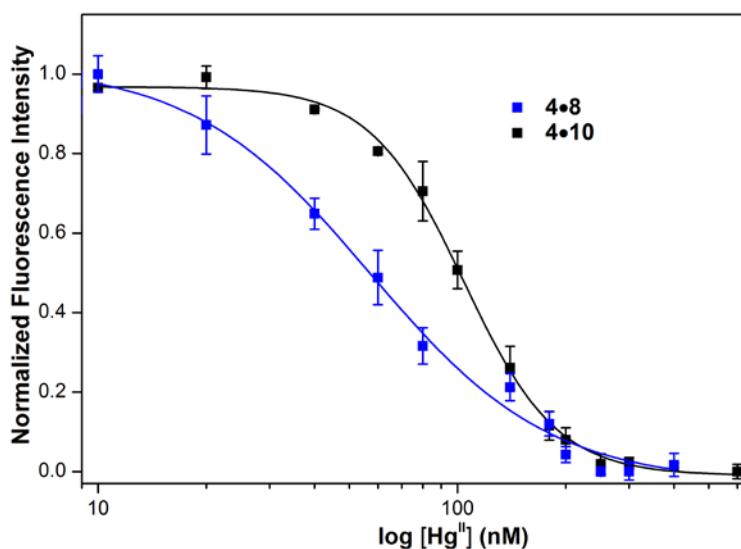
Figure 17. Curve fit for the titration of duplexes **4•8** and **4•10** with Hg^{II}. Normalized fluorescence intensity at respective emission maximum (λ_{em} = 439 nm for **4•8** and λ_{em} = 440 nm for **4•10**) plotted against log $[Hg^{II}]$ (see experimental section for details).

169

5.3 Conclusions

We have developed a new fluorescent nucleoside probe, 5-methoxybenzofuran-modified uridine analogue, whose emission properties are highly sensitive to polarity and viscosity changes. Even though this probe is practically non-emissive in water, a small decrease in solvent polarity enhances its quantum yield significantly. This property of the probe will be highly useful in studying ligand or protein binding to nucleic acids that result in small changes in micropolarity. The nucleoside triphosphate served as a good substrate for RNA polymerase and was amenable to incorporate into RNA transcripts with moderate to good efficiency. Rewardingly, the nucleoside analogue incorporated into RNA transcripts was minimally perturbing, and specifically signalled the presence of dT-U and U-U mismatches with enhancement in fluorescence. This property of the probe enabled the detection and estimation of the mercury-mediated base pairing in duplexes at nanomolar concentrations. The difference in binding affinity exhibited by Hg^{II} ion to mismatches as confirmed by using our probe may have implications on the interference of metals ions on replication, transcription and reverse transcription. Taken together, our results demonstrate that the proficiency of the emissive nucleoside analogue should complement existing probes in analyzing nucleic acids.

5.4 Experimental Section

5.4.1 Materials

N,N,N',N'-tetramethylethylenediamine, *n*-butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) dichloride and all reagents (Bio-Ultra grade) used for buffer preparation were purchased from Sigma-Aldrich. 5-methoxybenzofuran (**1a**) was synthesized following a previously reported procedure.[22] 5-iodouridine was synthesized following the previously established method.[27] $CoCl_2.6H_2O$, $FeCl_2.4H_2O$, $CuSO_4$, $ZnCl_2$, $NiCl_2.6H_2O$, $PbCl_2$, $CdCl_2.2.5\ H_2O$, $MnCl_2$, $CaCl_2.2H_2O$, $MgCl_2.6H_2O$ and $Hg(ClO_4)_2.6H_2O$ were used for respective metal ion source and purchased from either Sigma- Aldrich or Alfa Aesar. T7 RNA polymerase, ribonuclease inhibitor (RiboLock), NTPs, RNase A and RNase T1 were acquired from Fermentas, Thermo Fisher Scientific. $PoCl_3$ was purchased from Acros Organic and was freshly distilled before use. Radiolabeled α-^{32}P ATP (2000 Ci/mmol) was obtained from the Board of Radiation and Isotope Technology, Government of India. Calf intestinal alkaline phosphatase (CIP) and snake venom phosphodiesterase I were acquired from Invitrogen and Sigma-Aldrich, respectively. DNA ONs **5**, **6**, **7** and **8** were purchased from Integrated DNA Technology. RNA oligonucleotide **9** and **10** were purchased

from Dharmacon RNAi Technologies, deprotected according to the provider's procedure. All custome DNA and RNA oligonucleotides were purified by denaturing polyacrylamide gel desalted on Sep-Pak Classic C18 cartridges (Waters Corporation). Autoclaved Millipore water was used in all biophysical analysis.

5.4.2 Instrumentation

Mass analysis was performed either on an Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer or on a Water Synapt G2 High Definition mass spectrometers. HPLC analysis was done using Agilent Technologies 1260 Infinity HPLC. Reverse-phase flash chromatographic (C18 RediSepRf column) purifications were carried out using Teledyne ISCO, Combi Flash Rf. Absorption spectra were recorded on a Shimadzu UV-2600 spectrophotometer. Steady-state fluorescence spectra were recorded on a Fluoromax-4 spectrophotometer. UV-thermal melting analysis of ONs was conducted on Cary 300 Bio UV-Vis spectrophotometer. CD analysis was performed on a JASCO J-815 CD spectrometer. NMR spectra of small molecules were recorded on a Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. NMR spectra of ON duplexes were recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe and processed using Bruker TopSpin Software.

5.4.3 Synthesis of 5-(5-methoxybenzofuran)-uridine and corresponding triphosphate

5.4.3.1 Tributyl(5-methoxybenzofuran-2-yl)stannan (1b)

TMEDA (0.58 g, 5.00 mmol, 1.2 equiv) was added to a solution of 5-methoxybenzofuran (**1a**) (0.62 g, 4.16 mmol, 1.0 equiv) in dry THF (20 ml) and the mixture was allowed to stir for 30 min at -78 °C. n-BuLi (2.40 ml, 2M solution in hexane, 5.00 mmol, 1.2 equiv) was added dropwise to it at -78 °C and reaction was brought to room temperature over a period of 2 h. The reaction mixture was again cooled to -78 °C and n-Bu$_3$SnCl (1.35 ml, 5.00 mmol, 1.2 equiv) was added dropwise to it. The reaction mixture was stirred for 1 h. The reaction was quenched with ammonium chloride solution (25 ml) and extracted two times with diethyl ether (2 x 25 ml). Organic layer was dried over sodium sulphate and evaporated to oily liquid. Crude liquid was purified by normal-phase flash column chromatography (petroleum ether) to afford **1a** as clear oil (1.28 g, 70%). TLC (10% EtOAc in petroleum ether) $R_f = 0.7$; ^{1}H NMR (400 MHz, CDCl$_3$): δ (ppm) 7.39 (d, $J = 9.8$ Hz, 1H), 7.03 (d, $J = 2.4$ Hz, 1H), 6.86–6.83 (m, 2H), 3.85 (s, 3H), 1.64–1.57 (m, 6H), 1.41–1.32 (m, 6H), 1.24–1.07 (m, 6H), 0.91 (t, $J = 7.4$ Hz, 9H); ^{13}C NMR (100 MHz, CDCl$_3$): δ (ppm) 166.6, 155.6, 153.9, 128.7, 118.2,

112.2, 111.4, 102.7, 56.1, 29.1, 27.3, 13.8, 10.3; MALDI-TOF: m/z Calculated for $C_{21}H_{34}FKO_2Sn$ [M+K]$^+$ = 477.12, found = 477.02.

5.4.3.2 5-(5-methoxybenzofuran)-uridine (1)

A mixture of 5-iodo-uridine[27] (0.25 g, 0.67 mmol, 1 equiv) and *bis*(triphenylphosphine)-palladium(II) chloride (0.02 g, 0.034 mmol, 0.05 equiv) and tributyl(5-fluorobenzofuran-2-yl)stannane (0.73 g, 1.67 mmol, 2.5 equiv) was dissolved in degassed anhydrous dioxane (8 ml). The mixture was heated at 90 °C for 2 h under N_2 atmosphere and filtered through celite pad. The celite pad was washed with methanol (2x15 ml). The filtrate was evaporated and resulting residue was purified by reverse phase column chromatography (50 % methanol in water) to afford product **1** as white solid compound (0.17 g, 65%).TLC (15% methanol in DCM) R_f = 0.40; ^{1}H NMR (400 MHz, d_6 -DMSO): δ (ppm) 8.83 (s, 1H), 7.4 (d, *J* = 8.8 Hz, 1H), 7.28 (br, 1H), 7.14 (d, *J* = 2.8 Hz, 1H), 6.87 (dd, J_1 = 8.8 Hz, J_2 = 2.4 Hz), 5.87 (d, *J* = 4 Hz, 1H), 5.50 (br, 1H), 5.37 (br, 1H), 5.15 (br, 1H), 4.15–4.09 (m, 2H), 3.96–3.95 (m,1H), 3.81 (br, 1H), 3.77 (s, 3H), 3.68–3.65 (m, 1H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 160.4, 155.7, 149.8, 149.7, 147.9, 137.2, 129.5, 112.6, 111.2, 104.9, 104.2, 103.6, 88.8, 84.7, 74.4, 69.5, 60.1, 55.6; HRMS: m/z Calculated. for $C_{18}H_{18}N_2NaO_8$ [M+Na]$^+$ = 413.0961, found = 413.0968.

5.4.3.3 5-(5-methoxybenzofuran)-uridine triphosphate (2)

To a solution of 5-methoxybenzofuran (77 mg, 0.20 mmol, 1.0 equiv) in trimethyl phosphate, freshly distilled POCl$_3$ (46 µL, 0.49 mmol) was added in ice-cold condition. The solution was stirred for 24 h at ~ 4 °C. After 24 h, it was observed that the starting material was not consumed completely. Bis(tributylammonium) pyrophosphate solution in DMF (0.5 M, 1.9 ml, 0.99 mmol) and tributylamine (0.52 ml, 2.17 ml, 11 equiv) were speedily and simultaneously added to the reaction mixture in ice-cold condition. The reaction mixture was stirred for 30 min at 4 °C and quenched with 1 M triethy ammonium bicarbonate buffer (TEAB, pH 7.5, 15 ml) and washed with ethyl acetate (2 x 15 ml). The aqueous layer was evaporated to dryness and purified using an DEAE sephadex-A25 anion exchange column (10 mM–1M TEAB buffer, pH 7.5) followed by reverse phase flash chromatography (C18 Redi*SepRf*, 0–50% acetonitrile in 100 mM triethylammonium acetate buffer, pH 7.2, flow rate 6 ml/ min, 55 min). Evaporation of the appropriate fraction resulted into the desired triphosphate **2** as tetraethyl ammonium salt (31 mg, 15 %). ^{1}H NMR (400 MHz, D$_2$O): 7.82

(s, 1H), 7.39–7.38 (m, 1H), 6.93–6.83 (m, 3H), 5.79 (br, 1H), 4.41–4.34 (m, 5H), 3.79 (s, 3H); ^{13}C NMR (100 MHz, D$_2$O): δ (ppm) 161.63, 154.98, 150.41, 150.36, 148.47, 136.38, 129.29, 112.89, 111.77, 106.11, 105.16, 103.89, 89.61, 83.13, 73.52, 69.80, 65.68, 55.83; ^{31}P NMR (162 MHz, D$_2$O): -6.46 (br, P$_\gamma$), -11.51 (d, J = 14.42 Hz, P$_\alpha$), -22.32 (br, P$_\beta$); MALDI-TOF: m/z Calculated for C$_{18}$H$_{21}$N$_2$NaO$_{17}$P$_3$ [M+Na]$^+$ = 653.00, found = 653.10.

5.4.4 Photophysical analysis of 5-methoxybenzofuran-modified uridine analogue 1 in different solvents.

5.4.4.1 UV-absorption and steady-state fluorescence studies

UV-absorption spectra of nucleoside **1** (25 µM) was recorded on a Shimadzu UV-2600 spectrophotometer in quartz cuvette (Hellma, path length 1 cm). Each UV sample contained 2.5% DMSO. Steady-state fluorescence study of nucleoside **1** (5 µM) in different solvent was performed in micro fluorescence cuvette (Hellma, path length 1 cm) on Fluoromax-4 spectrofluorometer. In this study samples were excited at their lowest energy absorption maximum (Table 1). Each fluorescence sample contained 0.5% DMSO. Anisotropy values (r) of nucleoside **1** in different solvents were determined by analysing the data using software provided with the instrument Fluoromax-4 spectrofluorometer. Each anisotropy value was an average of 10 successive measurements. All UV-absorption, steady-state fluorescence and anisotropy measurements were performed in triplicate.

5.4.4.2 Time-resolved fluorescence study

Time resolved fluorescence study of nucleoside **1** was performed using time correlated single photon counting (TCSPC) setup (Horiba Jobin Yvon, U.S.A.). Nucleoside **1** in methanol, dioxane, ethylene glycol and glycerol was excited by using 375 nm diode laser source (IBH, UK, NanoLED-375L). Fluorescence signal at respective emission maxima was collected. All studies were done in triplicate and lifetimes were calculated by fitting the decay profile using IBH DAS6 software. The χ^2 value for all the curve fits was found to be nearly one.

5.4.4.3 Quantum yield

Quantum yield of nucleoside **1** in different solvents was determined relative to 2-aminopurine as the standard. Following equation was used to calculate the quantum yield.[28]

$$\Phi_{F(x)} = \left(A_s/A_x\right)\left(F_x/F_s\right)\left(n_x/n_s\right)^2 \Phi_{F(s)}$$

Where s is the standard, x is the modified nucleoside, A is the absorbance at excitation wavelength, F is the area under the emission curve, n is the refractive index of the solvent, and Φ_F is the quantum yield. Quantum yield of 2-aminopurine in water is 0.68.

5.4.5 Transcription reaction with α-^{32}P ATP

Promoter-template duplexes were formed by annealing a 1:1 solution of DNA template **T1**–**T5** (5 µM) and RNA polymerase consensus 18 mer promoter DNA sequence in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) at 90 °C for 3 min and allowing it to attain room temperature. The solution was placed on an ice bath for 30 min and stored at -40 °C. Transcription was performed at 37 °C in 40 mM Tris-HCl buffer (pH 7.8) using 250 nM annealed promoter-template duplexes, 10 mM NaCl, 10 mM MgCl$_2$, 10 mM of dithiothreitol (DTT), 2 mM spermidine, 1 U/µL, RNase inhibitor (Riboblock), 1 mM GTP, CTP, UTP and or modified UTP **2**, 20 µM ATP, 5 µCi α-^{32}P ATP and 3 U/µL (total 60 units) T7 RNA polymerase in a total 20 µL reaction volume. After 3.5 h, the reaction was quenched using 20 µL of loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8). The samples were heated at 75 °C for 3 min and then cooled on an ice bath. The samples (4 µL) were loaded on a sequencing 18% denaturing polyacrylamide gel and were electrophoresed at a constant power (11 W) for nearly 4 h. The bands corresponding to the radioactive products were imaged using an X-ray film. The relative transcription efficiency was calculated using GeneTools software from Syngene. The % of incorporation of modified UTP **2** into RNA oligonucleotide by T7 RNA polymerase was determined considering the transcription efficiency in presence of all natural NTPs as 100 %. All reactions were performed in duplicate and the errors in yields were ≤ 2%. See Figure 7.

5.4.6 Large scale transcription

In order to isolate and characterize modified transcript **4**, large scale transcription reaction with a total volume of 250 µL was performed with the template **T1**. The transctiption reaction was performed in presence of 2 mM GTP, ATP, CTP and modified UTP **2**, 20 mM

MgCl$_2$, 0.4 U/µL of RNase inhibitor (RiboLock), 800 units T7 RNA polymerase and 300 nM promoter template duplex. The transcription reaction was performed at 37 °C for 12 h and the white precipitation of pyrophosphate was observed. The reaction volume was reduced to almost one third of the total volume using speed vac. 40 µL loading buffer (10 mM Tris HCl, 7 M urea, 100 mM EDTA, pH 8.0) was added to the residual mixture and loaded onto to the 20% polyacrylamide gel electrophoresis and run at 25 W constant power for 6 h. Gel was UV shadowed, the appropriate band corresponding to transcript was cut and transferred to the Poly-Prep column (Bio-Rad). The gel pieces were crushed; the transcript was extracted with 0.5 M ammonium acetate for 12 h and desalted using Sep-Pak classic C18 cartridges (Waters). Around 10-12 nmole of transcript **4** (ε_{260} = 91560 M^{-1}cm^{-1}) was isolated from each reaction. The purity of the transcript was examined by HPLC and the transcript was characterized by MALDI TOF mass analysis.

5.4.7 Enzymatic digestion of transcript 4

Transcript **4** (2.5 nmole) was incubated with snake venom phosphodiesterase I (0.01 U), calf intestinal alkaline phosphatase (1 U/µL) and RNase A (0.25 µg), 50 mM Tris-HCl buffer (pH 8.5, 40 mM MgCl$_2$, 0.1 mM EDTA) in a total volume of 100 µL for 12 h at 37 °C. Afterwards, RNase T1 (0.2 U/µL) was added and the sample was incubated for additional 4 h at 37 °C. The resulting mixture obtained from the digestion was analyzed by RP-HPLC using Phenomenex-Luna C18 column (250 × 4.6 mm, 5 micron) at 260 and 330 nm. Mobile phase A: 50mM TEAA buffer (pH 7.3), mobile phase B: acetonitrile. Flow rate: 1 ml/min. Gradient: 0−10% B in 20 min, 10−100% B in 10 min. Further to characterise the ribonucleosides in digest, the fraction corresponding to the individual ribonucleoside was collected and analyzed by mass spectroscopy.

5.4.8 Steady-state fluorescence studies of transcript 4 and its duplexes in presence and absence of metal ions

Transcript **4** (1 µM) or a 1:1 mixture of transcript **4** (1 µM) and its respective complementary DNA or RNA strand was heated at 90 ℃ for 3 min in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM NaNO$_3$. All samples were annealed either in presence or absences of 1 equivalent (1 µM) respective metal ions. Samples were allowed to come to room temperature over a period of 2 h and kept at 4 ℃ for 1 h. Samples were excited at 340 nm with excitation and emission slit width 7 nm and 8 nm, respectively.

5.4.9 Thermal melting analysis of ON duplexes

ON duplexes (5 µM) were formed in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM $NaNO_3$ by annealing as mentioned earlier, either in presence or absence of 1 equivalent Hg^{II} (5 µM). Thermal melting analysis was performed using Cary 300 Bio UV-Vis spectrophotometer. The temperature was increased from 18 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 260 nm.

5.4.10 ^{1}H NMR analysis of ON duplexes

Duplex **4•8** (50 µM) with dT-**1** mismatch and duplex **4•10** (50 µM) with U-**1** mismatch were formed in sodium cacodylate buffer (10 mM, pH 7.0) containing 500 mM $NaNO_3$ similarly like steady-state fluorescence sample. After recording the ^{1}H NMR of the duplexes without Hg^{II} ion, the concentration of Hg^{II} in the sample was increased gradually by adding higher concentration (2.5 mM) of Hg^{II} ion such that final concentration of Hg^{II} was 25 µM (0.5 equivalent) and 50 µM (1 equivalent). In each titration point, after addition of Hg^{II} ion, samples were annealed as mentioned earlier. ^{1}H NMR spectra was recorded at 20 °C on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with BB(F) Double Channel Probe.

5.4.11 Fluorescence binding assay

RNA ON **4** (10 µM) was annealed with equimolar concentration of either DNA ON **8** or RNA ON **10** to form respective duplexes. Then the duplexes (10 µM) were diluted to 0.18 µM with 1X buffer (10 mM sodium cacodylate buffer, pH 7.0 with 500 mM $NaNO_3$). A series of duplex samples (**4•8** or **4•10**, 0.18 µM) containing increasing concentration of the Hg^{II} (10 to 600 nM) was prepared and incubated for 1 h before fluorescence analysis. Samples were excited at 340 nm with an excitation and emission slit width of 10 nm and 12 nm, respectively. The measurements were made in triplicate at 20 °C. The spectrum corresponding to a blank without any ON duplex but containing the respective Hg^{II} concentration was subtracted from each spectrum. From the dose-dependent quenching curves, the apparent dissociation constants (K_d) for the binding of Hg^{II} to respective duplexes were determined by fitting normalized fluorescence intensity (F_N) *versus* log of ligand concentration plot to Hill equation (Origin 8.5).[7f]

$$F_N = \frac{F_i - F_s}{F_0 - F_s}$$

F_i is the fluorescence intensity at each ligand concentration. F_0 and F_s are the fluorescence intensity in the absence of ligand (L) and at saturation point, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$F_N = F_0 + (F_s - F_0)\left(\frac{[L]^n}{[K_d]^n + [L]^n}\right)$$

5.5 References

1. (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579–2619.(b) Srivatsan, S. G.; Sawant, A. A. *Pure Appl. Chem.* **2011**, *83*, 213–232. (c) Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem.* **2013**, *53*, 366–378.
2. (a) Seo, Y. J.; Ryu, J. H.; Kim, B. H. *Org. Lett.* **2005**, *7*, 4931–4933. (b) Okamoto, A.; Saito, Y.; Saito, I. *J. Photochem. Photobiol. C* **2005**, *6*, 108–122. (c) Tainaka, K.; Tanaka, K.; Ikeda, S.; Nishiza, K.-I.; Unzai, T.; Fujiwara, Y.; Saito, I; Okamoto, A. *J. Am. Chem. Soc.* **2007**, *129*, 4776–4784. (d) Gardarsson, H.; Kale, A. S.; Sigurdsson, S. T. *ChemBioChem* **2011**, *12*, 567–575 (e) Ming, X.; Seela, F. *Chem. Eur. J.* **2012**, *18*, 9590–9600.
3. (a) Valis, L.; Amann, N.; Wagenknecht, H.-A. *Org. Biomol. Chem.* **2005**, *3*, 36–38. (b) Shipova, E.; Gates, K. S. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2111–2113. (c) Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784–10785.
4. (a) Greco, N. J.; Sinkeldam, R. W.; Tor, Y. *Org. Lett.* **2009**, *11*, 1115–1118. (b) Zhang, Q; Wang, Y.; Meng, X.; Dhar, R.; Huang, H. *Anal. Chem.* **2013**, *85*, 201–207.
5. (a) Srivatsan, S. G.; Tor, Y. *Tetrahedron* **2007**, *63*, 3601–3607. (b) Riedl, J.; Pohl, R.; Ernsting, N. P.; Orság, P.; Fojta, M.; Hocek, M. *Chem. Sci.* **2012**, *3*, 2797–2806. (c) Dziuba, D.; Pospísil, P.; Matyašovský , J.; Brynda, J.; Nachtigallová, D.; Rulísek, L.; Pohl, R.; Hof, M.; Hocek, M. *Chem. Sci.* **2016**, *7*, 5775–5785.
6. Xie, Y.; Dix, A. V.; Tor, Y. *J. Am. Chem. Soc.* **2009**, *131*, 17605–17614.
7. (a) Pawar, M. G.; Srivatsan, S. G. *Org. Lett.* **2011**, *13*, 1114–1117. (b) Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J.* **2011**, *17*, 12820–12827. (c) Tanpure, A. A.; Srivatsan, S. G. *ChemBioChem* **2012**, *13*, 2392–2399. (d) Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G. *Bioconjugate Chem.* **2013**, *24*, 1367–1377. (e) Sabale, P. M.; Nuthanakanti, A.; Srivatsan, S. G. *Ind. J. Chem.* 2013, *52A*, 1004–1013. (f) Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149. (g) Sabale, P. M.; Tanpure, A. A.; Srivatsan, S. G. *Org. Biomol. Chem.* **2018**, *16*, 4141–4150. (h) Manna, S.; Sarkar, D.; Srivatsan, S. G. *J. Am. Chem. Soc.* **2018**, *140*, 12622–12633.
8. (a) Ono, A.; Torigoe, H.; Tanaka, Y.; Okamoto, I. *Chem. Soc. Rev.* **2011**, *40*, 5855–5866. (b) Tanaka, Y.; Kondo, J.; Sychrovský, V.; Šebera, J.; Dairaku, T.; Saneyoshi, H.; Urata, H.; Torigoeh, H.; Ono, A. *Chem. Commun.* **2015**, *51*, 17343–17360. (c) Johannsen, S.; Paulus, S.; Düpre, N.; Müller, J.; Sigel, R. K. O. *J. Inorg. Biochem.* **2008**, *102*, 1141–115.
9. Clarkson, T. W.; Magos, L. *Crit. Rev. Toxicol.* **2006**, *36*, 609–662.

10. (a) Urata, H.; Yamaguchi, E.; Funai, T.; Matsumura, Y.; Wada, S.-i. *Angew. Chem. Int. Ed.* **2010**, *49*, 6516–6519. (b) Funai, T.; Nakamura, J.; Miyazaki, Y.; Kiriu, R.; Nakagawa, O.; Wada, S.-i., Ono, A.; Urata, H. *Angew. Chem. Int. Ed.* **2014**, *53*, 6624–6627. (c) Park, K. S.; Jung, C.; Park, H. G. *Angew. Chem. Int. Ed.* **2010**, *49*, 9757–9760.

11. (a) Ono, A.; Togashi, H. *Angew. Chem. Int. Ed.* **2004**, *43*, 4300–4302. (b) Ma, D.-L.; Chan, D. S.-H.; Man, B. Y.-W.; Leung, C.-H. *Chem. -Asian J.* **2011**, *6*, 986–1003.

12. (a) Torigoe, H.; Kawahashi, K.; Takamori, A.; Ono, A. *Nucleosides, Nucleotides Nucleic Acids* **2005**, *24*, 915–917. (b) Lin, Y.-W.; Ho, H.-T.; Huang, C.-C.; Chang, H.-T. *Nucleic Acids Res.* **2008**, *36*, e123. (c) Torigoe, H.; Ono, A.; Kozasa, T. Transition *Met. Chem.* **2011**, *36*, 131–144.

13. (a) Wang, Y. S.; Cheng, C.-C.; Chen, J.-K.; Ko, F.-H.; Chang, F.-C. *J. Mater. Chem. A* **2013**, *1*, 7745–7750. (b) Dave, N.; Chan, M. Y.; Huang, P.-J. J.; Smith, B. D.; Liu, J. *J. Am. Chem. Soc.* **2010**, *132*, 12668–12673. (c) Huang, P.-J. J.; Liu, J. *Chem. – Eur. J.* **2011**, *17*, 5004–5010. (d) He, D.; He, X.; Wang, K.; Zhao, Y.; Zou, Z. *Langmuir* **2013**, *29*, 5896–5904.

14. (a) Carell, T.; Behrens, C.; Gierlich, J. *J. Org. Biomol. Chem.* **2003**, *1*, 2221–2228. (b) Isobe, H.; Yamazaki, N.; Asano, A.; Fujino, T.; Nakanishi, W.; Seki, S. *Chem. Lett.* **2011**, *40*, 318–319.

15. (a) Okamoto, I.; Iwamoto, K.; Watanabe, Y.; Miyake, Y.; Ono, A. *Angew. Chem. Int. Ed.* **2009**, *48*, 1648–1651. (b) Yamaguchi, H.; Šebera, J.; Kondo, J.; Oda, S.; Komuro, T.; Kawamura, T.; Dairaku, T.; Kondo, Y.; Okamoto, I.; Ono, A.; Burda, J. V.; Kojima, C.; Sychrovský, V.; Tanaka, Y. *Nucleic Acids Res.* **2014**, 4094–4099. (c) Miyake, Y.; Togashi, H.; Tashiro, M.; Yamaguchi, H.; Oda, S.; Kudo, M.; Tanaka, Y.; Kondo, Y.; Sawa, R.; Fujimoto, T.; Machinami, T.; Ono, A. *J. Am. Chem. Soc.* **2006**, *128*, 2172–2173.

16. (a) Kuklenyik, Z.; Marzilli, L. G. *Inorg. Chem.* **1996**, *35*, 5654–5662. (b) Tanaka, Y.; Oda, S.; Yamaguchi, H.; Kondo, Y.; Kojima, C.; Ono, A. *J. Am. Chem. Soc.* **2007**, *129*, 244–245. (c) Tanaka, Y.; Ono, A. *Dalton Trans.* **2008**, 4965–4974. (d) Tanaka, Y.; Yamaguchi, H.; Oda, S.; Nomura, M.; Kojima, C.; Kondo, Y.; Ono, A. *Nucleosides, Nucleotides Nucleic Acids* **2006**, *25*, 613–624.

17. (a)Torigoe, H.; Ono, A.; Kozasa, T. *Chem. Eur. J.* **2010**, *16*, 13218–13225. (b) Torigoe, H.; Miyakawa, Y.; Ono, A.; Kozasa, T. *Thermochim. Acta* **2012**, *532*, 28–35.

18. (a)Wang, Z.; Lee, J. H.; Lu, Y. *Chem. Commun.* **2008**, 6005–6007. (b) Teh, H. B; Wu, H.; Zuo, X.; Li, S. F. Y. *Sensors and Actuators B* **2014**, *195*, 623–629. (c) Liu, C.-W.; Huang, C.-C.; Chang, H.-T. *Anal. Chem.* **2009**, *81,* 2383–2387.

19. (a) Zhang, L.; Li, T.; Li, B.; Li, J.; Wang, E. *Chem. Commun.* **2010**, *46*, 1476–1478. (b) Li, H.; Zhai, J.; Tian, J.; Luo, Y.; Sun, X. *Biosensors and Bioelectronics* **2011**, *26*, 4656–4660. (c) Wang, H.; Wang, Y.; Jin, J.; Yang, R. *Anal. Chem.* **2008**, *80*, 9021–9028. (d) Zhang, J. R.; Huang, W. T.; Xie, W. Y.; Wen, T.; Luo, H. Q.; Li, N. B. *Analyst* **2012**, *137*, 3300–3305.

20. (a) Chiang, C.-K.; Huang, C.-C.; Liu, C-W.; Chang, H.-T. *Anal. Chem.* **2008**, *80*, 3716–3721. (b) Chan, D. S.-H.; Lee, H.-M.; Che, C.-M.; Leung, C.-H.; Ma, D.-L. *Chem. Commun.* **2009**, 7479–7481.(c) Zhang, X.; Li, Y.; Su, H.; Zhang, S. *Biosensors and Bioelectronics* **2010**, *25*, 1338–1343.

21. (a) Mata, G.; Schmidt, O. P.; Luedtke, N. W. *Chem. Commun.* **2016**, *52*, 4718–4721. (b) Schmidt, O. P.; Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2016**, *138*, 14733–14739. (c) Schmidt, O. P.; Benz, A. S.; Mata, G; Luedtke, N. W. *Nucleic Acids Res.* **2018**, *46*, 6470–6479.
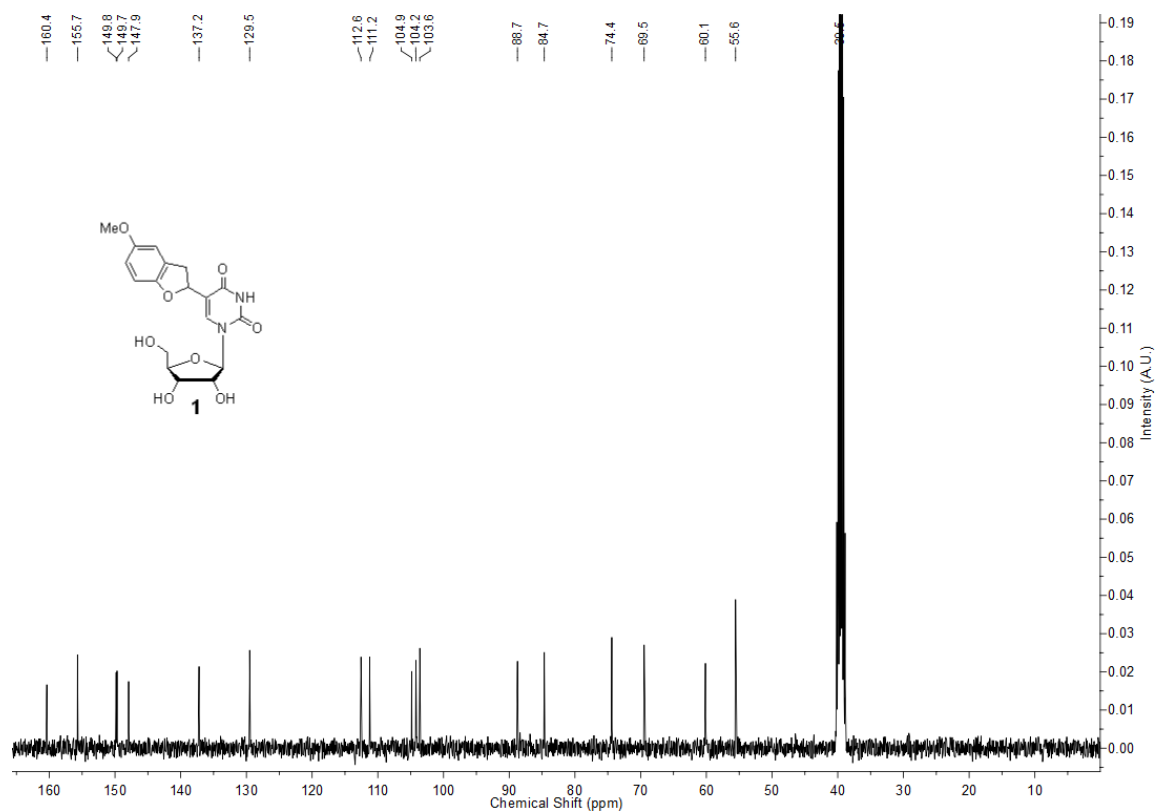
22. Bonini, C.; Cristiani, G.; Funicello, M.; Viggiani, L. *Synth. Commun.* **2006**, *36*, 1983–1990.
23. Reichardt, C. *Chem. Rev.* **1994**, *94,* 2319–2358.
24. Sinkeldam, R. W.; Wheat, A. J.; Boyaci, H.; Tor, Y. *ChemPhysChem* **2011**, *12*, 567–570.
25. Haidekkera, M. A.; Theodorakis, E. A. *Org. Biomol. Chem.* **2007**, *5*, 1669–1678.
26. (a) Kawai, M.; Lee, M. J.; Evans, V T.; Nordlund, M. *J. Fluoresc.* 2001, 11, 23–32; b) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 37–41. (c) C. A. M. Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. *J. Phys. Chem.* **1996**, *100*, 5541–5553
27. Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem.* **1994**, *5*, 508–512.
28. Lavabre, D.; Fery-Forgues, S. *J. Chem. Educ.* **1999**, *76*, 1260–1264.

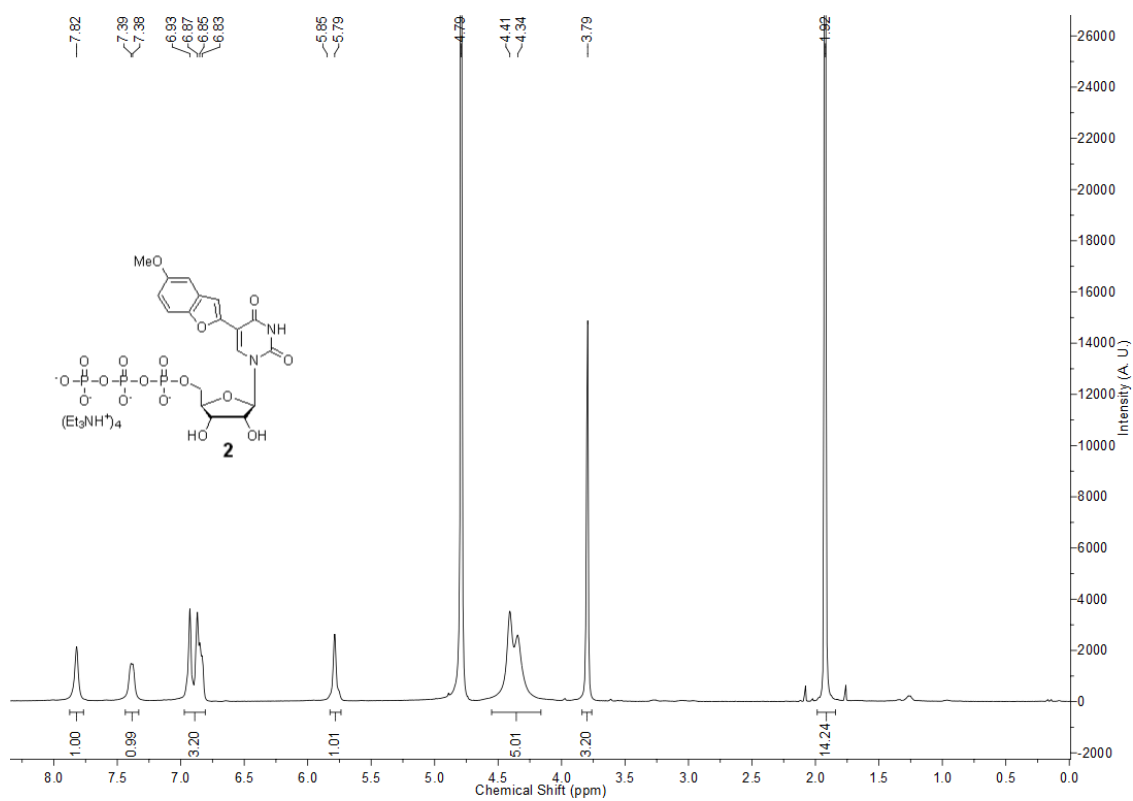## 5.6. Appendix-III: Characterization data of synthesized compounds
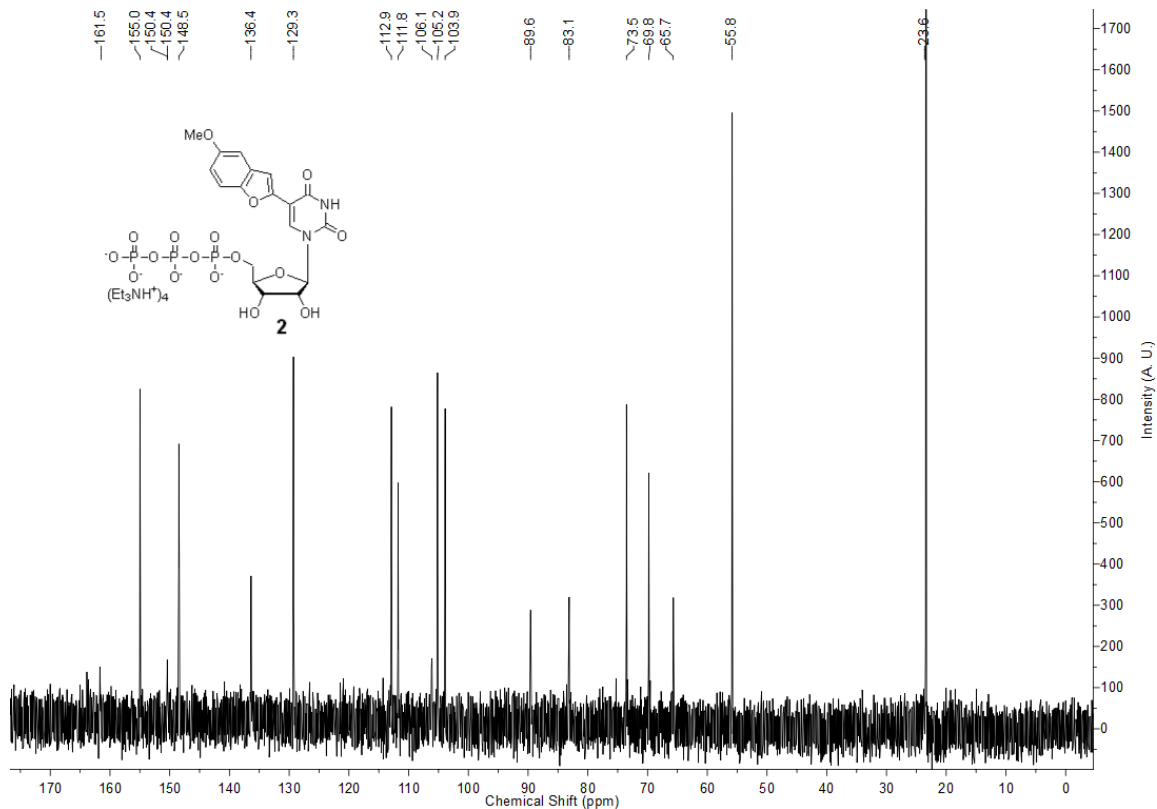
^{1}H NMR of **1** (400 MHz, d_6-DMSO)
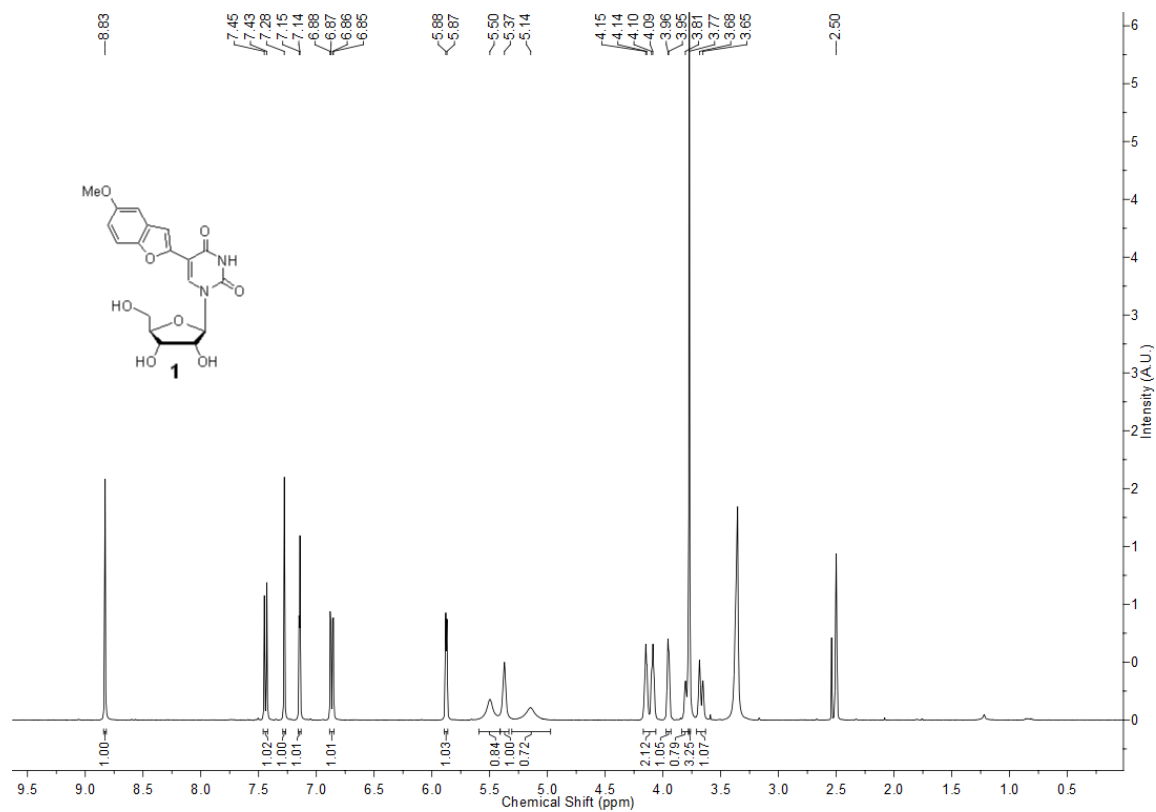
^{13}C NMR of **1** (100 MHz, d_6-DMSO)

^{1}H NMR of **2** (400 MHz, D$_2$O)
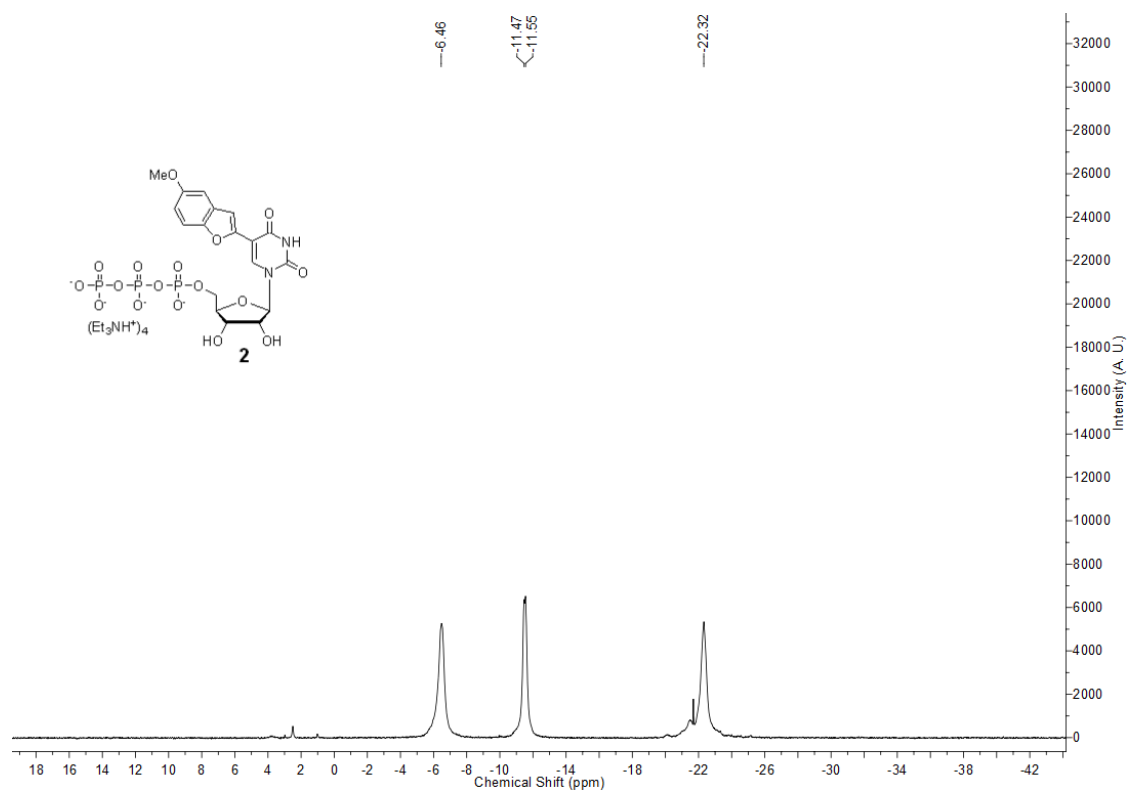

^{13}C NMR of **2** (100 MHz, D$_2$O)

181

^{31}P NMR of **2** (162 MHz, D$_2$O)

182

General Conclusions and Future Perspective

Using simple physical organic concept here in this thesis we have developed structurally non-invasive and environment-sensitive nucleoside probes to investigate nucleic acid structure and recognition. The benzofuran-modified fluorescent uridine analog and a cell-model, reverse micelle (RM) enabled us to understand how a confined environment can affect the G-quadruplex (GQ) topology adopted by human telomeric (H-Telo) DNA and its ligand binding affinities. Further, a novel dual-app probe, 5-fluorobenzofuran-modified uridine analog containing a fluorophore and an in-cell NMR compatible ^{19}F isotope label provided the opportunity to study H-Telo DNA in real time by fluorescence and in cell by NMR. The fluorescence component of the probe efficiently differentiated the binding affinity of different ligands to different GQ topologies. The in-cell ^{19}F NMR studies provided the information about GQ topology of H-Telo DNA in cell, which has remained unclear till date.

This probe system is general and can be utilized in investigating other individual GQ-forming sequences as well as other structural motifs in the genome. Currently, my colleagues are utilizing this first generation probe and also newly developed second generation multi-app probes to study structural motifs such as GQs and i-motif (iMs) of oncogenes and viruses. Since GQ and iM forming sequences often co-exist, the signatures obtained using these probes for individual structures will enable determination of the relative propensity of a promoter sequence to adopt duplex, GQ and iM.

In the last part of the thesis, I have demonstrated the development of a new fluorescent nucleoside probe by attaching methoxybenzofuran to uridine and utilization of the probe in investigation of metal-mediated base-pair formation in nucleic acid duplexes. The addition of an auxochrome (methoxy group), made the nucleoside probe highly micropolarity-sensitive. The responsiveness of the probe will be used in understanding protein-nucleic acid and small molecule-nucleic acid interactions.