Probing nucleic acid conformations and recognition by fluorescence and ^{19}F NMR using dual-purpose nucleoside analogs

A thesis submitted in partial fulfillment of the requirements
Of the degree of
Doctor of Philosophy

By

Saddam Husen Yusuf Khatik
ID. 20163460

Indian Institute of Science Education and Research, Pune
2023

This dissertation is dedicated

to my mother

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pashan, Pune–411 008

Prof. Seergazhi G. Srivatsan
Department of Chemistry

CERTIFICATE

Certified that the work incorporated in the thesis entitled "***Probing nucleic acid conformations and recognition by fluorescence and*** ^{19}***F NMR using dual-purpose nucleoside analogs***" submitted by **Mr. Saddam Husen Yusuf Khatik** 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.

S. G. Srivatsan

Date: 01/05/2023
Place: Pune

Signature of Supervisor
Prof. Seergazhi G. Srivatsan

DECLARATION

I declare that this written submission represents my ideas in my own words and where 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 be 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.

Saddam Husen Yusuf Khatik
ID: 20163460

Date: 00.05.2023
Place: Pune

Acknowledgments

The journey of obtaining my PhD degree was filled with many hurdles, but that help me to grow personally, intellectually and professionally. I really enjoyed this journey and learn a lot and it was all possible due to the support and encouragement from numerous people. At the end of this journey, I would like to acknowledge all those people for their direct or indirect contributions, which made this thesis possible. Foremost, I deeply express my gratitude towards my PhD supervisor Prof. S. G. Srivatsan for his valuable suggestions, continuous support and encouragement throughout my PhD degree. He has a pivotal role in my development as a PhD scholar. Before joining the PhD program, I was an organic chemist and had a phobia of biology but his immense knowledge, his awesome way of transferring scientific knowledge and his motivational words devoted me to the understanding of nucleic acid chemistry and their biology. Because of that, I am now enough confident and comfortable to work at the interface of chemistry and biology. I am very grateful for having him as my PhD mentor.

I would like to extend my gratitude towards my Research Advisory Committee (RAC) members, Dr. Jeetender Chugh and Dr. Siddhesh S. Kamat for their valuable comments and suggestions during RAC meetings. I would like to acknowledge Dr. P. I. Pradeepkumar and his student Ms. Sruthi Sudhakar for providing the computational data for one of my research project. Also, I would like to thank Dr. Jeet Kalia and his student Mr. Satyajeet Mishra for providing the Frog oocytes for my in-cell NMR experiments.

I am very grateful for getting an opportunity to pursue my PhD degree in IISER Pune. Most importantly, I am thankful to IISER Pune and India Alliance for providing me research fellowship. I want to express my sincere gratitude to the former Director of IISER Pune, Prof. K. N. Ganesh for establishing this wonderful institute with world-class research facilities. I would like to acknowledge our current director, Prof. Sunil Bhagwat and chair chemistry, Prof Nirmalya Ballav for all facilities. I like to acknowledge all faculty members and research fellows of chemistry department for their support during my PhD. Also, I acknowledge all non-teaching and technical staff especially Sandeep Mishra, Nitin, Sandeep Kanade, Mahesh, Tushar, Mayuresh and Sneha for their help.

I am thankful to all my former labmates Dr. Maroti, Dr. Anupam, Dr. Arun, Dr. Pramod, Dr. Ashok, Dr. Sudeshna, Dr. Jerrin, Dr. Manisha, Akanksha, Uddhav, Dr. Vyankat and Dr. Cornelia for setting a wonderful lab culture and present labmates Pulak, Sarupa, Swagata, Amit, Arindam, Harshkumar and Dr. Apeksha for their direct or indirect support and in maintaining cheerful lab atmosphere. Particularly, I thank Dr. Pramod, Dr. Sudeshna, Dr. Jerrin and Dr. Vyankat for their support in various experiments. Also, I am thankful to all my IISER Pune colleagues and NCL friends for their support during my PhD, whose names I might have missed unintentionally.

Finally and most importantly special thanks to my mother, my wife and my entire family for their unconditional love, support and their sacrifices, which made this difficult journey enjoyable.

Saddam Husen Yusuf Khatik

Chapter 2 is a reprint of the data published in the journal: **Khatik, S. Y.**; Sudhakar, S.; Mishra, S.; Kalia, J.; Pradeepkumar, P. I.; Srivatsan, S. G. Probing juxtaposed G-quadruplex and hairpin motifs using a responsive nucleoside probe: a unique scaffold for chemotherapy. *Chem. Sci*. **2023**, Just Accepted, doi.org/10.1039/D3SC00519D.

Also, the synthesis and photophysical properties of the modified nucleoside are a part of data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. Environment-Sensitive Nucleoside Probe Unravels the Complex Structural Dynamics of i-Motif DNAs. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.

The thesis author is a main author and researcher for this work.

Chapter 3 is a reprint of part of the data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. Environment-Sensitive Nucleoside Probe Unravels the Complex Structural Dynamics of i-Motif DNAs. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.

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

Chapter 4 is a reprint of part of the data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. Environment-Sensitive Nucleoside Probe Unravels the Complex Structural Dynamics of i-Motif DNAs. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.

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

Table of Content

Chapter 1: Biophysical tools to probe nucleic acid structure, dynamics and function

Chapter 2: Probing juxtaposed G-quadruplex and hairpin motifs using a responsive nucleoside probe: a unique scaffold for chemotherapy

Chapter 3: Nucleoside probe reports the formation of human telomeric i-motif structures

Chapter 4: 3FBF-dU probe unravels the complex structural dynamics of Braf i-motif DNAs

Chapter 5: Detection of aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA using 3FBF-U

Abbreviations

λ	Wavelength
μM	Micromolar
μL	Microliter
ε	Molar extinction coefficient
2-AP	2-Aminopurine
3-MI	3-Methyl Isoxanthopterin
6-MI	6-Methyl Isoxanthopterin
Abs	Absorbance
ACN	Acetonitrile
Ag	Silver
ATP	Adenosine Triphosphate
C	Cytosine
CD	Circular Dichroism
COSY	Correlated Spectroscopy
CTP	Cytidine Triphosphate
Cy	Cysteine
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
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ESI-MS	Electrospray Ionisation Mass Spectroscopy
em	Emission
ex	Excitation
EPR	Electron Paramagnetic Resonance
FAM	Fluorescein Amidites
FI	Fluorescence Intensity
FRET	Fluorescence Resonance Energy Transfer
G	Guanosine
GQ	G-quadruplex
GTP	Guanosine Triphosphate
HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
H-Telo	Human Telomeric
iM	Intercalated-Motif
in vitro	Outside living organism
in vivo	Inside living organism
Kd	Dissociation Constant
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time of flight
max	Maximum
MeOH	Methanol
$(MeO)_3PO$	Trimethyl Phosphate

MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligram
MHz	Megahertz
mM	Milimolar
nm	Nanometer
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NTP	Nucleoside Triphosphate
ON	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PELDOR	Pulsed Electron Double Resonance
Pd	Palladium
PDS	Pyridostatin
$POCl_3$	Phosphorus Oxychloride
ppm	Parts per million
RC	Random Coil
RNA	Ribonucleic acid
RNase	Ribonuclease
Rf	Retention Factor
ss	Single-stranded
T	Thymine
TCA	Trichloroacetic acid
tC	1,3-diaza-2-oxophenothiazine-2′-deoxycytidine
tC_{nitro}	7-nitro-1,3-diaza-2-oxophenothiazine-2′-deoxycytidine
tC^{o}	1,3-diaza-2-oxophenoxazine-2′-deoxycytidine
TCSPC	Time-Correlated Single Photon Counting
TDS	Thermal difference spectrum
TEAA	Triethylammonium Acetate
TERRA	Telomeric Repeat-Containing RNA
TFT	Trifluorotoluene
THF	Tetrahydrofuran
TLC	Thin layer chromatography
Tm	Thermal melting
TMEDA	N,N,N',N' Tetramethylethylenediamine
TO	Thiazole Orange
TOCSY	Total Correlation Spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
U	Uridine/Uracil
UTP	Uridine Triphosphate
UV	Ultraviolet

Synopsis

Probing nucleic acid conformations and recognition by fluorescence and ^{19}F NMR using dual-purpose nucleoside analogs

Background and Aim: In addition to a double-stranded structure, nucleic acids are known to adopt various alternative structures (bulges, hairpins, branched junctions, triplex and tetraplex, etc.) and they are directly associated with the biological functions of nucleic acids.[1] Understanding the nucleic acid structure-function relationship has great importance in structural biology and can be implemented in the development of nucleic acid therapeutics. In this context, tremendous efforts have been taken to understand the function of nucleic acids by determining their structures, dynamics and how they interact with proteins using different techniques (NMR, EPR, X-ray crystallography, fluorescence, etc).[2] However, to employ these techniques, nucleic acids are needed to be labeled with an appropriate tag, as they don't contain any intrinsic label. In addition to structural understanding, these techniques are useful to identify small molecule binders specific to different structures as potential therapeutic agents. However, the majority of studies provide nucleic acid structural information in *in vitro* conditions. As structure and dynamics are very sensitive to the surrounding environment like ionic conditions, molecular crowding, confinement and pH, it is not necessary that a nucleic acid sequence would adopt a structure in cell similar to the one observed in *in vitro* conditions. Recently, some structure-specific antibodies and fluorescence light-up probes have been developed to visualize the formation of nucleic acid structures in cellular conditions.[3] Also, in-cell NMR and EPR techniques are very useful to detect the preferred nucleic acid conformations, their dynamics and interactions with proteins in complex cellular environments.[4] However, these techniques are not so useful to distinguish certain conformations (different tetraplex structures) when they exist in a complex equilibrium. In this regard, we anticipated that the development of a minimally perturbing nucleoside probe, which could sense its microenvironment and produce unique signatures for different topologies both *in vitro* and in cellular models, would be highly useful in gaining insights into the structure and recognition properties of nucleic acid motifs.

In this thesis, we have designed and synthesized dual-labeled nucleoside analogs that contain fluorescence and ^{19}F NMR labels. The analogs were employed to examine nucleic acid conformations, dynamics and interactions with ligands. Modified nucleosides were synthesized by attaching 5-trifluoromethylbenzofuran heterocyclic ring at the 5 positions of 2´

deoxyuridine or uridine analogs. These analogs are highly sensitive to changes in solvent polarity and viscosity and are structurally non-invasive when incorporated into DNA and RNA ONs. Interestingly, the modified nucleoside produced a definite fluorescence and ^{19}F NMR signatures for the individual G-quadruplex (GQ) structures formed by the EGFR G-rich DNA sequence. Importantly, ^{19}F NMR signatures were useful to detect a physiologically relevant EGFR GQ structure under cellular conditions. Further, the nucleoside analog was used to monitor the formation of different human telomeric i-motif (iM) structures and their dynamics under different conditions. Rewardingly, the nucleoside analog provided a valuable information on the complex iM structural polymorphism form in the Braf promoter region and could even detect lowly populated iM conformations. Similarly, the modified ribonucleoside was employed to observe aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA. Collectively, these dual-labeled nucleoside analogs provide a simplified solution for the investigation of complex nucleic acid structural equilibrium and could help to identify topology-specific binders.

The thesis is organized as detailed below.

Chapter 1: Biophysical tools to probe nucleic acid structure, dynamics and function

The formation of non-canonical nucleic acid structures in particular the tetraplex structures (GQ and iM) and their biological functions have discussed in this chapter. Their structure-function relationship has been studied using various biophysical techniques namely CD, UV-absorption, NMR, EPR, X-ray crystallography and fluorescence.[2] The working principle of these techniques, current progress and available challenges in studying tetraplex structures and their interactions with proteins and ligands are briefly explained. Additionally, the usefulness of structure specific-antibodies, fluorescence light-up probes, in-cell NMR and in-cell EPR techniques to investigate the different GQ/iM structures in cellular conditions are discussed. Further, the development and utility of micro-environment-sensitive fluorescent nucleoside analogs to probe the different nucleic acid structures *in vitro* and *in vivo* are reviewed in detail. Their limitations in probing nucleic acid structures and motivation for the development of our dual-app nucleoside analogs are elaborated in the last part of the chapter.

Chapter 2: Probing juxtaposed G-quadruplex and hairpin motifs using a responsive nucleoside probe: a unique scaffold for chemotherapy

Guanine (G) rich DNA and RNA sequences are known to fold into G-quadruplex (GQ) structures.[1c] Bioinformatics study indicated that potential GQ-forming sequences are widespread in the human genome, mainly found in the telomere region of the chromosome, promoter regions of several proto-oncogenes, and the untranslated region of mRNA.[5] An ample amount of data demonstrated that the GQ is an important structural element involved in maintaining the stability of the genome and regulating the expression of an oncogene.[6] Hence, many GQ stabilizing ligands have been developed to downregulate gene expression and are considered antitumor agents.[7] However, the majority of ligands poorly distinguish between different topologies of GQ due to similarity in their structural skeleton. Alternatively, a non-canonical GQ structure that contains an additional structural element in the loop like GQ-duplex is considered an attractive target.[8] Since specific GQ targeting can be achieved by developing a ligand that could simultaneously target the GQ core and its proximal duplex motif.[9] In this chapter, we have studied the G-rich sequence from the promoter region of the epidermal growth factor receptor (EGFR) gene, which is known to fold into a mixture of two unique GQ structures (parallel and hybrid GQ) containing an additional hairpin structure in their third loop. Hence, it is necessary to determine their structural equilibrium under *in vitro* and *in vivo* conditions. Several biophysical tools have been developed to investigate GQ structures and their dynamics, but it remains a major challenge to differentiate different topologies *in vivo* due to the complex GQ structural equilibrium.

Here, we have designed and synthesized a dual-label micro-environment responsive nucleoside analog namely, 5-trifluoromethylbenzofuran-modified deoxyuridine (dU*) to investigate the GQ structural equilibrium adopted by the EGFR promoter region. Wherein, fluorescence and ^{19}F NMR properties of the modified nucleoside were sensitive to changes in the solvent polarity and solvent. Since the incorporation of modified nucleoside in the EGFR G-rich sequence could experience a different atmosphere in the parallel GQ and hybrid GQ topology, it displays distinct fluorescence and ^{19}F NMR spectral pattern for the individual topologies. These spectral properties were helpful to determine the relative populations of parallel GQ and hybrid GQ structures under different conditions. When a probe was placed in the mutated EGFR G-rich sequence, it elucidated the role of a hairpin domain in EGFR GQs equilibrium and reveal that it is more important for the formation of parallel GQ topology. Remarkably, the ^{19}F NMR signature of modified EGFR ON exhibits that parallel GQ is a predominant component in intracellular ionic conditions. However, in cellular conditions (in frog egg lysate and extract), it was observed that EGFR ON folds into a single hybrid GQ topology. Additionally, dU* is a valuable tool to identify GQ-binding ligands and determine how they interact with coexisting GQ topologies.

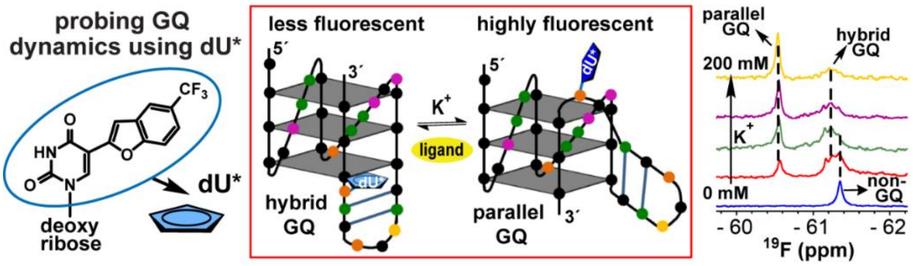


Figure 1. Probing the EGFR GQ structural equilibrium and their ligand interactions using dU* analog.

Chapter 3: Nucleoside probe reports the formation of human telomeric i-motif structures

Much like a GQ structure, iM is another non-canonical structure formed by a cytosine (C) rich sequence which is complementary to the GQ forming sequence.[10] Since the formation of iM structure requires slightly acidic conditions, their existence in cellular environments was less anticipated. However, recent studies using the in-cell 1D NMR analysis and the structure-specific antibody (iMab) provide evidence for the formation of iM structures in cellular conditions.[10] Further, the discovery of proteins and ligands that specifically interact with the iM structure and could modulate gene expression highlighted their biological importance. Hence, it is essential to develop a biophysical tool that could provide an understanding of the formation of iM structures and their dynamics under different conditions.

In this chapter, we have utilized a dU* analog to monitor the formation H-Telo iM structures and their dynamics under different conditions (pH and temperature). The H-Telo iM is known to fold into a mixture of 3´E (minor) and 5´E (major) iM topology and their equilibrium is very sensitive to pH and temperature conditions. Here, the fluorescence property of modified nucleoside was useful to monitor the iM folding/unfolding process upon changing the pH. Importantly, the modified nucleoside displayed distinct ^{19}F NMR peaks for the unfolded state, 3´E iM and 5´E iM conformations. These signatures were effective to determine their relative population with respect to pH and temperatures. Our study highlighted that the unfolded structure is a major component at higher pH (7.5), but at lower pH (5), 5´E iM conformation becomes a predominant structure. However, the study suggested that the H-Telo C-rich sequence could fold into an iM structure until pH 6.2 only and above that it remains in the unfolded state.

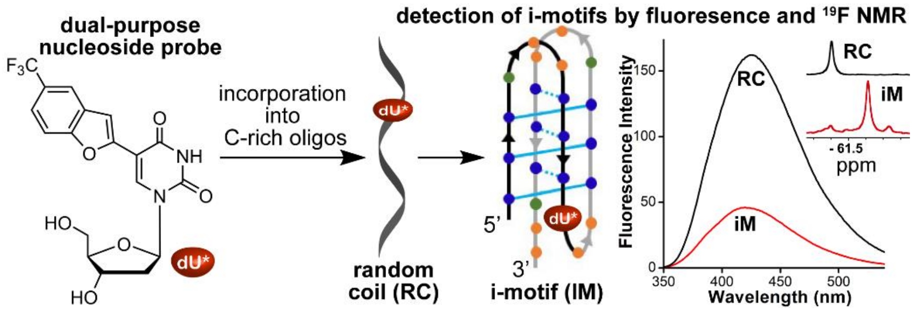


Figure 2. Modified nucleoside analog reveals the formation of different H-Telo iM structures.

Chapter 4: TFBF-dU probe unravels the complex structural dynamics of Braf i-motif DNAs

The stability of the iM structure depends on various factors such as molecular crowding, negative superhelicity and certain modifications, which could support the formation of iM structures at neutral pH.[10] Recent studies highlighted that the stability of iM structures majorly relies on the sequence content, particularly on the length of C-stretches and intervening loop residues.[11] It increases with an increase in the length of C-stretches and ON with C-stretches of 5 to 6 cytosines in length could even form an iM structure near neutral/physiological conditions. However, the stability of iMs doesn't further shift for C-stretches with more than 6 cytosines in length. Further, the stability of iM structures reduces with an increase in the length of loop residues but the formation of the secondary structure in the loop (hairpin) increases iM stability irrespective of its length.[11b] Several C-rich sequences from the promoter region of the oncogenes contain longer C-stretches of varying length, thus they could fold into iM structures near physiological or cellular conditions. Although, these sequences could fold into more than one iM structure using the different combinations of C-tract and loop residues. It has been found that the iM displays a high degree of structural polymorphism and it is very sensitive to surrounding conditions due to the minute energy differences. Formation of the iM structure highly depends on the extent of protonation. At low pH, thermodynamically stable conformation would be preferred, but at high pH/neutral pH, iM structure with less number of tetrad would form due to low protonation.[11b] Hence, it is imperative to determine what structure longer C-rich sequences adopt at near neutral/physiological conditions.

In this context, we have used the dU* nucleoside to investigate the iM structural polymorphism exhibited by the longer C-rich sequence from the promoter region of the Braf gene. Incorporation of the modified nucleoside into Braf ON fluorescently reports the formation of iM structure near physiological pH (~7.2). Interestingly, ^{19}F NMR analysis revealed that the Braf sequence folds into multiple iM structures and their structural equilibrium is highly sensitive to changes in pH and temperature. At lower pH, it adopts one major iM structure along with multiple minor structures, wherein the population of minor iM structures significantly improved with decrease in the major iM structure upon increasing pH. At higher pH, the minor iM structures are preferred due to the less degree of protonation. This result indicates that the thermodynamically stable iM structure that formed at lower pH need not exist at higher pH. Further, the incorporation of the modified analog in mutated Braf ONs was useful to identify cytosine residues that are involved in the formation of major iM structure and also to determine the loop size of the major iM structure.

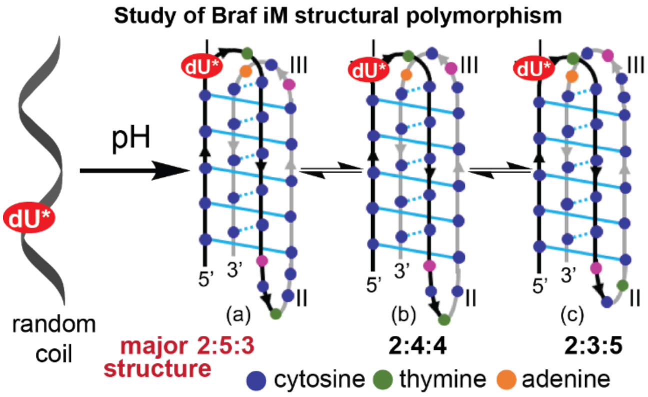


Figure 3. Investigation of the Braf iM structural polymorphism using the modified analog.

Chapter 5: Detection of aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA using 3FBF-U

RNA is known to fold into secondary, tertiary and quaternary structures and undergoes conformational transitions to perform certain functions.[12] In this chapter, we have explored the utility of dual-channel ribonucleoside probe to investigate RNA conformations. We have synthesized 5-trifluoromethyl-benzofuran-modified uridine (U*) and its triphosphate (U*TP). Initial photophysical and ^{19}F NMR studies of U* in different solvents validated its micro-environment-sensing ability similar to dU* analog. Further, the modified nucleotide (U*TP) serves as a good substrate for *in vitro* transcription reaction using T7 RNA polymerase and incorporated into RNA with moderate to good efficiency. Interestingly, the photophysical properties and ^{19}F NMR signal of U* analog in short RNA were able to distinguish single-stranded RNA from its duplex and also changes in its flanking bases. Inspired by these results, we incorporated modified nucleotide into one of the therapeutically important RNA which is the bacterial ribosomal decoding site (A-site) to monitor RNA-ligand interactions. In protein synthesis, the selection of an accurate tRNA is a crucial step to get the correct protein.[13] The crystal structure analysis suggested that the A-site which is part of 16sRNA contact with cognate codon-anticodon pairing of mRNA and tRNA and undergoes a conformational transition which gives the signal for protein synthesis.[14] Particularly, A_{1492} and A_{1493} adenine residues part of the small bulge of A-site are flexible and experience conformational transitions during cognate codon-anticodon interaction. Interestingly, naturally occurring aminoglycosides bind with the bulge part of A-site and fix the conformations of adenine residues (A_{1492}, A_{1493}) similar to those formed upon cognate codon-anticodon interactions. This misleads the process of tRNA selection and hence hampers protein synthesis. It was found that the aminoglycosides interact with A_{1408}, A_{1492} and A_{1493} residues through direct H-bonding and a noncanonical $U_{1406}{\circ}U_{1495}$ pair through water-mediated H-bonding.[14] Therefore, we incorporated U* at the U_{1406} position in the short A-site RNA and used to detect the aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site. Altogether, our dual-labeled micro-environment sensitive probe could be utilized to monitor the RNA structure and dynamics and also, could assist to develop a small-molecule screening platform.

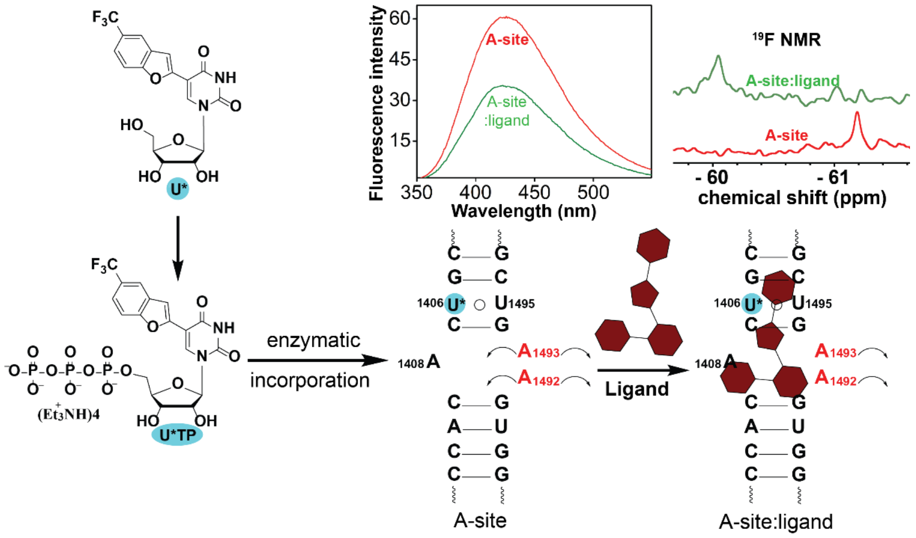


Figure 4. Incorporation of the modified nucleotide (U*TP) into the A-site RNA using an enzymatic reaction reports aminoglycoside antibiotics-induced conformational changes in the A-site by fluorescence and ^{19}F NMR.

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) Choi, J.; Majima, T. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909. (d) Tateishi-Karimata, H.; Sugimoto, N. *Nucleic Acids Res*. **2021**, *49*, 7839–7855.
2. (a) Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. *Nucleic Acids Res*. **2018**, *46*, 9264–9275. (b) Lacroix, L.; Séosse, A.; Mergny, J.-L. *Nucleic Acids Res*. **2011**, *39*, e21–e31. (c) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc*. **2006**, *128*, 9963–9970. (d) Parkinson, G. N.; Lee, M. P. H.; Neidle, S.

Nature, **2002**, *417*, 876–880. (e) Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.

3. (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182−186. (b) 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. (c) Stollar, B. D.; Raso, V. *Nature* **1974**, *250*, 231–234. (d) Riesen, A. J. V.; Le, J.; Slavkovic, S.; Churcher, Z. R.; Shoara, A. A.; Johnson P. E.; Manderville, R. A. *ACS Appl. Bio. Mater.* **2021**, 4, 6732–6741.
4. (a) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. *Nucleic Acids Res*. **2011**, *39*, 5768–5775. (b) Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantírek, L.; Dötsch, V. *Angew. Chem. Int. Ed*. **2014**, *53*, 10300–10314.
5. Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K. *Mol. Cancer* **2021**, *20*, 40–57.
6. (a) Sun, Z.-Y.; Wang, X.-N.; Cheng, S.-Q.; Su, X.-X.; Ou, T.-M. *Molecules* **2019**, *24*, 396–424. (b) Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. *Pharmaceuticals* **2021**, *14*, 769–808.
7. Ma, Y.; Iida, K.; Nagasawa, K. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 3–17.
8. (a) Lim, K. W.; Jenjaroenpun, P.; Low, Z. J.; Khong, Z. J.; Ng, Y. S.; Kuznetsov, V. A.; Phan, A. T. *Nucleic Acids Res*. **2015**, *43*, 5630–5646. (b) Vianney, Y. M.; Weisz, K. *Nucleic Acids Res*. **2022**, *50*, 11948–11964.
9. (a) Asamitsu, S.; Obata, S.; Phan, A. T.; Hashiya, K.; Bando, T.; Sugiyama, H. *Chem. Eur. J.* **2018**, *24*, 4428–4435. (b) Yang, M.; Carter, S.; Parmar, S.; Bume, D. D.; Calabrese, D. R.; Liang, X.; Yazdani, K.; Xu, M.; Liu, Z.; Thiele, C. J.; Schneekloth Jr, J. S. *Nucleic Acids Res*. **2021**, *49*, 7856–7869.
10. (a) Day, H. A.; Pavlou, P.; Waller, Z. A. E. *Bioorg. Med. Chem.* **2014**, *22*, 4407−4418. (b) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2018**, *46*, 8038−8056.
11. (a) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res*. **2017**, *45*, 2951−2959. (b) Školáková, P.; Renčiuk, D.; Palackỷ, J.; Krafčík, D.; Dvořáková, Z.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. *Nucleic Acids Res.* **2019**, *47*, 2177−2189.
12. Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. *Nat. Rev. Mol. Cell Biol*. **2019**, *20*, 475−489.
13. Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *TRENDS Biochem. Sci.* **2003**, *28*, 259−266.
14. Matt, T.; Akbergenov, R.; Shcherbakov, D.; Böttger, E. C. *Isr. J. Chem*. **2010**, *50*, 60–70.

List of Publications

1. **Khatik, S. Y.**; Sudhakar, S.; Mishra, S.; Kalia, J.; Pradeepkumar, P. I.; Srivatsan, S. G. Probing juxtaposed G-quadruplex and hairpin motifs using a responsive nucleoside probe: a unique scaffold for chemotherapy. *Chem. Sci*. **2023**, Just Accepted, doi.org/10.1039/D3SC00519D.

2. **Khatik, S. Y.**; Srivatsan, S. G. Environment-Sensitive Nucleoside Probe Unravels the Complex Structural Dynamics of i-Motif DNAs. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.

3. Nuthanakanti, A.; Ahmed, I.; **Khatik, S. Y.**; Saikrishnan, K.; Srivatsan, S. G. Probing G-quadruplex topologies and recognition concurrently in real time and 3D using a dual-app nucleoside probe. *Nucleic Acids Res*. **2019**, *47*, 6059–6072.

4. **Khatik, S. Y.**; Srivatsan, S. G. Detection of aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA using 3FBF-U. (manuscript under preparation)

Chapter 1:

Biophysical tools to probe nucleic acid structure, dynamics and function

1.1 Introduction

Nucleic acid act as a carrier of genetic information and encloses a blueprint for the biosynthesis of RNA and proteins in living things. Additionally, the ability of nucleic acids (DNA or RNA) to adopt secondary and tertiary structures enables them to serve as catalysts and control cellular processes such as replication, transcription and translation.[1] Sequencing data revealed that more than 50% of the total human genomic DNA are accounted for repetitive DNA sequences.[2] Apart from conventional double helix structures, these repeat DNA sequences could fold into different structures like hairpin, multi-helix junctions, triplex, left-handed Z-form, tetraplex, etc. using canonical or non-canonical base pairing under particular conditions.[1c,3] These motifs undergo conformational transitions depending on the sequence, and interactions with proteins/metabolites or metal ions to perform various functions.[3] Importantly, some of the non-canonical structures are particularly found in or near the promoter region of the oncogenes and are known to regulate their expressions.[5] Alteration in their structure either due to changes in specific environmental conditions or mutations could lead to a disease state.[6]

Among these structures hairpins, multi-helix junctions, triplex and Z-form are discovered long before but still, their biological functions are not fully understood.[3] Particularly, RNA adopts hairpin, and multi-helix junctions structures and they are important for the formation of tertiary and quaternary structures.[4b,7] Also, RNA triplex structures are crucial in cellular nucleoprotein complexes to perform catalytic and regulatory functions. In the case of Z-DNA structure, it involves in gene expression, recombination and regulation processes.[8] Recently, several Z-form DNA-specific proteins have been identified, these proteins specifically bind to Z-DNA and regulate transcription or gene inhibition processes.[9] Recent studies reveal that the formation of these structures could associate with different disease states. However, information related to the formation of these structures under cellular conditions and their exact biological function is limited.

Another important class of structure is tetraplex structures namely G-quadruplex (GQ) and i-Motif (iM).[1c] These structures received much attention because of their existence in the functional regions of the genome and imperative biological functions.[5] The bioinformatics analysis suggests that potential GQ and iM forming sequences are not randomly distributed in the genome, but are particularly located in the human telomeric region, the promoter region of oncogenes.[5,10] Also, almost 40% of human promoter region contains a minimum of one GQ motif indicating that the formation of GQ could play a crucial role in gene expression.[5c,10b] Immunoglobulin-based assay highlighted that the relative population of the GQ and iM structures changes with cell cycle,[11,12] where GQ formation is associated with the replication phase and iM formation is detected in the transcription phase. Further, biochemical experiments underlined the involvement of these structures in various cellular processes such as replication, transcription and translation.[5,13] To perform various functions nucleic acid undergoes a conformational transition, which depends on the sequence, ionic condition, molecular crowding, pH and interactions with proteins. Also, under cellular conditions, proteins show different specificity and processivity against various structures.[14] Hence, it is very important to distinguish and harness the information of different structures to know their structure-function relationship under *in vitro* and cellular conditions.

To date, several biophysical techniques namely circular dichroism (CD), UV-vis absorption, fluorescence, nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR) and X-ray crystallography techniques have been developed to probe nucleic acid structures.[15] While, the native nucleic acid doesn't contain intrinsic label compatible with fluorescence, NMR, EPR and X-ray techniques, because of that several responsive nucleoside labels have been developed to investigate nucleic acid structure, dynamics and its interaction with proteins and ligands.[15b–e] Although these techniques provide valuable information on the nucleic acid structure-function relationship under *in vitro* conditions, still it remains a major challenge to gather these pieces of information under *in vivo* conditions due to the complex cellular environment. As nucleic acid structures are very sensitive to changes in their surrounding conditions, it is hard to say whether the structure formed in *in vitro* conditions would remain intact under the cellular milieu. Recently, the development of structure-specific antibodies and fluorescent light-up probes has greatly enhanced our understanding of the formation of particular nucleic acid structures and their location in complex cellular conditions.[11,12,16] Further, in-cell NMR and EPR techniques were found to be very useful to investigate the nucleic acid structure, dynamics and their interactions with proteins and ligands.[17] However, these techniques have some limitations that make them difficult to employ in the study of certain structures under cellular conditions.

In this chapter, the formation of non-canonical tetraplex nucleic acid structures i.e. GQ and iM, their biological functions and available biophysical techniques to study these structures are discussed. Extensive research has been done in the field of tetraplex structures, due to the availability of these structures in the regulatory region of the genome. Various available techniques, in particular, base-modified labels to investigate the non-canonical structure, their stability, dynamics and interactions with proteins and ligands are discussed in detail. Also, current progress in the development of structure-specific antibodies, ligands and other available techniques to detect tetraplex structures in a cellular environment is explained. Further, the key challenges in probing and targeting tetraplex structures and motivation for the present work are explained in the last part of this chapter.

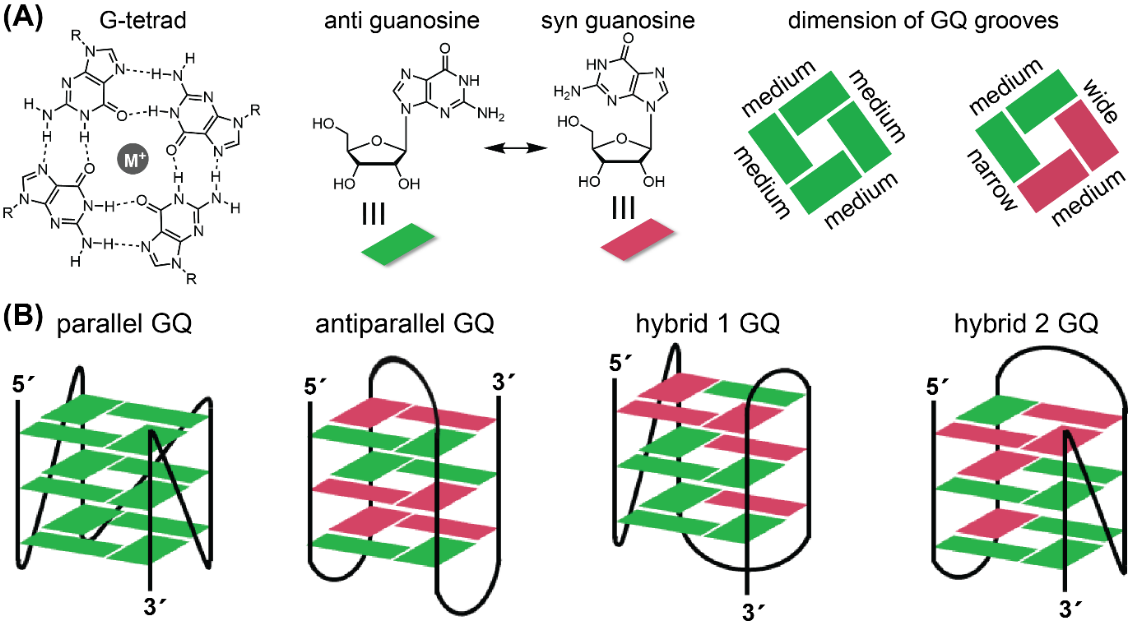


Figure 1. The structure of basic components of G-quadruplex (**A**) planer guanine tetrad, *anti* and *syn* glycosidic conformations of the guanosine are represented in green and ruby color, respectively, and the dimension of GQ grooves depends on the conformation of guanosine residues involved in G-tetrad formations. (**B**) Loop orientations and conformation of guanosine residues in different GQ topologies, such as parallel, antiparallel, hybrid 1 and hybrid 2 GQ.

1.2 G-quadruplex

G-quadruplex is a non-canonical structural motif formed by guanine (G)-rich DNA/RNA sequences. GQ composed of two or more G-tetrads that formed from the association of four guanosines through Hoogsteen hydrogen bonding are stacked one above the other in the presence of metal cations (K^+, Na^+, etc., Figure 1).[1c] Computational and sequencing analysis demonstrated that around 700,000 potential GQ-forming sequences (PQS) exist in the human genome.[5c] These GQ forming sequences contain at least four G-tract with a minimum length of three guanosines in each. They are particularly enriched in the telomere, the promoter regions of the oncogene, and the untranslated region of the RNA. GQ is a right-handed helical structure with four grooves and could fold into various topologies such as parallel stranded, antiparallel stranded and hybrid-type with mixed parallel-antiparallel stranded GQs (Figure 1).[1c,18a] Depending on the strand orientations, (i) GQ adopts three different loops namely lateral loop, diagonal loop and propeller loop, (ii) guanosine residues of the G-tetrad adopt *anti* glycosidic conformations in the parallel orientation, but they embrace a mixture of *syn* and *anti* glycosidic conformations in the antiparallel orientation,[18] and (iii) parallel orientation generates grooves with equal width (medium size), conversely antiparallel or mixed orientation creates both narrow and wide grooves.[1c,18a] Basically, the stability of GQ is defined by the number of G-tetrads that exist in the structure. Moreover, the stability and final topology of GQs depend on the number and nature of loop residues.

1.2.1 DNA G-quadruplex

Human telomeric (H-Telo) overhang is 50 to 200 nucleotides in length and composed of $(TTAGGG)_n$ repeating units. This G-rich sequence folds into the GQ structure and protects the chromosomal end from exonuclease degradation and end-to-end fusion.[19a] The H-Telo sequence containing four repeating units is the most studied GQ-forming sequence and could adopt different topologies depending on the salt (Na^+, K^+) and molecular crowding conditions. In the presence of Na^+ ion, it adopts a basket type of antiparallel GQ structure,[19b] but it folds into a parallel GQ structure under crystallization conditions containing K^+ ion and dehydrating molecular crowding agent.[19c] However, it exhibits structural polymorphism under physiological conditions (presence of K^+ ions). Under intracellular concentrations of K^+ ions, the H-Telo G-rich sequence assembled into a mixture of hybrid-type 1 and hybrid-type 2 GQ structures, wherein hybrid-type 2 was observed as the predominant topology.[19d,e] These structures differ in their loop orientations, propeller loop in the hybrid 1 and hybrid 2 structures is located at the 5′-end and 3′-end, respectively. Additionally, a distinct loop-capping was observed in these structures, hybrid 1 GQ contains an A-A-A triple capping at the 5´-end,[19d] on the other hand, T-A-T triple capping was found at the 3´-end of the hybrid 2 GQ structure.[19e] In the case of longer telomeric sequences containing 4−8 repeating units, they adopt higher-order GQ structures.

Apart from the telomeric regions, G-rich sequences are abundant in the promoter region of the oncogenes, specifically 1 kb upstream from the transcription start site (TSS). Interestingly, the percentage of PQS is high near TSS, which gradually reduces upon going away from the TSS. This survey indicates that the promoter GQ structures could involve in the process of gene regulation. It was observed that promoter G-rich sequences comprising multiple G-tracts of different lengths could fold into various GQ structures using a distinct combination of G-tracts and intervening loop residues. One such sequence is 27 nucleotides long G-rich segment from the *cMYC* promoter region, which contains five G-tracts of different lengths. In this sequence, the formation of three possible parallel GQ structures using 1234, 2345 and 1245 G-tracts has been reported, in which a parallel GQ formed by 2345 G-tracts is the major component (Figure 2A).[20] Further, the hTERT promoter region contains 68 nucleotides long G-rich sequence, which contains twelve G-tracts and could adopt more complex GQ structures (Figure 2B).[21a] In this sequence, there are three PQS namely PQS1, PQS2 and PQS3 present. Wherein, isolated PQS1 folds into a mixture of parallel and hybrid GQ structures, and another isolated PQS3 adopts a parallel topology.[21b,c] On the other hand, PQS2 alone could not fold into a stable GQ structure. In the case of full-length sequence, the model structure demonstrated the formation of three stacked parallel GQ structures. However, another model structure displayed a very unusual structure comprising a long hairpin structure joining the two stacked parallel and antiparallel GQs.[21a] A recent study presented experimental proofs supporting the formation of stacked, three parallel GQ structures in a full-length sequence.[21d] Similarly, the promoter region of *BCL-2*, *cKIT*, *KRAS*, *HRAS*, *VEGF*, and *PDGF-A*, etc oncogenes fold into different GQ structures.[22]

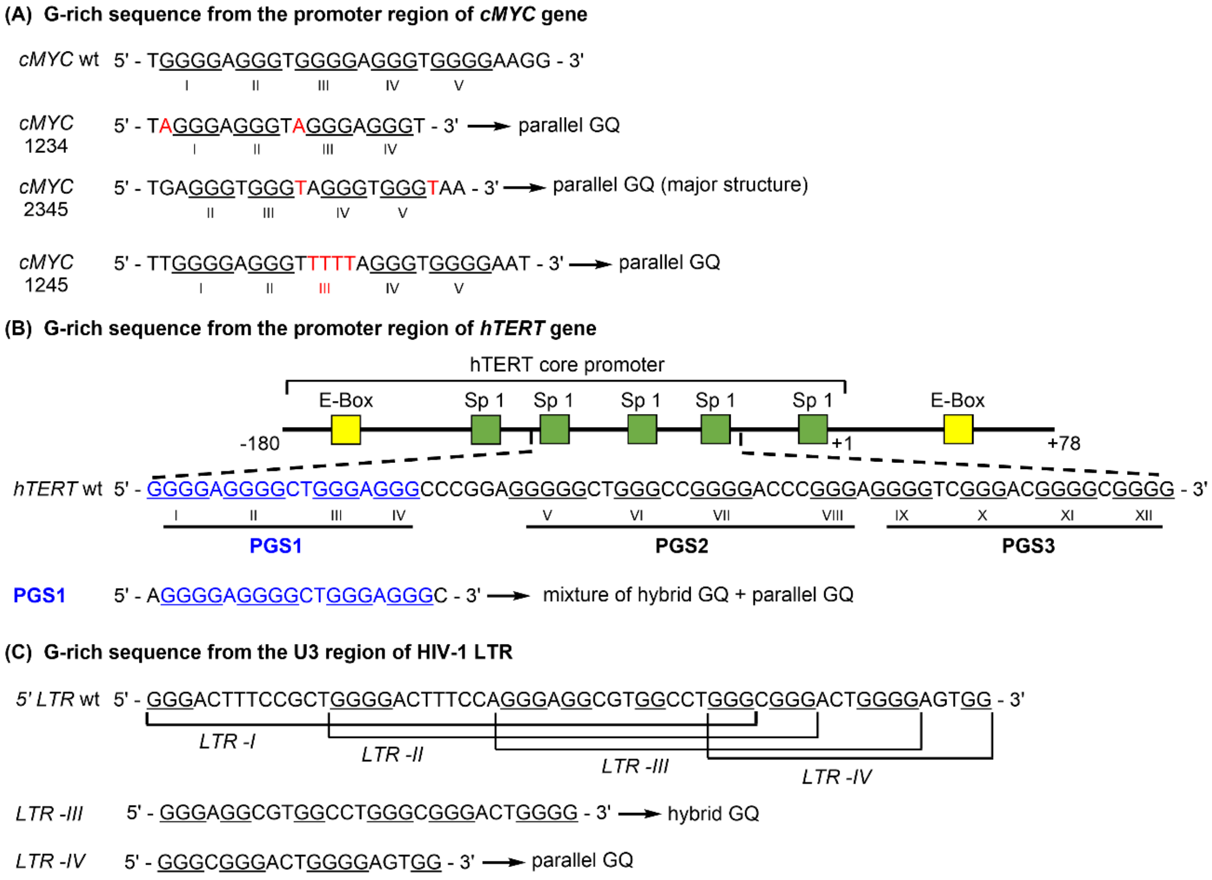


Figure 2. G-rich sequences from the promoter region of (**A**) cMYC gene, (**B**) hTERT gene and (**C**) LTR region of HIV-1 contains multiple G-tracts and are known to fold into more than one GQ structure.

Extensive research has been conducted to understand the GQ structure and its importance in cancer treatment. A recent study highlighted that PQS are present in the genome of pathogens like bacteria and viruses.[23] They play important roles in the virus life cycle and hence, targeting these structures with ligands could emerge as an antiviral strategy. One of the important viral targets is the HIV-1 virus, 5′-long terminal repeat (5′-LTR) of the HIV proviral DNA promoter contains a long G-rich sequence (-100 to -48) (Figure 2C).[24a] The LTR G-rich comprises three PQS, namely LTR-II, LTR-III and LTR-IV. Structural analysis of isolated LTR-III G-rich sequence suggested the formation of a hybrid GQ structure containing a hairpin structure in the 2nd loop.[24b] Similarly, the LTR-IV G-rich sequence folded into a parallel GQ topology with a single-thymidine bulge.[24c] However, the full-length sequence adopts the LTR-III GQ (hybrid topology) structure under physiological conditions, but addition of the GQ binding ligand induces LTR-IV GQ structure. Biochemical studies highlighted that the formation of GQ (LTR-III GQ) in this region acts as a transcription repressor because this region represents the binding site of transcription factors.[24d] On the other hand, stabilization of the LTR-IV GQ topology resulted in the activation of transcription.

1.2.2 RNA G-quadruplex

There are two main structural differences between DNA and RNA GQ structures, (i) the absence of a methyl group in the uracil base and (ii) the presence of a 2′-hydroxy group in the ribose ring of RNA. Interestingly, the 2′-hydroxy group plays a crucial role in determining the topology of RNA GQ, it exerts a steric hindrance on the glycosidic torsion angle resulting in the formation of anti-glycosidic conformation of guanosine and hence, RNA predominantly folds into a parallel GQ topology irrespective of the environmental conditions.[18a,25a] Also, the 2′-hydroxy group forms intramolecular hydrogen bonding with O_4' atom of the ribose, phosphate group, backbone oxygens, and with the N_2 of central guanosines involved in GQ formation. As a result of that RNA GQ contains fewer accompanying water molecules compared to DNA GQ.[18a,25b] Hence, due to the presence of the 2′-hydroxy group, RNA GQs are more stable than DNA GQs.[25b] The formation of RNA GQ was observed in the telomeric repeat-containing RNA (TERRA) and non-coding regions of the mRNA (5′ and 3′ untranslated regions) like NRAS, Zic-1, BCL-2, TRF2, and VEGF.[25c]

1.3 G-quadruplex binding ligands

G-quadruplex is implicated in a wide range of cellular functions, like telomere preservation, replication, transcription, translation, DNA damage repair and epigenetic regulation processes.[26a,b] Hence, small molecule ligands that induce or stabilize GQ structures are considered promising antitumor agents. They suppress the tumor growth using three different pathways (i) by blocking the telomeres activity in the telomer region, which is upregulated in

85 to 90% of cancer cells, (ii) by downregulating the gene expression, and (iii) by inducing DNA damages or point mutations during replication/transcription, which increases genome instability and leads to the cell apoptosis in tumor cells (Figure 3).[26a] Over the last two decades, several ligands have been developed to stabilize DNA or RNA GQ structures, and some of them displayed good activity against the tumor cells. Most of the ligand contains aromatic cores that interact with the planer G-tetrad of GQ and some charged side chains which can interact with grooves and loop residues (Figure 4A). Ligands could interact with GQ structure through different binging modes (external stacking, intercalation, and groove binding) and stabilizes a certain topology or transform it into another topology.[26c] The stabilization of GQ structure in the presence of ligands is known to inhibit the helicases-mediated GQ unwinding process and hamper the progression of polymerase, which induces genome instability (Figure 3).[5c,26a] However, GQ exhibits structural polymorphism and shares similarities in their structural skeleton, owing to that it remains a major challenge to develop a ligand possessing an affinity for a particular GQ topology. Extensive efforts have been made to enhance the selectivity and affinity of ligands, leading to the discovery of a few ligands that interact specifically with one target.[5c,26d]

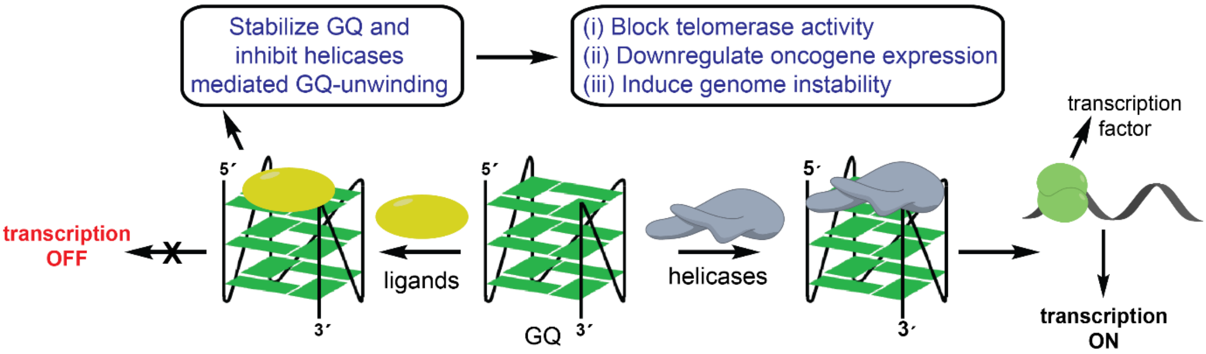
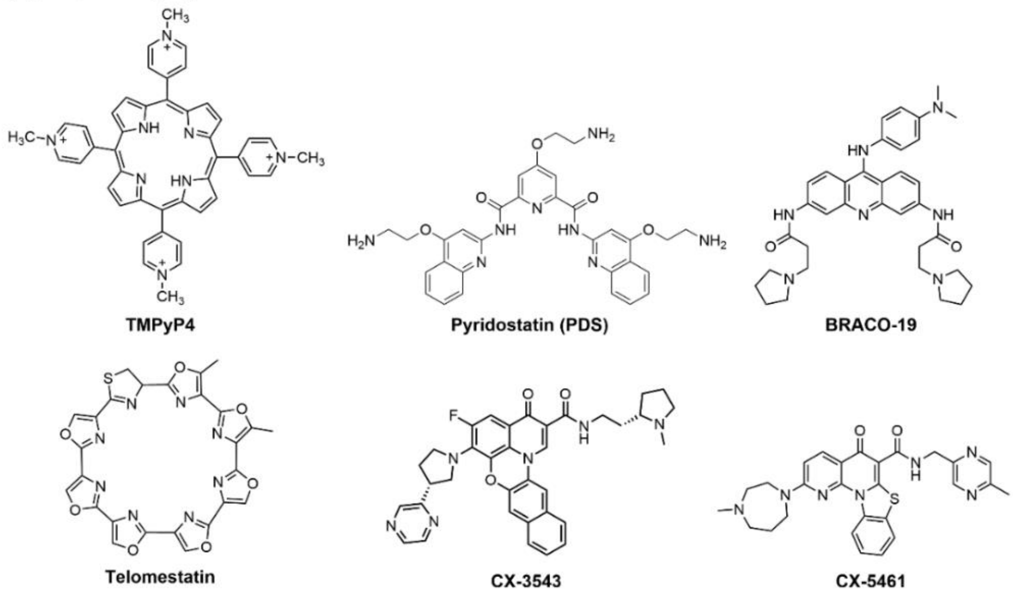


Figure 3. Schematic picture of the involvement of GQ structure in the gene regulatory process in the presence and absence of ligand.

(A) GQ binding ligands



(B) iM binding ligands

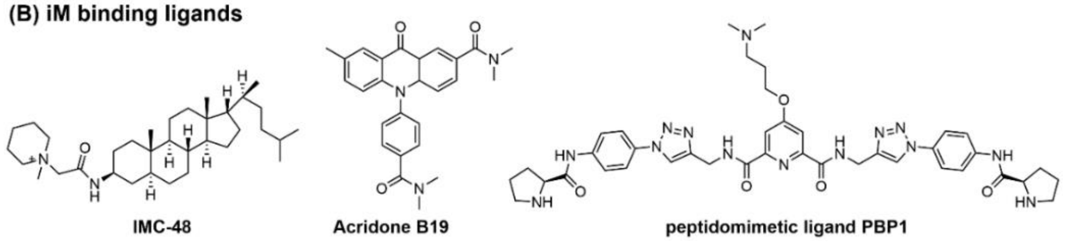


Figure 4. Structure and common name of some of the representative (**A**) GQ binding ligands and (**B**) iM binding ligands.

1.4 i-Motif

The complimentary sequence to the GQ forming sequence is a cytosine (C)-rich sequence, and it folds into another tetraplex structure called i-motif (iM). It contains two intercalated parallel-stranded duplexes in an antiparallel direction formed by hemi-protonated cytosine-cytosine (C:C$^+$) base pairs (Figure 5A).[27] Notably, the iM structure is more compact compared to other nucleic acid structures (B-DNA, GQ). It is a right-handed helical structure with a twist angle of 12 - 20° and the distance between two adjacent base pairs is 3.1 Å. These values are significantly smaller than those observed to B-DNA (36° and 3.4 Å).[27b,28a] Also, formation of the iM structure generates two wide and two narrow grooves, wherein a narrow groove is extremely narrow (3.1 Å against ~5.7 Å for B-DNA).[28a] This compactness in the iM structure creates steric clashes between inter-strands present at short distances. Hence, formation of the RNA iM structure is less stable, as the presence of 2´-OH group in the ribose ring of RNA increases additional steric clashes.[28b] The stability of an iM structure depends on the sequence, pH, molecular crowding and negative superhelicity.[27,29] A sequence that could adopt an iM structure with higher number C:C$^+$ base pairs and longer central loops shows elevated stability compared to fewer number C:C$^+$ base pairs and shorter central loops. [29a,b]

The formation of iM structures requires slightly acidic conditions because protonation of one of the paired cytosine is required to form C:C$^+$ base pairs.[27] Hence, the existence of an iM structure under cellular conditions was less anticipated. Recently, the development of iM-specific antibody and in-cell NMR-based studies demonstrated proof of the existence and stabilities of an iM structure under the cellular environment.[30] Hence, specific binding of proteins and small molecules with an iM structure could control the gene expression (Figure 5B). Recent studies displayed that the binding of hnRNP LL, hnRNP K and BmILF proteins with iM structures formed in the promoter regions of *BCL-2*, *cMYC* and *Bombyx mori* genes, respectively resulted in the activation of transcription.[31] Additionally, the binding of ligands like IMC-48, IMC-76, acridone derivatives, peptidomimetic ligand PBP1, etc with iM structure could either upregulate or downregulate the transcription process (Figure 4B and 5B).[27,31a,32] These evidences highlighted the importance of an iM structure in the regulation of gene expression. Hence, targeting the iM and GQ structures simultaneously with ligands could control the regulation of gene expression effectively.

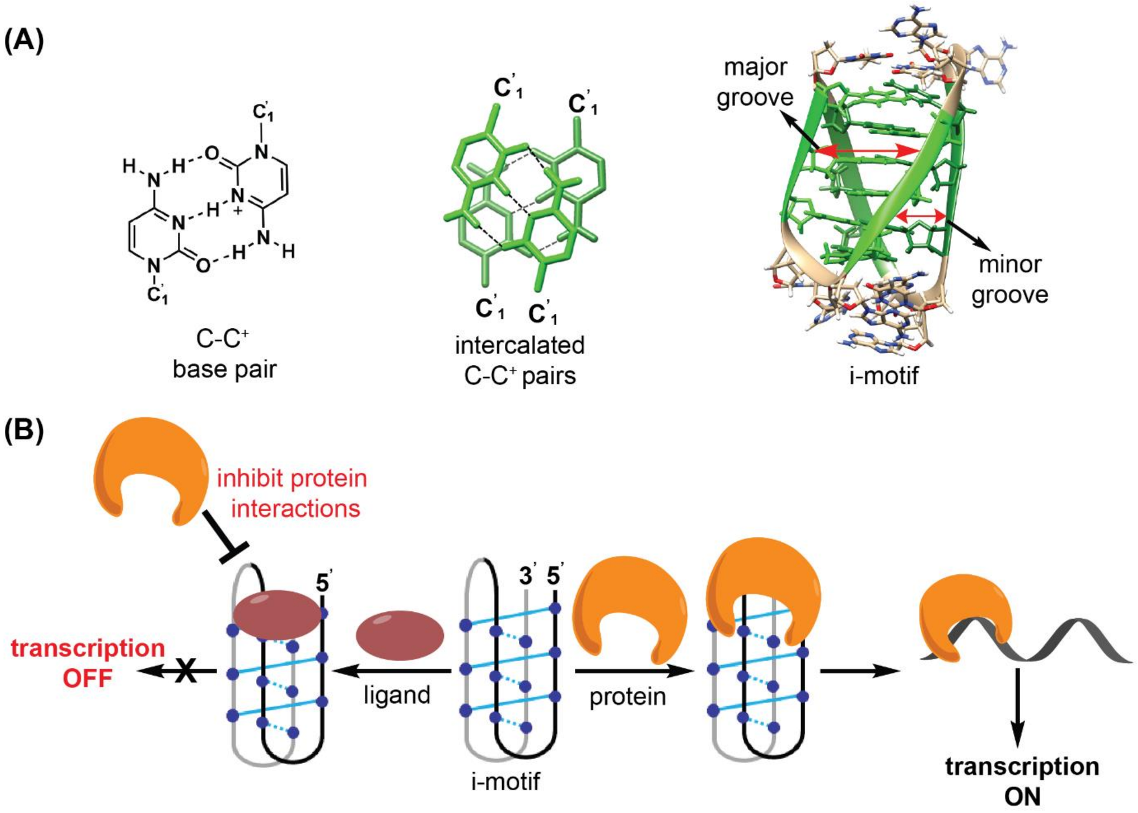


Figure 5. (**A**) Structure of hemiprotonated cytosine-cytosine$^+$ base pair, intercalated iM tetrad, and structure of the 5´E human telomeric iM (PDB: 1EL2).[27c] (**B**) Schematic presentation of the importance of iM structure in cellular processes in the presence and absence of ligands.

1.5 Techniques to study tetraplex structures and their interactions with ligands

1.5.1 CD spectroscopy

Nucleic acids display distinguishable CD profile for different structures (A-form, B-form, Z-form, GQ, iM, etc).[33a,b] Since, nucleic acid conformations differ in the strand orientations and conformation of nucleosides (syn or anti), which gives them different chirality.[33a,b] Mainly, purine and pyrimidine bases of nucleic acids are responsible for the CD absorption profile. CD spectra of B-DNA duplexes show a positive signal at ~260–280 nm and a negative signal at ~245 nm, but A-DNA/RNA shows a positive and negative band at ~260 nm and ~210 nm, respectively. Similarly, tetraplex structures also exhibit characteristic CD signatures. The parallel GQ topology of DNA/RNA displays a strong positive band at ~260 nm and a negative band at ~240 nm with a half intensity of the positive band.[33c] For the antiparallel GQ topology, the CD shows a positive band at ~290 nm and a negative band at ~260 nm. Similarly, the hybrid GQ displays a strong positive band at ~290 nm, shoulder at ~260 nm and a small negative band at ~240 nm. Further, the i-motif structure exhibits a positive band at ~285–290 nm and a negative band at ~260 nm.[31a] Hence, the CD technique is highly beneficial to differentiate various nucleic acid structures. However, it provides only a qualitative perception of nucleic acid conformations and is difficult to understand when multiple structures exist in the equilibrium.

1.5.2 UV spectroscopy

UV-thermal melting study: Nucleic acid shows UV absorbance in the region of 200 nm to 300 nm due to the purine and pyrimidine bases. The involvement of these bases in hydrogen bonding and base stacking during the formation of secondary structures results in hypochromicity.[34a] Hence, the UV-thermal melting technique has been used to determine melting temperature (T_m) of nucleic acid structures by recording changes in the absorbance at 260 nm. Wherein, the melting of these structures breaks the hydrogen bonding and unstacked nucleobases which exposes them to solvents and increases absorbance. However, some structures (GQ and iM) upon melting display a decrease in the absorbance at 295 nm. Surprisingly, for these structures, changes in the absorbance are significantly high (50−80%) at 295 nm compared to 260 nm (4%) wavelength.[34b] This might be due to the higher contribution of n to л* transition at this wavelength, which could be possible because of higher stacking interactions in the folded tetraplex (GQ and iM) structures.[34c] Hence, the melting temperature of a GQ structure is usually determined by recording changes in the absorbance at

295 nm.[34b] However, when multiple GQ structures exist in the equilibrium, the UV-thermal melting technique could not provide the melting temperature of individual components.

Thermal difference spectra (TDS): Nucleic acid structures exhibit different absorbance in the folded and unfolded state. Therefore, the TDS of nucleic acid structures can be derived by subtracting the UV absorption spectrum of folded state from the unfolded state recorded at temperatures below and above its melting temperature (T_m).[35] Similar to CD, the TDS generates a unique signature for different nucleic acid structures. TDS of a GQ structure displays two positive bands at ~243 nm and ~273 nm and one negative band at ~295 nm. Similarly, TDS of an iM structure shows a positive band at ~239 nm and a negative band at ~295 nm.[35a] TDS is a simple and complementing technique to identify the formation of GQ and iM structures, but it can’t be used to distinguish between different GQ topologies.

1.5.3 NMR spectroscopy

NMR spectroscopy is a very powerful technique to investigate nucleic acid structures, it provides a high-resolution structure in a solution state at the atomic level.[36] Apart from that, it is useful to explore the nucleic acid structure folding kinetics, dynamics and their interactions with proteins and ligands. In the folding process, nucleic acids form different types of hydrogen bonding (Watson-Crick, Hoogsteen, C-C^+ hemiprotonated), which can be distinguished by characteristics of 1D NMR signals for imino protons. The formation of a duplex structure that involves Watson-Crick H-bonding is identified by the presence of imino proton signals between 12−14 ppm. The GQ structures formed by Hoogsteen hydrogen bonding show imino proton signal in the range of 10−12 ppm. Similarly, the iM structures containing hemiprotonated C-C^+ base pairs display imino proton signals within the range of 15−17 ppm.[37] However, some sequences show broad and merged peaks due to the signal overlapping arising from the dynamic equilibrium of multiple structures. Particularly, the promoter region of the oncogenes is known to fold into more than one GQ and iM structure, and they could exist in the complex structural equilibrium. Hence, it is difficult to analyze the GQ/iM structures formed by these sequences. To get good resolved imino proton peaks, mutated sequences are used such that they prefer to fold into one topology and thereby enabling the characterization of their structure. Initially, the GQ structure was mainly studied by using the NOESY technique which identifies interacting pairs of protons.[36b] However, structural assignment using the NOESY was based on the model approach and hence could mislead the structural interpretation. To overcome this, a site-specific low-enrichment isotope labeling approach has been used to assign imino protons and other resonances (Figure 6).[36d] To obtain the solution structure of GQ and iM, the combination of NOESY, COSY, TOCSY, ^{13}C-^{1}H HMBC and ^{13}C-^{1}H and ^{15}N-^{1}H HSQC experiments could be performed using either native or isotope-labeled ON sequences. However, the sequence with a maximum length of 100 nucleotides only can be studied using this technique, because large nucleic acid structures show complex resonance overlapping.[35a]

Trantírek and Dötsch group developed an in-cell NMR method to analyze the nucleic acid structures in *Xenopus laevis* (frog) oocytes. Interestingly, using the 1D NMR technique, they compared the formation of human telomeric GQ structure under *in vitro*, *ex vivo* (PEG, Ficol 70, *Xenopus laevis* egg extract) and in-cell conditions.[38a,b] They demonstrated that the telomeric sequence might adopt a mixture of antiparallel, hybrid and parallel topologies under cellular conditions. Recently, Trantírek and coworkers have expanded the in-cell NMR approach to detect the DNA-ligand interactions in the nuclei of human cells.[38c] Additionally, using the 1D NMR technique in live HeLa cells, they highlighted that transfection of the pre-folded iM structure formed by the promoter sequences remained stable in the nuclei of living human cells.[30b,c] However, the major concern is that ONs are not stable under cellular environments for longer periods and resolution of the imino proton signals under *ex vivo* and *in vivo* is very poor. Hence, it is not a viable technique to study the more complex structures under cellular conditions. Also, the promoter region of the several oncogenes could adopt an equilibrium of multiple GQ structures, in this case, the NMR technique is not suitable to monitor their equilibrium.

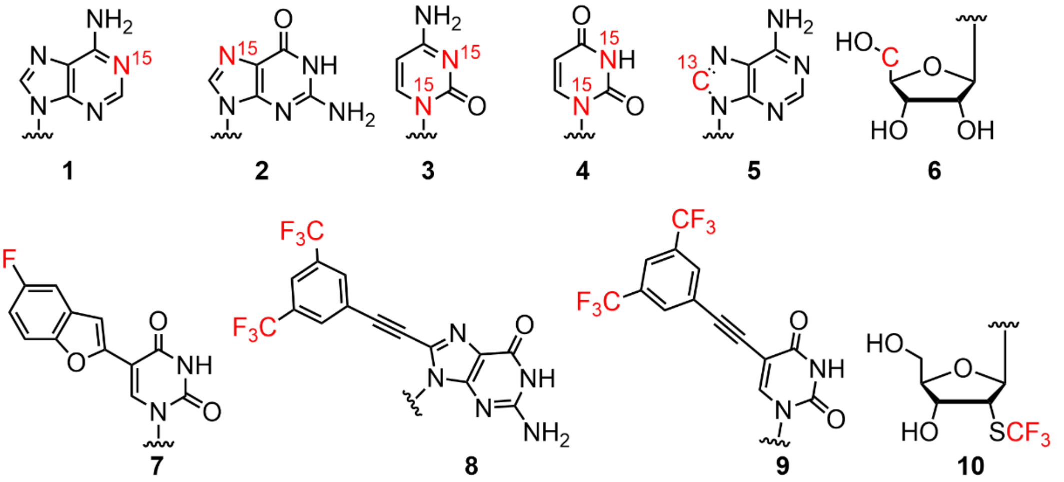


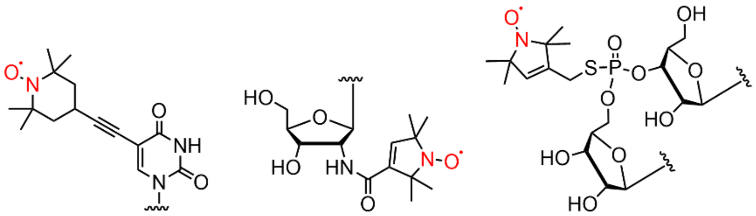
Figure 6. NMR label (^{15}N, ^{13}C and ^{19}F) containing nucleoside analogs to investigate nucleic acid structures using the NMR technique.

Recently, ^{19}F NMR technique is emerging as a very powerful method to analyze nucleic acid structures, their dynamics and interactions with proteins and ligands.[39] Although, it doesn't provide structural information at the atomic level much like ^{1}H and ^{13}C NMR techniques, but it is very useful to monitor the structural dynamics and their interactions with proteins and ligands under different conditions. For this purpose, several ^{19}F-labeled modified nucleosides have been developed and incorporated into the DNA/RNA using a solid phase synthesis or an enzymatic method (Figure 6, **7**−**10**).[39] In our lab, we have developed dual-labeled 5-fluoro-benzofuran-modified deoxyuridine and uridine analogs.[40] Here, the modified deoxyuridine analog was utilized to investigate the human telomeric GQ structures and their interactions with ligand using the ^{19}F NMR technique. Interestingly, in-cell ^{19}F NMR analysis indicated that the labeled H-Telo DNA adopts a hybrid-type 2 GQ topology in live cells. Next, a 5-fluoro-benzofuran-modified uridine analog was used to monitor the different RNA conformations. Similarly, Xu and co-workers have developed 3,5-bis(trifluoromethyl) benzene-modified deoxyguanosine and deoxyuridine probes and employed them to investigate the human telomeric GQ structure and thrombin binding aptamer.[41a,b] Additionally, they introduced a 3,5-bis(trifluoromethyl)benzene core at the 5´ position of H-Telo DNA and RNA sequences and utilized to identify the GQ topology under the cellular environment.[41c,d] Using ^{19}F NMR analysis in live cells, they demonstrated that the labeled TERRA RNA assembled into the higher-order parallel GQ structure in the live cells.

1.5.4 EPR spectroscopy

Electron paramagnetic resonance spectroscopy (EPR) is another important technique, used to understand the nucleic acid structure and dynamics. Generally, the NMR technique uses the magnetic properties of nuclear spin, but EPR utilizes the magnetic moment of the unpaired electron.[42] Natural nucleic acids are diamagnetic and don't contain unpaired electrons. The absence of spin labels in biomolecules eliminates the possibility of any background signal and hence the EPR technique is highly sensitive compared to the NMR technique. Importantly, the spin-label-modified nucleoside should be chemically inert and thermodynamically and kinetically stable under *in vitro* and *in vivo* experimental conditions. In this regard, several nitroxide spin-labeled modified nucleosides have been developed using different strategies (Figure 7).[42a] The nitroxide free-radical moiety can be attached to the nucleobase, sugar, or phosphate backbone of the DNA/RNA and utilized to monitor the nucleic acid structure, dynamics and their interactions with proteins and ligands.[43] Generally, pulsed electron double resonance (PELDOR) or double electron–electron resonance (DEER) EPR experiments are used to understand the inter-nitroxide distances and to monitor the global conformational changes. Particularly, PELDOR has been used to monitor the RNA dynamics, RNA-ligand interactions and protein-nucleic acid interactions.[17b] Using the DEER EPR technique, the Drescher group has determined the h-Telo GQ conformations under K^+ ion conditions.[43] Recently, the Clever group has used metal complexes as a rigid spin-label to determine the dimeric GQ structure using pulsed dipolar EPR spectroscopy.[44a] Importantly, they were able to distinguish the head-to-head and tail-to-tail GQ dimers using their intermolecular distance measurement. Similarly, the Qin group has modified the phosphate group by nitroxide spin-labeled and used to investigate the formation of H-Telo GQ structures under different conditions (salt, molecular crowding, ligand).[44b] In this study, they used distance measurement data to assign the different GQ conformations and demonstrated that the H-Telo sequence folds into a mixture of parallel and antiparallel GQ structures.

The absence of spin-labeled biomolecules, high sensitivity and the ability to measure long-range dipolar coupling between the spin-label makes the EPR a very promising technique to investigate the nucleic acid structure under cellular conditions.[42] Additionally, this technique is not limited by the size of bio-macromolecule like the NMR technique and could analyze coexisting structures. Recently, in-cell EPR analysis of the H-Telo GQ forming sequence has indicated that it folds into a mixture of parallel and 3-tetrad antiparallel conformations in the cellular conditions. However, one of the major drawbacks while working with EPR is that shock-frozen cells at cryogenic temperature require to perform in-cell EPR experiments because free radicals get quickly reduced in cellular environments.[45]



11 **12** **13**

Figure 7. A spin label containing nucleoside analogs for nucleic acid structural analysis using the EPR technique.

1.5.5 X-ray crystallography

X-ray crystallography is another important technique, which provides precise structural information in a 3D array.[46] The crystal structure of nucleic acids is established from the non-covalent interactions between nucleic acids, wherein free space occupied by water molecules. In the crystal structure, mapping the electron density is difficult as all parts of the sequence could show similar diffraction from the presence of similar atoms (hydrogen, carbon, nitrogen, and oxygen). Hence, heavy atom derivatizations of the nucleic acid are widely used method to map the electron density.[46] Bromine, Iodine and Selenium chemical modifications are generally used as heavy elements (Figure 8),[46–48] also cationic forms of alkali metals, transition metals and lanthanides can be appointed into the crystal structure by the soaking method.[49] Importantly, the incorporation of a heavy element into the crystal structure should not affect its native structure. A heavy element containing a larger number of electrons gives pronounce changes in the diffraction pattern compared to the native crystal. Hence, Selenium modified nucleosides have been incorporated into nucleic acids and used for crystallization (Figure 8).[48] However, nucleic acids show structural polymorphism, so it is hard to predict whether they would adopt similar topology under crystallization conditions (synthetic molecular crowding and dehydrating) and a complex cellular environment.

14 15 16 17 18

Figure 8. A heavy element (Br, I, Se) containing nucleoside analogs to study nucleic acid structures by X-ray crystallization technique.

1.5.6 Fluorescence spectroscopy

Fluorescence spectroscopy is another important, highly sensitive and widely used technique to study nucleic acid structures and folding dynamics *in vitro* and *in vivo* conditions. However, native nucleic acids are non-fluorescent, hence several fluorescent probes (covalent fluorescence label or non-covalent fluorescence binder) have been developed. Local environment of the nucleic acid structure changes, when they undergo conformational transitions and fluorescent labels are designed to report these changes in terms of their spectral properties (intensity, emission maximum, quantum yield, lifetime and anisotropy).[50] Here, we have discussed the different fluorescence labeling strategies used to investigate the GQ and iM structures, dynamics and their interactions with ligands *in vitro* and *in vivo.*

1.5.6.1 Covalent fluorescence label

1.5.6.1.1 Single-molecule Förster resonance energy transfer (smFRET)

smFRET is a highly useful technique that enables the real-time monitoring of individual nucleic acid structures with high resolutions.[51] This technique involves the measurement of changes in the distance between the donor and acceptor dyes using energy transfer efficiency. It displays a distinct energy transfer efficiency pattern for different GQ topologies (parallel GQ, antiparallel GQ and hybrid GQ) as their end-to-end distance are not similar. Using the different combinations of donor and acceptor dyes, smFRET has been used to monitor the GQ dynamics, their kinetics, also their interactions with proteins and ligands.[51] It has also used to detect the other nucleic acid structural transitions like triplex to duplex and B-DNA to Z-DNA, etc. However, data reproducibility of a particular target could vary with some parameters like the location of the donor and acceptor in the DNA construct, annealing conditions and imaging surface.[51a] This makes it time-consuming and not a very straightforward technique.

1.5.6.1.2 Fluorescent nucleosides

The incorporation of a fluorescent nucleoside analog at a particular site of interest into the tetraplex structures could photophysically provide a better understanding of the structure and interaction with proteins and ligands. This will further help to develop a binding assay to identify GQ/iM-specific binders. In this context, several fluorescent nucleoside analogs have been developed and employed to investigate nucleic acid structures.[52] These analogs are broadly categorised into the following types.

Isomorphic fluorescent nucleoside probes

2-aminopurine (2AP) has been widely used to monitor different nucleic acid structures, dynamics and interactions with proteins and ligands.[53a] It is a structural isomer of adenine and could form a Watson-Crick hydrogen bonding with thymine. It has a high fluorescence quantum yield and its photophysical properties are sensitive to changes in solvent polarities. Majima and co-workers have explored the utility of 2AP to monitor the formation of human telomeric GQ structure. They substituted a loop adenosine residue with 2AP in the H-Telo GQ forming sequence, which didn’t affect the formation of a native GQ structure and displayed higher fluorescence and lifetime for the antiparallel GQ structure compared to its duplex structure. In the duplex structure, 2AP interacts with the adjacent G-residue through stacking interactions and displays quench fluorescence due to the electron transfer process. However, in the GQ structure, distorted conformations of loop residues reduce 2AP stacking interactions with G-residue and expose it more to solvents. Hence, 2AP displayed enhanced fluorescence in the GQ structure.[53b] Similarly, Sugiyama and co-workers utilized a 2AP analog to detect the formation of GQ and iM structure in the 5´ end of the retinoblastoma (Rb) gene. Remarkably, 2AP exhibited high fluorescence and lifetime for tetraplex structures (GQ/iM) compared to the duplex structure. The fluorescence outcome of 2AP in the tetraplex structure could be due to the reduced stacking interactions with neighbouring bases. Although 2AP is a good fluorescent probe, its excitation and emission wavelengths are in the UV range, also its quantum yield is dramatically reduced when incorporated into ONs and this decreases its sensing ability.

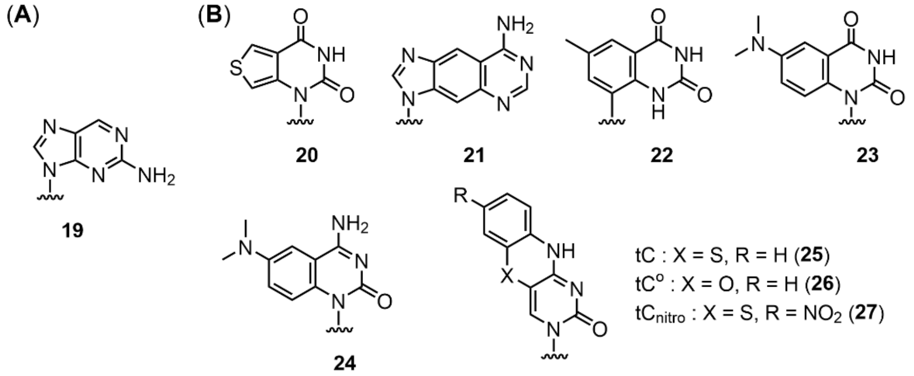


Figure 9. Chemical structures of (**A**) isomorphic fluorescent nucleoside and (**B**) size expanded fluorescent nucleoside analogs.

Expanded fluorescent nucleosides

Tor and co-workers have developed a fused thiophene derivative of uridine (Figure 9B, **20**), which exhibits fluorescence solvatochromism and a high quantum yield.[54a] It was utilized to detect mismatches in duplexes and the formation of abasic sites in the RNA. Kool and co-workers have reported benzo-fused purine and pyrimidine fluorescent nucleoside analogs (Figure 9B, **21**, **22**) and demonstrated their use as a new genetic system that could be orthogonal to the natural one.[54b] Additionally, Luedtke and co-workers developed and utilized modified thymidine (**23**) and modified cytidine (**24**) fluorescent analogs to study the formation of different metal base pairs (T-Hg-T and C-Hg-T) and their thermodynamics using fluorescence measurements.[54c] Wilhelmsson and Albinsson co-developed a family of tricyclic 2´deoxycytidine nucleoside analogs (tC, tC^{O} and tC$_{nitro}$, **25**–**27**).[54d,e] Among these tC and tC^{O}, both nucleosides and their corresponding modified DNA after incorporation portray high quantum yields. On the other hand, tC$_{nitro}$ is a non-fluorescent analog. Since the lowest energy absorption maxima of the tC$_{nitro}$ overlaps with the emission wavelength of tC^{O}, they have used these modified nucleosides as a fluorescence resonance energy transfer (FRET)-pair.[54f] The efficiency of this FRET base pair was evaluated by determining the distance between base pairs located at different positions in DNA ONs. Further, Juskowiak and co-workers used tC nucleoside to modify a C-rich sequence, which was utilized as a pH-sensitive aptasensor.[54g]

Extended-conjugation-based fluorescence nucleosides

Tor's group synthesized fluorescent nucleoside analogs by conjugating heterocyclic rings to the purine and pyrimidine bases. They conjugated different heterocyclic rings like furan, thiophene, oxazole, and thiazole with a pyrimidine base and found that furan-modified pyridine (Figure 10, **28**) exhibits probe-like properties i.e. emission maxima in the visible region and photophysical properties sensitive to changes in solvent polarities.[52,55] Further, the probe was utilized to (i) detect abasic sites,[56a] (ii) monitor RNA-aminoglycosides interactions[56b] and (iii) RNA-protein interactions.[56c] Srivatsan and co-workers have also developed a series of fluorescent nucleosides by attaching a selenophene, indole, N-methyl indole, benzothiophene, benzofuran and 5-fluorobenzofuran heterocycles with pyrimidine bases (Figure 10, **29**−**34**).[55,57] Out of them, selenophene, benzothiophene, benzofuran and 5-fluoro-benzofuran modified uridine and deoxyuridine display probe-like properties.[40a,b,57a,b,c] They have utilized benzofuran-modified pyrimidine analogs to detect the abasic sites in DNA and RNA.[57c,d] Further, they expanded the utility of a benzofuran-modified deoxyuridine analog to investigate the different nucleic acid structures and were able to distinguish duplex, iM and different GQ topologies using fluorescence and lifetime measurements. However, this probe is not useful for cell-based analysis as its emission maxima is in the UV region. They overcome this problem by developing a 5-fluorobenzofuran-modified deoxyuridine analog and were able to identify the physiologically relevant H-Telo GQ topology under the cellular condition using in-cell ^{19}F NMR analysis.[40a] Similarly, Manderville and co-workers have developed several fluorescent guanosine analogs by attaching furan, pyrrole, thiophene, benzofuran, indole, benzothiophene and benzonitrile moieties to guanosine nucleoside (Figure 10, **35**−**41**).[55] These fluorescent analogs display moderate to high quantum yield and fluorescence solvatochromism and viscochromism. They have utilized these fluorescent nucleoside analogs to investigate the human telomeric GQ structures and the GQ structure of thrombin-binding aptamer.

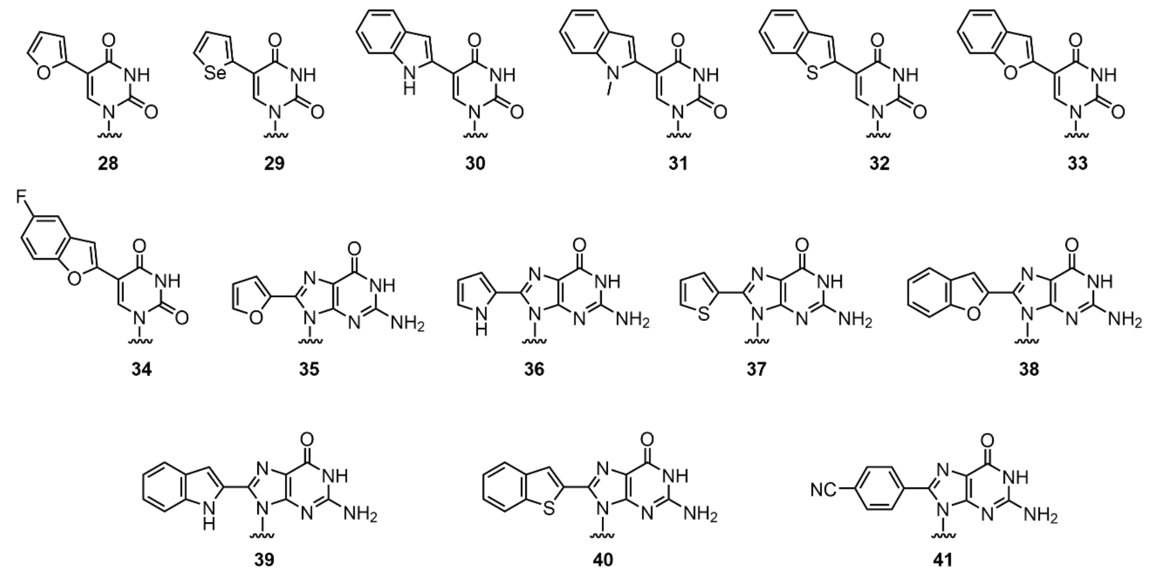


Figure 10. Chemical structures of the extended conjugation-based fluorescent nucleoside analogs.

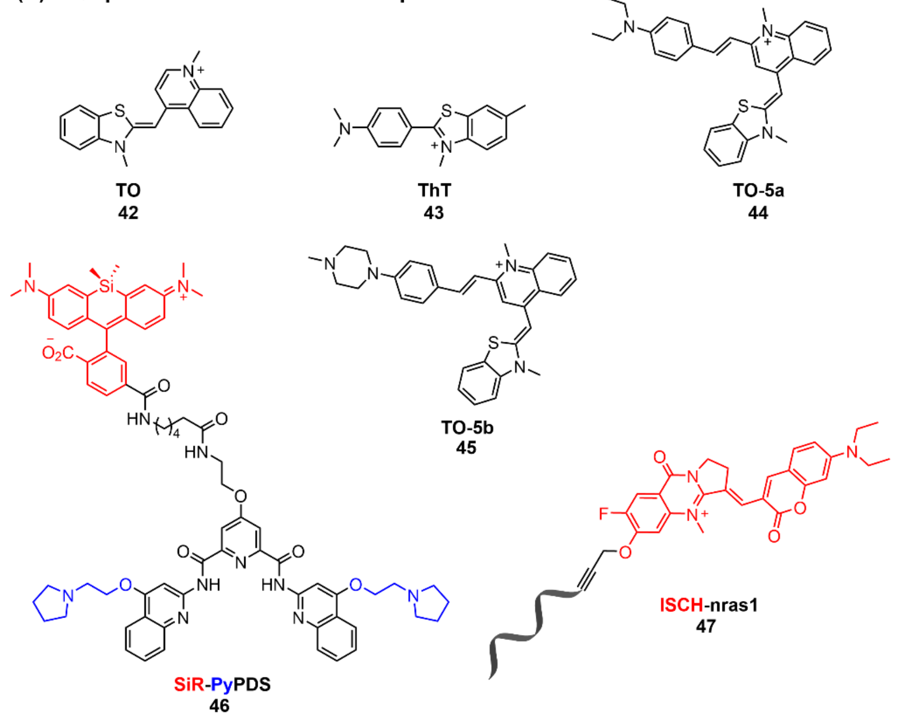
1.5.6.2 Non-covalent fluorescent binder

The structure-specific fluorescent binder is a powerful tool to determine the formation and localization of bio-macromolecules in living cells. The ideal fluorescent probe to investigate the GQ and iM should have the following properties (i) specific to a particular structure, (ii) turn-on fluorescence nature upon binding, (iii) high binding affinity, (iv) water soluble, (v) cell-permeable, (vi) high quantum yield and (vii) non-toxic.[50] In this regard, cationic fluorogenic dyes, namely thiazole orange (**TO**, **42**) and thioflavin T (**ThT**, **43**) are used as fluorescent turn-on GQ sensors to monitor the formation of GQ structure and their interactions with ligands (Figure 11A). Importantly, they specifically bind to GQ structure over single-stranded and double-stranded nucleic acid structures.[50a] These dyes are weakly fluorescent in an aqueous solution with or without non-GQ structures but display a very high fluorescence upon binding to the GQ structure. Recently, Shangguan and co-workers have developed four **TO** derivatives by attaching different styryl groups, out of them **TO-5a** (**44**) derivative displayed better selectivity for GQ structure and another derivative **TO-5b** (**45**) exhibited higher affinity for antiparallel GQ. Further, using live cell imaging, they demonstrated that **TO-5b** can enter the cell and stain the DNA and RNA GQ in the nuclei and nucleoli of live cells.[58a] Recently, Balasubramanian and co-workers have developed a GQ-specific fluorescent probe (**SiRPyPDS**, **46**) and utilized it to detect the single-molecule of individual GQ structures in live cells using live cell fluorescent imaging (Figure 11A). Interestingly, they have monitored the population of GQs and also GQ folding and unfolding dynamics in living cells.[58b] Tan and co-workers have developed a fluorogenic hybridization probe (**ISCH-nras1**,

47) to visualize the formation of GQ structure in the 5′-UTR of *NRAS* mRNA. This probe comprised two parts, (i) a fluorescence light-up probe (**ISCH-1**, red color) which acts as a GQ sensor and (ii) a DNA ON complementary to the adjacent region of the GQ forming sequence (Figure 11A). Using this probe strategy, they accurately spotted *NRAS* RNA GQ structure under *in vitro* and *in vivo* conditions.[58c]

Very few fluorescent ligands have been discovered that specifically bind to the iM structure. Pei and co-workers have very first time reported an iM-specific fluorescence light-up probe namely neutral red (Figure 11B).[59a] It is a phenazine dye and could exist in neutral (**NR**, **48**) and protonated forms (**NRH**, **49**) (Figure 11B), and upon binding with the H-Telo iM structure displayed significant enhancement in the fluorescence. They estimated that the protonated form of the probe could only bind with an iM structure. Hence, the iM structure that exists at near-neutral or physiological pH would not be detected by this probe due to its

(A) GQ-specific fluorescence turn-on probes



(B) iM-specific fluorescence turn-on probes

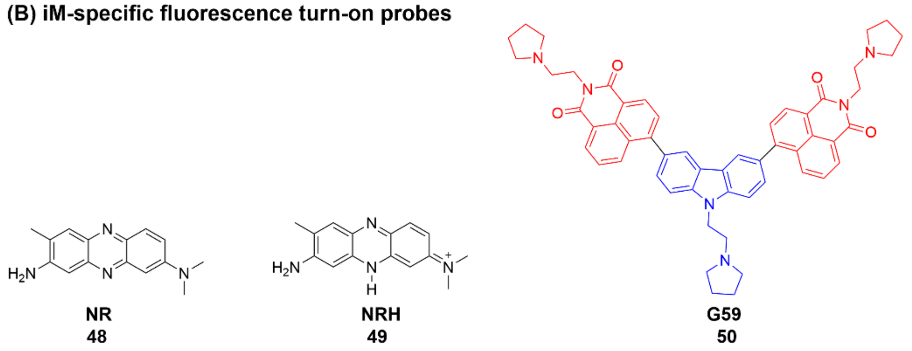


Figure 11. GQ and iM structure-specific non-covalent fluorescence light-up probes

lower pKa value (6.81). Recently, Li and co-workers have developed a fluorescence turn-on probe (**G59**, **50**) that selectively recognizes the iM structure form by the cMYC promoter sequence (Figure 11B).[59b] **G59** comprises a carbazole moiety (blue color) which is known to interact specifically with the cMYC iM structure and 1, 8-naphthalimide derivatives (red color) as a fluorescent probe (Figure 11B). The probe is non-fluorescent under different pH conditions. It is highly specific to the cMYC iM structure and displays very high fluorescence in the presence of cMYC iM structure but it is weakly fluorescent in the presence of GQ, duplex, single-stranded DNA or iM structures formed by other promoter sequences.

1.5.7 Antibody-based tools

Extensive biochemical and biophysical studies confirmed the formation of non-canonical tetraplex structures (GQ and iM) under *in vitro* conditions and a sufficient amount of data demonstrating their involvement in gene regulation processes. However, limited data are available on the existence of these structures in cellular environments. In this regard, the development of a structure-specific antibody that could localize the GQ structure under the cellular milieu has great potential to reveal the structure-function relationship of GQ and iM. Balasubramanian and co-workers identified a structure-specific antibody using a phage display method from the library of 2.3 x 10^{10} different single-chain antibody clones and visualized the DNA GQ structure *in vivo*.[11] Using the ELISA method, they demonstrated that the selected antibody called BG4 has a higher affinity (nanomolar) for intramolecular and intermolecular DNA GQ structure. The BG4 antibody has a similar affinity towards different GQ topologies but didn't bind to hairpin, single-stranded and duplex DNA structures. Further, they proposed that the formation of GQ structure might be associated with the DNA replication process because the number of BG4 foci increased by ~2.5 folds when cells enter into the S-phase during the cell cycle wherein replication occurs. However, GQ shows a high degree of structural polymorphism (parallel GQ, antiparallel GQ and hybrid GQ), so it is crucial to know the relevance of different GQ topologies in terms of their biological functions. Recently, Huang and co-workers reported a new single-chain variable fragment (scFv) antibody (D1), which shows high selectivity for parallel GQ topology *in vitro* and *in vivo*.[60] Using the ELISA experiment, they exhibited that the antibody D1 specifically interacts with the parallel GQ conformations formed by cMYC, KIT-1, KIT-2, HIF-1a, VEGF, KRAS, and RET ONs but doesn't interact with the hybrid GQ (H-Telo), antiparallel GQ (TBA and HRAS-1), duplex and single-stranded structures. Next, they transfected Cy5-labeled cMYC or TBA ONs into cells and cells stained using the DAPI and antibody D1. They observed that the antibody D1 foci were co-localized with only cMYC foci, but didn't co-localize with TBA foci. This confirmed that the antibody D1 specifically recognizes parallel GQ topology in cellular conditions.

Another important advancement in this field was done by Christ and co-workers, they isolated an i-motif-specific antibody (iMab) from the library of Garvan-2 human single-chain variable fragment (scFv) using the phage display method.[12] Initial ELISA-based study indicated that the iMab is highly specific to the H-Telo iM structure over duplex DNA, hairpin DNA, microRNA, streptavidin, neutravidin, hen egg-white lysozyme and neuropeptide Y. It exhibits a nanomolar binding affinity for the H-Telo iM structure and doesn't bind to a mutated H-Telo C-rich sequence that could not fold into an iM structure. Similarly, they showed that the iMab specifically binds with the iM structure formed by c-MYC, RET and VEGF C-rich sequences and no binding was observed for their corresponding GQ structures. Further, they employed an iMab to visualize the iM structure in human cells using the immunofluorescent staining technique. This study unravels that the formation of iM structure in cells depends on the cell cycle and is highest at the G1/S phase boundary.

1.6 Statement of the research problem

Significant progress has been made in the development of biophysical tools to investigate nucleic acid structure-function relationship. In particular, several modified nucleoside probes have been developed to understand nucleic acid structures and their interactions with ligands and proteins. However, nucleic acids display a high degree of structural polymorphism and regularly undergoes conformational transitions to perform various functions. Owing to this, limited information is available on what structure nucleic acid adopts under cellular conditions due to the lack of efficient probes or small molecule ligands that could differentiate different nucleic acid conformations. Hence, it is of high interest to develop a biophysical platform that could provide information on different structures adopted by a nucleic acid motif in cell-free and cellular environments, and how different structures interact with small molecule ligands and protein factors. Such an endeavour would greatly benefit in advancing nucleic acid diagnosis and therapeutics as screening assays could be devised to identify structure-specific binder.

Most of the nucleic acid studies use one label one technique approach; this traditional method could give a limited understanding of a targeted system. Here, we were interested in developing a dual-labelled nucleoside analog that would contain a fluorescence and a ^{19}F NMR tag for the following reasons. Conjugating a heterocyclic moiety to a nucleoside can produce microenvironment-responsive fluorescent nucleosides, which could report changes in the nucleic acid conformations through fluorescence. The fluorine tag on the nucleoside would enable nucleic acid analysis by using ^{19}F NMR technique. To develop a dual-labelled nucleoside probe, a 5-trifluoromethyl-benzofuran ring was attached at the C5-position of 2'-deoxyurindine and uridine nucleosides. The modified nucleosides were utilized in exploring the structural polymorphism and ligand/drug interaction of therapeutically relevant DNA and RNA motifs. Incorporation of the modified nucleoside into the G-rich sequence from the promoter region of the EGFR oncogene, which is known to fold into a mixture of parallel and hybrid GQ structures produces fluorescence and ^{19}F NMR signatures for the individual GQ topologies. These signatures were used to quantify the relative population of parallel GQ and hybrid GQ *in vitro* and to understand the effect of ligand binding on their structural equilibrium. Interestingly, ^{19}F NMR signatures helped in determine the GQ topology in a cellular model (frog egg lysate and extracts). Fluorescence and ^{19}F NMR properties of the modified nucleoside were utilized to monitor the formation of H-Telo iM and Braf iM structures under different conditions. Notably, the ^{19}F NMR label was useful in determining the relative population of different iM structures formed by the H-Telo sequence. Further, it aided in identifying the major iM structure from the complex iM structural equilibrium formed by the Braf C-rich sequence. Taking forward, we utilized the modified ribonucleotide analog to explore aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA transcript by fluorescence and ^{19}F NMR techniques. Minimally perturbing nature of the nucleoside analogs, their spectral properties and ability to distinguish different nucleic acid conformations highlight the usefulness our new probe system in nucleic acid analysis.

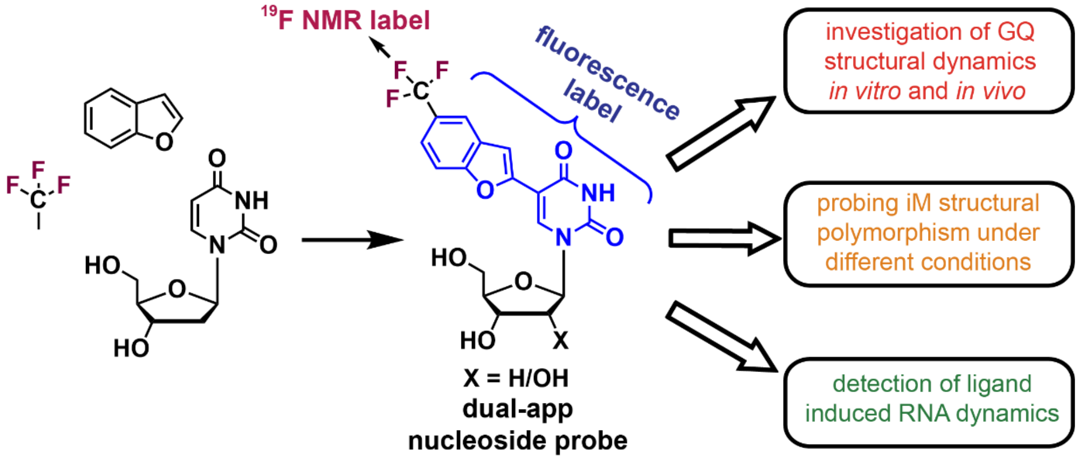


Figure 12. Design strategy for the development of dual-label nucleoside probes and their application in the investigation of nucleic acid structures, dynamics and their interaction with ligands.

1.7 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) Choi, J.; Majima, T. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909. (d) Tateishi-Karimata, H.; Sugimoto, N. *Nucleic Acids Res*. **2021**, *49*, 7839–7855.
2. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. et al. *Nature* **2001**, *409*, 860–921.
3. (a) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. *Cell. Mol. Life Sci.* **2010**, *67*, 43–62. (b) Wang, G.; Vasquez, K. M. *Mutat. Res.* **2006**, *598*, 103–119. (c) Bacolla, A.; Wells, R. D. *Mol. Carcinog*. **2009**, *48*, 273–285. (d) Wells, R. D.; Dere, R.; Hebert, M. L., Napierala, M.; Son, L. S. *Nucleic Acids Res*. **2005**, *33*, 3785–3798. (e) Rich, A.; Nordheim, A.; Wang, A. H. *Annu. Rev. Biochem*. **1984**, *53*, 791–846.
4. (a) Hall, K. B. *Curr. Opin. Chem. Biol*. **2008**, *12*, 612–618. (b) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. *Nat. Rev. Mol. Cell Biol*. **2019**, *20*, 474–489.
5. (a) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res*. **2007**, *35*, 406–413. (b) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res*. **2015**, *43*, 8627–8637. (c) Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K. *Mol. Cancer* **2021**, *20*, 40–57.
6. Grand, C. L.; Powell, T. J.; Nagle, R. B.; Bearss, D. J.; Tye, D.; Gleason-Guzman, M.; Hurley, L. H. *Proc. Natl. Acad. Sci. USA.* **2004**, *101*, 6140–6145.
7. (a) Buske, F. A.; Mattick, J. S.; Bailey, T. L. *RNA Biol*. **2011**, *8*, 427-439. (b) Butcher, S. E.; Pyle, A. M. *Acc. Chem. Res.* **2011**, *44*, 1302–1311.
8. (a) Blaho, J. A.; Wells, R. D. *J. Biol. Chem*. **1987**, *262*, 6082–6088. (b) Kmiec, E. B.; Angelides, K. J.; Holloman, W. K. *Cell* **1986**, *40*, 139–145. (c) Oh, D. B.; Kim, Y. G.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16666–16671.
9. (a) Bae, S.; Kim, D.; Kim, K. K.; Kim, Y.-G.; Hohng, S. *J. Am. Chem. Soc*. **2011**, *133*, 668–671. (b) Lee, A. R.; Park, C. J.; Cheong, H. K.; Ryu, K. S.; Park, J. W.; Kwon, M. Y.; Lee, J.; Kim, K. K.; Choi, B. S.; Lee, J. H. *Nucleic Acids Res*. **2016**, *44*, 2936–2948. (c) Kim, S. H.; Lim, S. H.; Lee, A. R.; Kwon, D. H.; Song, H. K.; Lee, J. H.; Cho, M.; Johner, A.; Lee, N. K.; Hong, S. C. *Nucleic Acids Res.* **2018**, *46*, 4129–4137. (d) Lee, A. R.; Park, C. J.; Cheong, H. K.; Ryu, K. S.; Park, J. W.; Kwon, M. Y.; Lee, J.; Kim, K. K.; Choi, B. S.; Lee, J. H. *Nucleic Acids Res*. **2016**, *44*, 2936–2948.
10. (a) Bedrat, A.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res*. **2016**, *44*, 1746–1759. (b) Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2008**, *36*, 1321–1333.

11. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182−186. 12. 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.
13. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.
14. (a) Tippana, R.; Hwang, H.; Opresko, P. L.; Bohr, V. A.; Myong, S. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 8448–8453. (b) Grün, J. T.; Schwalbe. H. *Biopolymers* **2022**, *113*, e23477–e23492. (c) Takahashia, S.; Brazierb, J. A.; Sugimoto, N. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 9605–9610.
15. (a) Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. *Nucleic Acids Res.* **2018**, *46*, 9264–9275. (b) Lacroix, L.; Séosse, A.; Mergny, J.-L. *Nucleic Acids Res.* **2011**, *39*, e21–e31. (c) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. (d) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature*, **2002**, *417*, 876–880. (e) Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.
16. (a) Stollar, B. D.; Raso, V. *Nature* **1974**, *250*, 231–234. (b) Riesen, A. J. V.; Le, J.; Slavkovic, S.; Churcher, Z. R.; Shoara, A. A.; Johnson P. E.; Manderville, R. A. *ACS Appl. Bio. Mater.* **2021**, 4, 6732–6741. (c) Summers, P. A.; Lewis, B. W.; Gonzalez-Garcia, J.; Porreca, R. M.; Lim, A. H. M.; Cadinu, P.; Martin-Pintado, N.; Mann, D. J.; Edel, J. B.; Vannier, J. B.; Kuimova, M. K.; Vilar, R.; *Nat. Commun*. **2021**, *12*, 162–172.
17. (a) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. *Nucleic Acids Res*. **2011**, *39*, 5768–5775. (b) Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantírek, L.; Dötsch, V. *Angew. Chem. Int. Ed*. **2014**, *53*, 10300–10314.
18. (a) Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867–5892. (b) Webba da Silva, M. *Chemistry* **2007**, *13*, 9738–9745. (c) Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. *Angew. Chem., Int. Ed.* **2009**, *48*, 9167–9170.
19. (a) Hwang, H.; Kreig, A.; Calvert, J.; Lormand, J.; Kwon, Y.; Daley, J. M.; Sung, P.; Opresko, P. L.; Myong, S. *Structure* **2014**, *22*, 842–853. (b) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282. (c) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880. (d) Dai, J. X.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Z. *Nucleic Acids Res*. **2007**, *35*, 2440–2450. (e) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2007**, *35*, 4927–4940.

20. (a) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048–2058. (b) Mathad, R. I.; Hatzakis, E.; Dai1, J.; Yang, D. *Nucleic Acids Res*. **2011**, *39*, 9023–9033. (c) Dickerhoff, J.; Onel, B.; Chen, L.; Chen, Y.; Yang, D. *ACS Omega* **2019**, *4*, 2533−2539.
21. (a) Chaires, J. B.; Trent, J. O.; Gray, R. D.; Dean, W. L.; Buscaglia, R.; Thomas, S. D.; Miller, D. M. *PLoS One* **2014**, *9*, e115580. (b) Lim, K. W.; Lacroix, L.; Yue, D. J.; Lim, J. K.; Lim, J. M.; Phan, A. T. *J. Am. Chem. Soc*. **2010**, *132*, 12331–12342. (c) Micheli, E.; Martufi, M.; Cacchione, S.; De Santis, P.; Savino, M. *Biophys. Chem*. **2010**, *153*, 43–53. (d) Monsen, R. C.; DeLeeuw, L.; Dean, W. L.; Gray, R. D.; T. Sabo, M.; Chakravarthy, S.; Chaires, J. B.; Trent, J. O. *Nucleic Acids Res*. **2020**, *48*, 5720–5734.
22. (a) Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. *J. Am. Chem. Soc.* **2014**, *136*, 1750–1753. (b) Danny Hsu, S.-T.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2009**, *131*, 13399–13409. (c) Cogoi, S.; Xodo, L. E. *Nucleic Acids Res*. **2006**, *34*, 2536–2549. (d) Cogoi, S.; Shchekotikhin, A. E.; Xodo, L. E. *Nucleic Acids Res*. **2014**, *42*, 8379–8388. (e) Wu, Y.; Zan, L. P.; Wang, X. D.; Lu, Y. J.; Ou, T. M.; Lin, J.; Huang, Z. S.; Gu, L. Q. *Biochim. Biophys. Acta* **2014**, *1840*, 2970–2977. (f) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. *Nucleic Acids Res.* **2007**, *35*, 7698–7713.
23. (a) Metifiot, M.; Amrane, S.; Litvak, S.; Andreola, M-L. *Nucleic Acids Res*. **2014**, *42*, 12352–12366. (b) Dumetz, F.; Merrick, C. J. *Molecules* **2019**, *24*, 1339–1349.
24. (a) Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palu, G.; Richter, S. N. J. Med. Chem. 2013, 56 (16), 6521−30. (b) Butovskaya, E.; Heddi, B.; Bakalar, B.; Richter, S. N.; Phan, A. T. *J. Am. Chem. Soc.* **2018**, *140*, 13654−13662. (c) De Nicola, B.; Lech, C. J.; Heddi, B.; Regmi, S.; Frasson, I.; Perrone, R.; Richter, S. N.; Phan, A. T. *Nucleic Acids Res*. **2016**, *44*, 6442−6451. (d) Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palù, G.; Fabris, D.; Richter, S. N.; *Nucleic Acids Res*. **2015**, *43*, 8884−8897.
25. (a) Fay, M. M.; Lyons, S. M.; Ivanov, P. *J Mol Biol*. **2017**, *429*, 2127–2147. (b) Zhang, D. H.; Fujimoto, T.; Saxena, S.; Yu, H. Q.; Miyoshi, D.; Sugimoto, N. *Biochemistry* **2010**, *49*, 4554–4563. (c) Bugaut, A.; Balasubramanian, S. *Nucleic Acids Res*. **2012**, *40*, 4727–4741.
26. (a) Sun, Z.-Y.; Wang, X.-N.; Cheng, S.-Q.; Su, X.-X.; Ou, T.-M. *Molecules* **2019**, *24*, 396–424. (b) Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. *Pharmaceuticals* **2021**, *14*, 769–

808. (c) Ma, Y.; Iida, K.; Nagasawa, K. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 3–17. (d) Asamitsu, S.; Bando, T.; Sugiyama, H. *Chem. Eur. J.* **2019**, *25*, 417–430.

27. (a) Day, H. A.; Pavlou, P.; Waller, Z. A. E. *Bioorg. Med. Chem.* **2014**, *22*, 4407−4418. (b) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2018**, *46*, 8038−8056. (c) Phan, A. T.; Guéron, M.; Leroy, J. L. *J. Mol. Biol.* **2000**, *299*, 123−144.
28. (a) Berger, I.; Egli, M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12116–12121. (b) Lacroix, L.; Mergny, J. L.; Leroy, J. L.; Helene, C. *Biochemistry* **1996**, *35*, 8715–8722.
29. (a) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res*. **2017**, *45*, 2951−2959. (b) Cheng, M.; Chen, J.; Ju, H.; Zhou, J.; Mergny, J.-L. *J. Am. Chem. Soc*. **2021**, *143*, 7792−7807. (c) Sun, D.; Hurley, L. H. *J. Med. Chem*. **2009**, *52*, 2863−2874. (d) Rajendran, A.; Nakano, S.-I.; Sugimoto, *N. Chem. Commun*. **2010**, *46*, 1299−1301. (e) Bhavsar-Jog, Y. P.; Dornshuld, E. V.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. *Biochemistry* **2014**, *53*, 1586−1594.
30. (a) 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. (b) 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. (c) Cheng, M.; Qiu, D.; Tamon, L.; Istvánková,ˇE.; Vísková,ˇP.; Amrane, S.; Guédin, A.; Chen, J.; Lacroix, L.; Ju, H.; Trantírek, L.; Sahakyan, A. B.; Zhou, J.; Mergny, J.-L. *Angew. Chem., Int. Ed*. **2021**, *60*, 10286−10294.
31. (a) Kang, H. J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc*. **2014**, *136*, 4172–4185. (b) Sutherland, C.; Cui, Y. X.; Mao, H. B.; Hurley, L. H. *J. Am. Chem. Soc*. **2016**, *138*, 14138–14151. (c) Niu, K.; Zhang, X.; Deng, H.; Wu, F.; Ren, Y.; Xiang, H.; Zheng, S.; Liu, L.; Huang, L.; Zeng, B. et al. *Nucleic Acids Res*. **2018**, *46*, 1710–1723.
32. (a) Cui, Y.; Koirala, D.; Kang, H. J.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H. *Nucleic Acids Res*. **2014**, *42*, 5755−5764. (b) Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T.-M.; Tan, J.-H.; Huang, Z.-S.; Li, D. *Chem. Commun*. **2018**, *54*, 2036−2039. (c) Brown, S. L.; Kendrick S. *Pharmaceuticals* **2021**, *14*, 96−120. (d) Debnath, M.; Ghosh, S.; Chauhan, A.; Paul, R.; Bhattacharyya, K.; Dash, J. *Chem. Sci*. **2017**, *8*, 7448−7456. (e) Sheng, Q.; Neaverson, J. C.; Mahmoud, T.; Stevenson, C. E. M.; Matthews, S. E.; Waller, Z. A. E. *Org. Biomol. Chem*. **2017**, *15*, 5669−5673. (f). Wright, E. P.; Day, H. A.; Ibrahim, A. M.; Kumar, J.; Boswell, L. J. E.; Huguin, C.; Stevenson, C. E. M.; Pors, K.; Waller, Z. A. E. *Sci. Rep*. **2016**, *6*, 39456.

33. (a) Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Nucleic Acids Res. 2009, 37, 1713–1725. (b) Woody, R. W. *Methods Enzymol.* **1995**, *246*, 34–71. (c) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Methods 2012, 57, 64–75.

34. (a) Jr. Tinoco, I. *J. Am. Chem. Soc.* **1960**, *82*, 4785–4790. (b) Rachwal, P. A.; Fox, K. R. *Methods* **2007**, *43,* 291–301. (c) Rich, A.; Kasha, M. *J. Am. Chem. Soc.* **1960**, *82*, 6197–6199.

35. (a) Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, e138–e144. (b) Evans, L.; Kotar, A.; Valentini, M.; Filloux, A.; Jamshidi, S.; Plavec, J.; Rahman, K. M.; Vilar, R. *RSC Chem. Biol.* **2023**, *4*, 94–100.

36. (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. (c) Yamaoki, Y.; Nagata, T.; Sakamoto, T.; Katahira, M. *Biophys. Rev.* **2020**, *12*, 411–417. (d) Dayie**,** T. K.; Olenginski, L. T.; Taiwo, K. M. *Chem. Rev.* **2022**, *122*, 10, 9357–9394.

37. (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.

38. (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. (c) Krafcikova, M.; Dzatko, S.; Caron, C.; Granzhan, A.; Fiala, R.; Loja, T.; Teulade-Fichou, M.-P.; Fessl, T.; Hänsel-Hertsch, R.; Mergny, J.-L.; Foldynova-Trantirkova**,** S.; Trantirek, L. *J. Am. Chem. Soc.* **2019**, *141*, 13281–13285.

39. (a) Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971−7982. (b) Himmelstoß, M.; Erharter, K.; Renard, E.; Ennifar, E.; Kreutz, C.; Micura, R. *Chem. Sci.* **2020**, *11*, 11322−11330. (c) Li, Q.; Chen, J.; Trajkovski, M.; Zhou, Y.; Fan, C.; Lu, K.; Tang, P.; Su, X.; Plavec, J.; Xi, Z.; Zhou, C. *J. Am. Chem. Soc.* **2020**, *142,* 4739−4748. (d) Baranowski, M. R.; Warminski, M.; Jemielity, J.; Kowalska, J. *Nucleic Acids Res.* **2020**, *48*, 8209−8224. (e) Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. *J. Org. Chem.* **2012**, *77,* 1026−1044. (f) Sakamoto, T.; Hasegawa, D.; Fujimoto, K. *Chem. Commun.* **2015**, *51*, 8749−8752.

40. (a) Manna, S.; Sarkar, D.; Srivatsan, S. G. *J. Am. Chem. Soc.* **2018**, *140*, 12622−12633. (b) Manna, S.; Sontakke, V. A.; Srivatsan, S. G. *ChemBioChem* **2022**, *23,* e202100601.

41 (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. *Analyst* **2017,** *142*, 4083–4088. (c) Bao, H.-

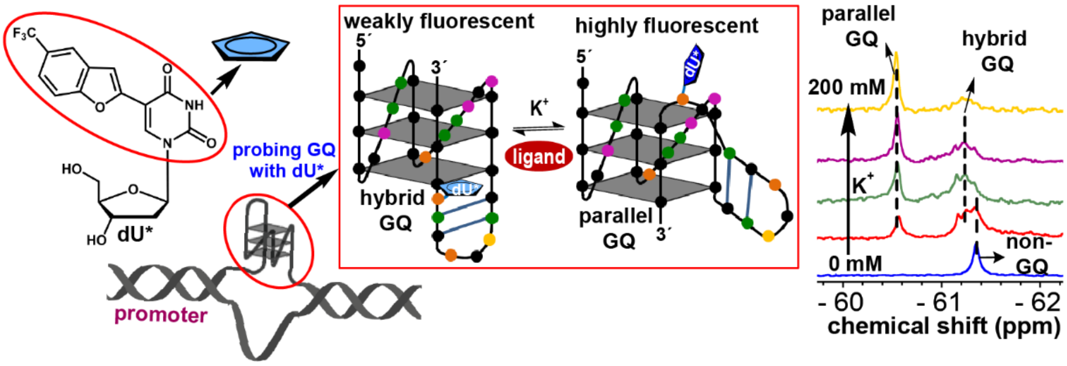
L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. *Nucleic Acids Res*. **2017**, *45*, 5501–5511. (d) Bao, H.-L.; Liu, H.-S.; Xu, Y. *Nucleic Acids Res*. **2019**, *47*, 4940−4947.

42. (a) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struc. Biol*. **2004**, *14*, 350–359. (b) Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.
43. Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M. *Angew. Chem. Int. Ed.* **2009**, *48*, 9728–9730.
44. (a) Stratmann, L. M.; Kutin, Y.; Kasanmascheff, M.; Clever, G. H. *Angew. Chem. Int. Ed*. **2021**, *60*, 4939– 4947. (b) Zhang, X.; Xu, C.-X.; Di Felice, R.; Sponer, J.; Islam, B.; Stadlbauer, P.; Ding, Y.; Mao, L.; Mao, Z.-W.; Qin, P. Z. *Biochemistry* **2016**, *55*, 360–372.
45. Azarkh, M.; Singh, V.; Okle, O.; Dietrich, D. R.; Hartig, J. S.; Drescher, M. *ChemPhysChem* **2012**, *13*, 1444–1447.
46. (a) Mooers, B. H. *Methods* **2009**, *47*, 168–176. (b) Pike, A. C. W.; Garman, E. F.; Krojer, T.; Delft, F.; Carpenter, E. P.; Acta. *Crystallogr. D. Struct. Biol*. **2016**, *72*, 303–318.
47. (a) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880. (b) Geng, Y.; Liu, C.; Zhou, B.; Cai, Q.; Miao, H.; Shi, X.; Xu, N.; You, Y.; Fung, C. P.; Din, R. U.; Zhu, G. *Nucleic Acids Res*. **2019**, *47*, 5395. (c) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. *Nucleic Acids Res*. **2012**, *40*, 4691–700.
48. (a) Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z. *J. Am. Chem. Soc.* **2002**, *124*, 24−25. (b) Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M. *J. Am. Chem. Soc*. **2002**, *124*, 14910–14916. (c) Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Höbartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R.; et al. *Nat. Struct. Mol. Biol*. **2005**, *12*, 218−224. (d) Höbartner, C.; Rieder, R.; Kreutz, C.; Puffer, B.; Lang, K.; Polonskaia, A.; Serganov, A.; Micura, R. *J. Am. Chem. Soc*. **2005**, *127*, 12035−12045.
49. Dauter, M.; Dauter, Z. *Methods Mol. Biol*. **2017**, *1607*, 349–356.
50. (a) Bhasikuttan, A. C.; Mohanty, J. *Chem. Commun*. **2015**, *51*, 7581−7597. (b) Suseela, V.; Narayanaswamy, N.; Pratihar, S.; Govindaraju, T. *Chem. Soc. Rev*. **2018**, *47*, 1098−1131. (c) Yuan, J.-H.; Shao, W.; Chen, S.-B.; Huang, Z.-S.; Tan, J.-H. *Biochem. Biophys. Res. Commun*. **2020**, *531*, 18−24.
51. (a) Maleki, P.; Budhathoki, J. B.; Roy, W. A.; Balci, H. *Mol. Genet. Genomics* **2017**, *292*, 483–498. (b) Fang, J.; Xie, C.; Tao, Y.; Wei, D. *Biochimie* **2023**, *206*, 1−11.
52. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem.* **2013**, *53,* 366–378.

53. (a) Dziuba, D.; Didier, P.; Ciaco, S.; Barth, A.; Seidel, C. A. M.; Mély, Y. *Chem. Soc. Rev.* **2021**, *50*, 7062−7107.
54. (a) Srivatsan, S. G.; Greco, N. J.; Tor, Y. *Angew. Chem. Int. Ed*. **2008**, *47*, 6661–6665. (b) Saitoa, Y.; Hudson, R. H. E. *J. Photochem. Photobiol. C.* **2018**, *36*, 48–73. (c) Schmidt, O. P.; Benz, A. S.; Mata, G; Luedtke, N. W. *Nucleic Acids Res*. **2018**, *46*, 6470–6479. (d) Preus, S.; Kilså, K.; Wilhelmssonb, L. M.; Albinsson, B. *Phys. Chem. Chem. Phys*. **2010**, *12*, 8881–8892. (e) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. *Nucleic Acids Res.* **2008**, *36*, 157–167. (f) Börjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. *J. Am. Chem. Soc.* **2009**, *131*, 4288–4293. (g) Bielecka, P.; Juskowiak, B. *Molecules* **2015**, *20*, 18511–18525.
55. Manna, S.; Srivatsan, S. G. *RSC Adv*. **2018**, *8*, 25673–25694.
56. (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.
57. (a) Pawar, M. G.; Srivatsan, S. G. *Org. Lett*. **2011**, *13*, 1114–1117. (b) Pawar, M. G.; Srivatsan S. G. *J. Phys. Chem. B.* **2013**, *117*, 14273–14282. (c) Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J.* **2011**, *17*, 12820–12827. (d) Tanpure, A. A.; Srivatsan, S. G. *Chembiochem*. **2012**, *13*, 2392–9156. (e) Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res*. **2015**, *43*, e149. (f) Sabale, P. M.; Tanpure, A. A.; Srivatsan, S. G. *Org. Biomol. Chem*. **2018**, *16*, 4141–4150.
58. (a) Zhang, L.; Liu, X.; Lu, S.; Liu, J.; Zhong, S.; Wei, Y.; Bing, T.; Zhang, N.; Shangguan, D. *ACS Appl. Bio. Mater.* **2020**, *3*, 2643–2650. (b) Antonio, M. D.; Ponjavic, A.; Radzevičius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L.-M.; Lee1, S. F.; Klenerman, D.; Balasubramanian, S. *Nat. Chem*. **2020**, *12*, 832–837. (c) 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. (b) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem. Commun*. **2004**, 1438– 1439. (c) Xu, Y.; Sugiyama, H.; *Nucleic Acids Res*. **2006**, *34*, 949–954.
59. (a) Xu, L.; Wang, ab J.; Sun, N.; Liu, M.; Cao, Y.; Wang, Z.; Pei, R. *Chem. Commun*. **2016**, *52*, 14330−14333. (b) Wei, Z.; Liu, B.; Lin, X.; Wang, J.; Huang, Z.-S.; Li, D. *Int. J. Mol. Sci*. **2022**, *23*, 3872−3883.
60. 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.; et al. *Cell Chem. Biol*. **2016**, 23, 1261–1270.

Chapter 2:

Probing juxtaposed G-quadruplex and hairpin motifs using a responsive nucleoside probe: a unique scaffold for chemotherapy



Chapter 2 is a reprint of the data published in the journal:

Also, synthesis and photophysical properties of the modified nucleoside is part of data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.
The thesis author is a main author and researcher for this work.

2.1 Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane protein belonging to the family of protein kinase receptors. The EGFR gene codes for a tyrosine kinase receptor, which is activated by physiological extracellular ligands initiating an important signal transduction pathway that is required for normal cell growth, differentiation and proliferation in mammalian cells.[1,2] Overexpression or mutations that elevate the activity of EGFR signalling pathway is directly linked to the progression of several cancers including that of lung, breast and glioblastoma.[3,4] Presently available therapeutic strategies to counter the effects of upregulation of EGFR activity employ tyrosine kinase inhibitors or monoclonal antibodies that interfere with the binding of ligands to the extracellular receptor domain.[5-7] However, their efficiency is limited by intrinsic or acquired resistance.[7,8] Alternative, we envision that targeting non-canonical nucleic acid structural elements that act as natural regulators could be a viable strategy to control the EGFR expression at replication and transcription levels.[9-11] One such class of structures is a G-quadruplex (GQ), which is formed by guanine rich sequences.[12,13] Four guanine bases in a sequence via Hoogsteen hydrogen bonding form a tetrad (G-tetrad) and two or more G-tetrads, stabilized by monovalent cations (K^+ or Na^+), stack one above the other to form GQ structures.[14] Compelling experimental data indicate that GQs present in DNA and RNA serve as gene regulatory elements.[15,16] Dysfunction of these elements in several genes are linked to tumour progression,[17,18] and hence, small molecule ligands that stabilize these structures and down regulate the gene expression are perceived as alternative therapeutic tools to mitigate cancer.[19–25] Many small molecules ligands developed so far, though show good selectivity between GQ and duplex structures, they seldom distinguish between different topologies of GQs as they have a similar tetrad skeleton.[26] Recently, a few ligands have been developed that bind to a specific GQ topology.[27-32] Nevertheless, targeting a specific GQ motif amongst others in the genome remains a major challenge. Alternatively, G-rich segments that harbour GQ-duplex or GQ-hairpin junctions are considered as druggable targets,[33,34] wherein ligand scaffolds capable of simultaneously binding to GQ and duplex regions increase the specific targeting of such motifs.[35-37] In the context of EGFR gene, a 30 nucleotide G-rich sequence upstream of the transcription start site (-272 position) forms two unique GQ topologies (parallel and hybrid-type), which are stabilized by a short hairpin located at the terminal loop.[38] This domain represents a new point of intervention to potentially attenuate the disease-causing activity of the gene. Therefore, it is important to first understand (i) the structural polymorphism of EGFR GQs, (ii) how the GQs interact with small molecules ligands and (iii) the functional role of GQs in cellular process (e.g., DNA replication).

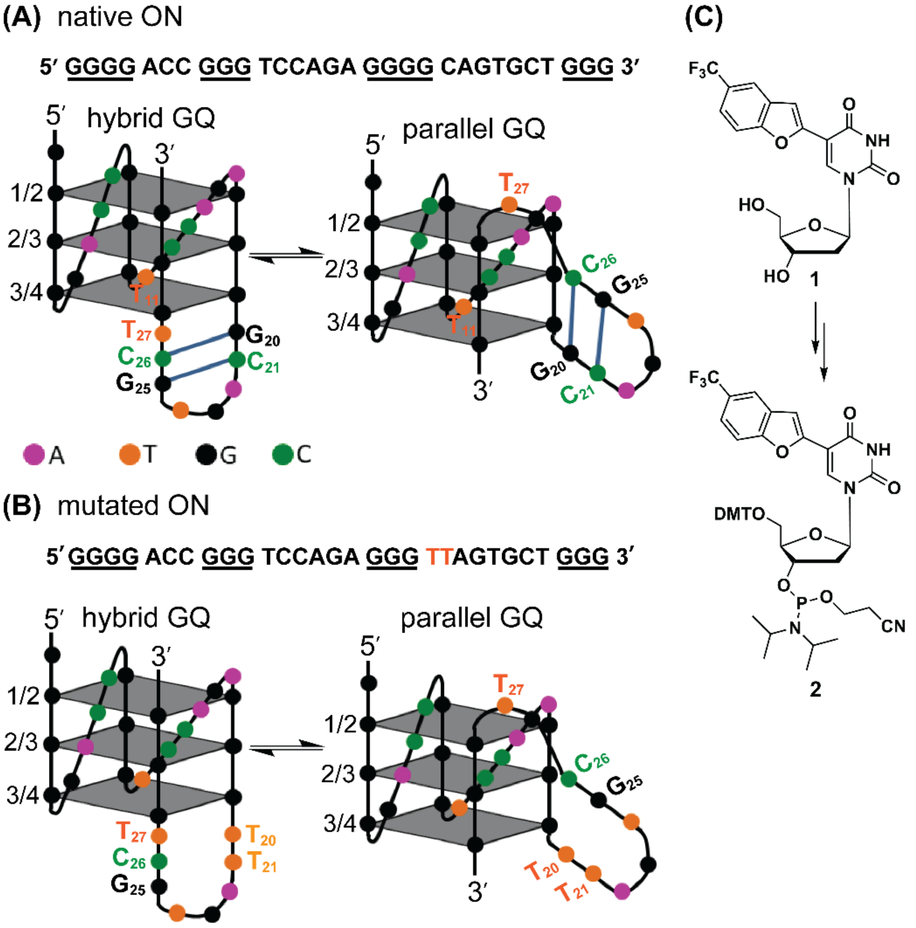


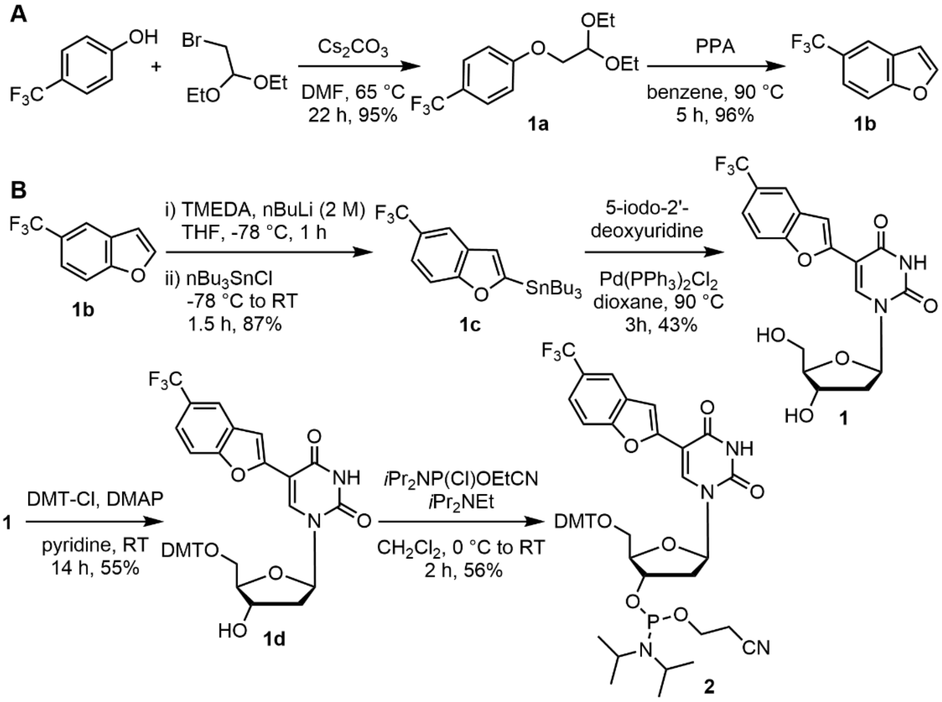
Figure 1. (**A**) EGFR-272 G-rich sequence folds into a mixture of hybrid and parallel GQ topologies with a hairpin structure in their third loop.[38] (**B**) Mutations (G_{20} and C_{21} to T_{20} and T_{21}) in the EGFR G-rich ON abolishes the formation of the hairpin structure in both the GQ topologies. (**C**) Chemical structure of modified nucleoside TFBF-dU (**1**) and its corresponding phosphoramidite (**2**) used in the synthesis of labeled ONs.

GQ exhibits a high degree of structural polymorphism, which apart from environmental conditions (e.g., ionic conditions, molecular crowding and confinement) depends on the sequence composition namely, number of G-tracts and loop nucleotides that connect them.[39,40] G-tracts composed of four or more guanine bases can form multiple dynamically interchangeable GQ conformations.[41-43] EGFR promoter region contains four G-tracts (4-3-4-3) that can support multiple GQ forms as guanine bases can pair differently (Figure 1A). As deduced from ^{1}H NMR and CD analysis, this sequence adopts two main intramolecular folded structures, namely hybrid-type and parallel GQs.[38] The third loop forms a hairpin junction in both the forms as a result of base-pairing between G_{20}-C_{26} and C_{21}-G_{25} (Figure 1A). Notably, the hairpin structure positively contributes to GQs stability. Therefore, molecular scaffolds that can target GQ and proximal hairpin simultaneously are envisioned to improve the druggability score of this new target.

Formation of GQ structures, and their stability and interaction with small molecule ligands are commonly studied *in vitro* by CD, UV-thermal melting, fluorescence, NMR and X-ray techniques using unlabelled oligonucleotides (ONs) or probe-labeled ONs.[44-47] Alternatively, fluorescent ligands, which show changes in spectral properties upon binding to GQ structures serve as useful sensors.[26,48,49] Notably, GQ-specific antibodies and fluorescent ligands have also been developed to visualize GQs in cellular environment.[50-53] However, when multiple GQ species are present, rarely these methods and probes provide useful information on the individual topologies and their structural equilibrium as they fail to distinguishing individual topologies. In this context, single-molecule approaches using force-based (magnetic tweezer and optical tweezer) and fluorescence-based (smFRET) techniques have been somewhat successful,[54-57] but these experimentations require sophisticated instrument setup. Importantly, these methods cannot be easily used to detect lowly populated GQs and extended to cell-based analysis.

Recently, we developed a highly environment-sensitive nucleoside analog by conjugating fluorobenzofuran moiety at the C5 position of 2'-deoxyuridine (FBF-dU).[58] The heterocycle modification imparts fluorescence and endows ^{19}F NMR label, thereby allowing a two-channel spectroscopic analysis of the GQ structures formed by the human telomeric overhang. While this analog is highly useful, assay conditions require very high concentrations of the labeled oligonucleotides (ONs, ~200 µM) and is not suitable for detecting lowly populated species coexisting in a dynamic equilibrium. In order to amplify the signal without compromising the sensitivity of the original probe system, we designed the second generation probe by conjugating trifluoromethyl benzofuran at the C5 position of 2'-deoxyuridine (Figure 1C, TFBF-dU **1**). Given the ability of EGFR promoter to adopt unique GQ architectures, whose structural polymorphism could be perturbed by environmental conditions (e.g., ionic conditions) and ligand binding, we decided to harness the true potential of TFBF-dU in studying this biologically important system in detail to validate its therapeutic potential.

Here, we report the design and synthesis of a new microenvironment-sensitive dual-app nucleoside analogue (Scheme 1) to probe GQ structures adopted by the wild-type and mutated EGFR promoter region, their population equilibrium and the influence of hairpin junction in driving GQ formation. TFBF-dU incorporated into EGFR G-rich sequences provides distinct spectral readouts for the individual GQ structures, thereby enabling the quantification of their relative population under different conditions. The probe helps in identifying the preferred GQ topology in the absence and presence of small molecule ligands under intracellular ionic conditions. The utility of the probe in determining the GQ structure adopted by the ON in a cell-like environment is also demonstrated by performing ^{19}F NMR in frog egg lysate and extract. Further, using polymerase stop assay we ascertained the implication of GQs and GQs bound to ligands in the DNA replication process.



Scheme 1. Synthesis of (**A**) 5-trifluoromethyl-benzofuran **1b** and (**B**) 5-(5-Trifluoromethyl-benzofuran-2- yl)-2′-deoxyuridine **1** and its corresponding phosphoramidite **2**. DMF: dimethylformamide, PPA: polyphosphoric acid, TMEDA: N,N,N′,N′-tetramethylethylenediamine; THF: tetrahydrofuran; DMT-Cl = 4,4′-dimethoxytrityl chloride; DMAP: 4-dimethylaminopyridine.

2.2 Results and Discussion

2.2.1 Design, Synthesis, and Environment-Sensitivity of Nucleoside Probe 1

Most nucleic acid studies employ ONs labeled with a probe that is suitable only for a given technique, which seldom gives a comprehensive understanding of the study system. Therefore, we envisioned to integrate two powerful readouts in a single nucleoside probe, namely, fluorescence and ^{19}F NMR, for the following reasons. Conjugating a heterocycle to a nucleobase can produce environment-sensitive fluorescent nucleoside analogues.[59–62] A few such analogues serve as good probes to investigate noncanonical nucleic acid structures, especially GQs.[63–65] On the other hand, the ^{19}F atom, due to high natural abundance, sensitivity to the environment, and "zero" background signal from cells,[66] is emerging as a very useful biophysical label to probe nucleic acid secondary structures and their interaction with ligands and proteins.[66–69] We combined these two features and developed fluorobenzofuran-modified nucleosides, which served as good tools to probe GQs of human telomeric overhang.[58] However, this modification is not ideally suited for capturing lowly populated structures as it requires high concentrations of the ON sample. Indeed, as demonstrated in the present study, a simple improvisation by conjugating the trifluoromethyl-benzofuran-2-yl (TFBF) core at the C5 position of 2′-deoxyuridine rewardingly produced a highly sensitive fluorescent nucleoside (**1**) with three chemically and magnetically equivalent ^{19}F atoms (Figure 1C). The heterocycle unit, 5-trifluoromethyl-benzofuran **1b**, was synthesized by using the steps shown in Scheme 1A. **1b** was then stannylated and reacted with 5-iodo-2′-deoxyuridine under Stille cross-coupling reaction conditions using Pd- $(PPh_3)_2Cl_2$ as a catalyst to afford the nucleoside analogue, 5- (5-trifluoromethyl-benzofuran-2-yl)-2′-deoxyuridine **1**, in moderate yields (Scheme 1B).

The lowest energy absorption maximum of nucleoside **1** exhibits hyperchromism as the solvent polarity is reduced from water to methanol to dioxane (Figure 2A, Table 1). The dipole moment of a nucleoside in a ground state and excited state could depend on the solvent polarity and viscosity. Generally, polar molecules (like nucleoside) exhibit less dipole moment in ground state in non-polar solvents and hence would require higher energy for excitation. Hence, TFBF-dU could display hyperchromic effect with decrease in the solvent polarity.[69b] In the fluorescence study, excitation of an aqueous solution of the analogue **1** displays an intense emission band centered at 427 nm and a quantum yield (Φ) of 0.05 (Figure 2A, Table 1). In less polar solvents, the fluorescence intensity and emission maximum (λem) are significantly quenched and blue-shifted, respectively. As nucleoside (polar molecule) would exhibit higher dipole moment in the excited state, hence, higher stoke shift was observed in a more polar solvent.[69b] To begin with, the nucleoside analogue was designed to have a molecular rotor element, that is, a rotatable bond between the TFBF moiety and nucleobase. If this element is active, then the fluorescence and ^{19}F NMR signals could be sensitive to the relative orientation of these rings about the rotatable bond.[70,71] To validate this notion, photophysical properties were determined in solvents (ethylene glycol and glycerol) of different viscosity but with a similar polarity. The λ_{em} of the nucleoside is very similar (~410 nm), but there is a significant enhancement in quantum yield (~2.5-fold) when the medium is changed from ethylene glycol to glycerol (Figure 2B). Furthermore, between water and glycerol there is a more than 8-fold increase in quantum yield indicating the presence of a molecular rotor element, which is rigidified in a more viscous medium. Higher fluorescence anisotropy in more viscous solvents also confirms the presence of a molecular rotor element in the nucleoside analogue (Table 1).

Time-resolved fluorescence analysis also reveals distinct excited-state lifetimes in different solvents wherein longer lifetimes result in higher quantum yields (Figure 3, Table 1).

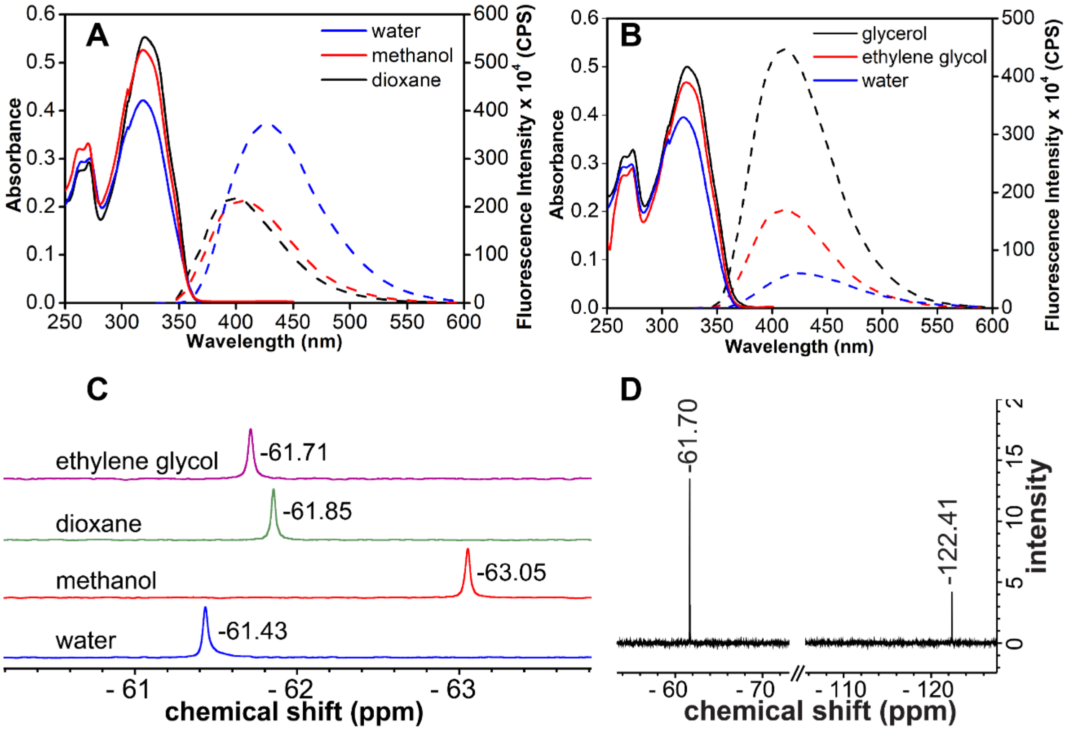


Figure 2. (**A** and **B**) UV absorption (25 µM, solid lines) and fluorescence (5 µM, dashed line) spectra of TFBF-nucleoside **1** in solvents of different polarity and viscosity, respectively. In the fluorescence study, samples were excited at respective lowest energy absorption maximum (Table 1). (**C**) ^{19}F NMR spectra of nucleoside **1** in different solvents. (**D**) A comparison of peak intensity of **1** (-61.7 ppm) and fluorobenzofuran-modified 2'-deoxyurindine (-122.4 ppm).

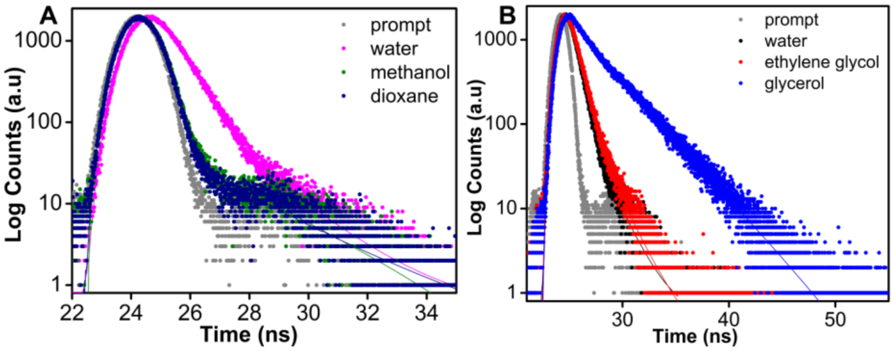


Figure 3. Fluorescence-decay profile of nucleoside **1** (5 µM) in solvents of different polarity (water, methanol, dioxane) and viscosity (water, ethylene glycol, glycerol). Instrument response (prompt) is shown in grey dots and curve fits are shown in solid lines.

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

Solvent	λ_{max}[a] (nm)	λ_{em} (nm)	Φ[b]	τ_{av}[b] (ns)	r[b]
water	319	427	0.05	0.72	0.07
methanol	319	405	0.02	0.27	n.d.
dioxane	320	400	0.02	0.23	n.d.
ethylene glycol	321	409	0.16	0.86	0.25
glycerol	322	410	0.42	2.73	0.35

[a]Wavelength given corresponds to the lowest energy absorption maximum. [b]Standard deviations for quantum yield (Φ), average lifetime (τ_{av}), and anisotropy (r) in different solvents are ≤ 0.002, ≤ 0.02 ns, and ≤ 0.001, respectively. n.d. = not determined.

The alignment of molecular dipoles in a medium of different polarity can either shield or deshield the ^{19}F atom.[71,72] Furthermore, the electron density on ^{19}F atoms will be different depending on the relative orientation of the heterocyclic ring with respect to the uracil ring. Much like fluorescence, distinct ^{19}F chemical shifts are exhibited by the analogue in solvents of different polarity and viscosity (Figure 2C). Notably, in comparison to fluorobenzofuran-modified 2′- deoxyurindine, TFBF-modified 2′-deoxyurindine **1** is found to be significantly more sensitive producing a 3-fold intense peak (Figure 2D). Collectively, these studies indicate that our nucleoside analogue incorporated into ON sequences can be used to probe their conformations simultaneously by fluorescence and ^{19}F NMR techniques.

Table 2 TFBF-dU modified and respective control unmodified telomeric repeat and EGFR ONs.

ON	5'------------sequence--------- 3'[a]
Telo1	TAGGGTTAGGGTTAGGGTTAGGGTT
Telo2	TAGGGTTAGGG**T***TAGGGTTAGGGTT
TeloC	CCCTAACCCTAACCCTAACCCT
3	GGGGACCGGG**T***CCAGAGGGGCAGTGCTGGG
4	GGGGACCGGGTCCAGAGGGGCAGTGC**T***GGG
5	GGGGACCGGGTCCAGAGGG**TT**AGTGC**T***GGG
6	GGGGACCGGGTCCAGAGGGGCAGTGCTGGG
7	GGGGACCGGGTCCAGAGGG**TT**AGTGCTGGG

[a]**T*** represents modified nucleoside **1**. **T** represents mutation points.

2.2.2 Synthesis of TFBF-dU-labeled ONs to detect GQs

To gain insights into the GQ folding dynamics and interaction of GQ structures with small molecule ligands, we decided to use TFBF-dU **1**. The important consideration we took into account is the placement of TFBF-dU in GQ-forming sequences. In comparison to G-tetrads, the loop residues that connect G-tetrads exhibit differences in conformation as well as interactions with adjacent bases in different GQ topologies (Figure 1A). Hence, replacing a loop dT residue with TFBF-dU would have a very minimum impact on the GQ formation and also would provide distinct spectral signatures for different topologies by the virtue of its ability to sense microenvironment changes (Figure 1A). Before incorporating into the EGRF sequence, which forms multiple GQs, we first studied the use of the nucleoside analog by employing a hybrid-type GQ formed by a human telomeric DNA repeat ON sequence (Telo1, Table 2).[73,74] For this purpose, the modified ON Telo2 in which one of the loop dTs is replaced with the nucleoside analog was synthesized by solid-phase ON synthesis method using phosphoramidite **2** (Figure 1C, Table 2). Similarly, in EGFR ONs **3** and **4**, the nucleoside analog **1** was placed in the second loop (T_{11}) and third loop (T_{27}), respectively (Figure 1A, Table 2). In addition, a mutated ON **5** (G_{20} and C_{21} to T) with the modification at T_{27} was synthesized to probe the role of hairpin in the formation of GQs (Figure 1B). All the ONs were purified by PAGE under denaturing conditions, and their purity and identity were confirmed by RP-HPLC and mass analyses, respectively (Figure 4–6, Table 3).

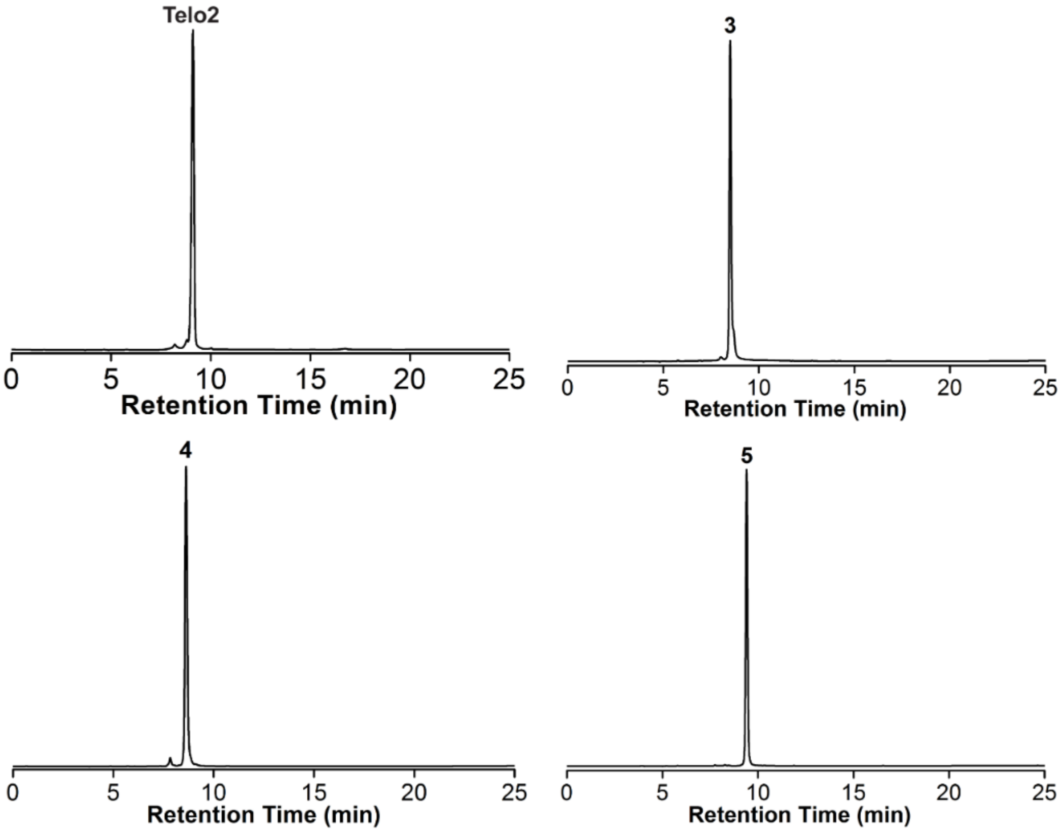


Figure 4. RP-HPLC chromatograms of TFBF-modified Telo2 and EGFR ONs **3**–**5** analyzed at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).

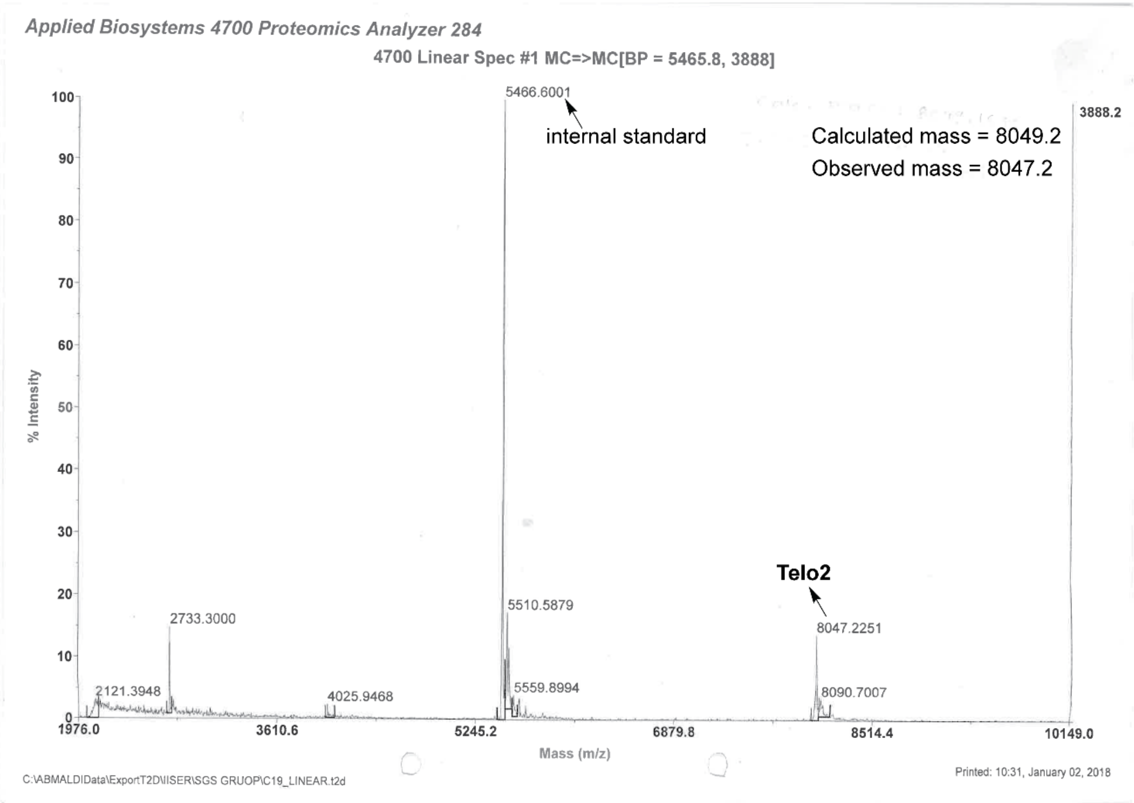
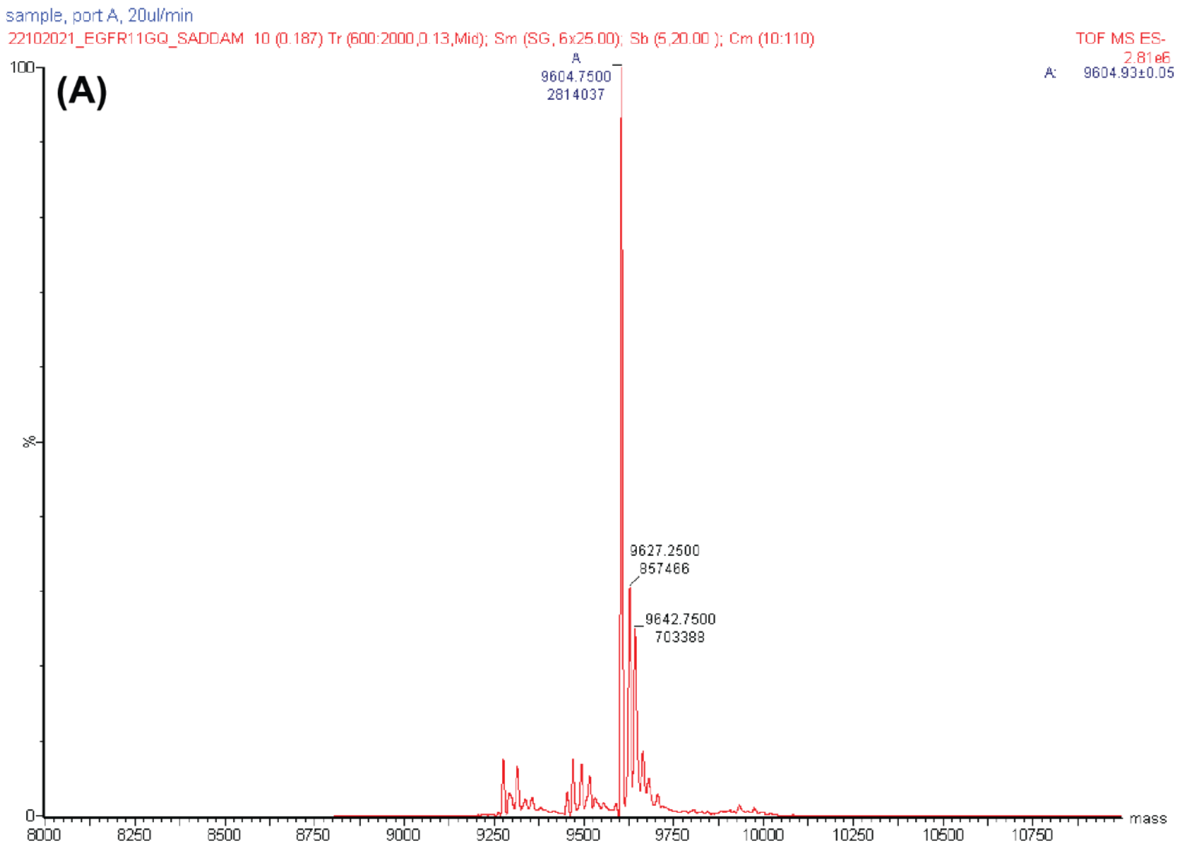


Figure 5. MALDI-TOF MS spectrum of TFBF-modified Telo2. Internal DNA ON standard m/z of +1 and +2 ions are 5466.6 and 2733.3.



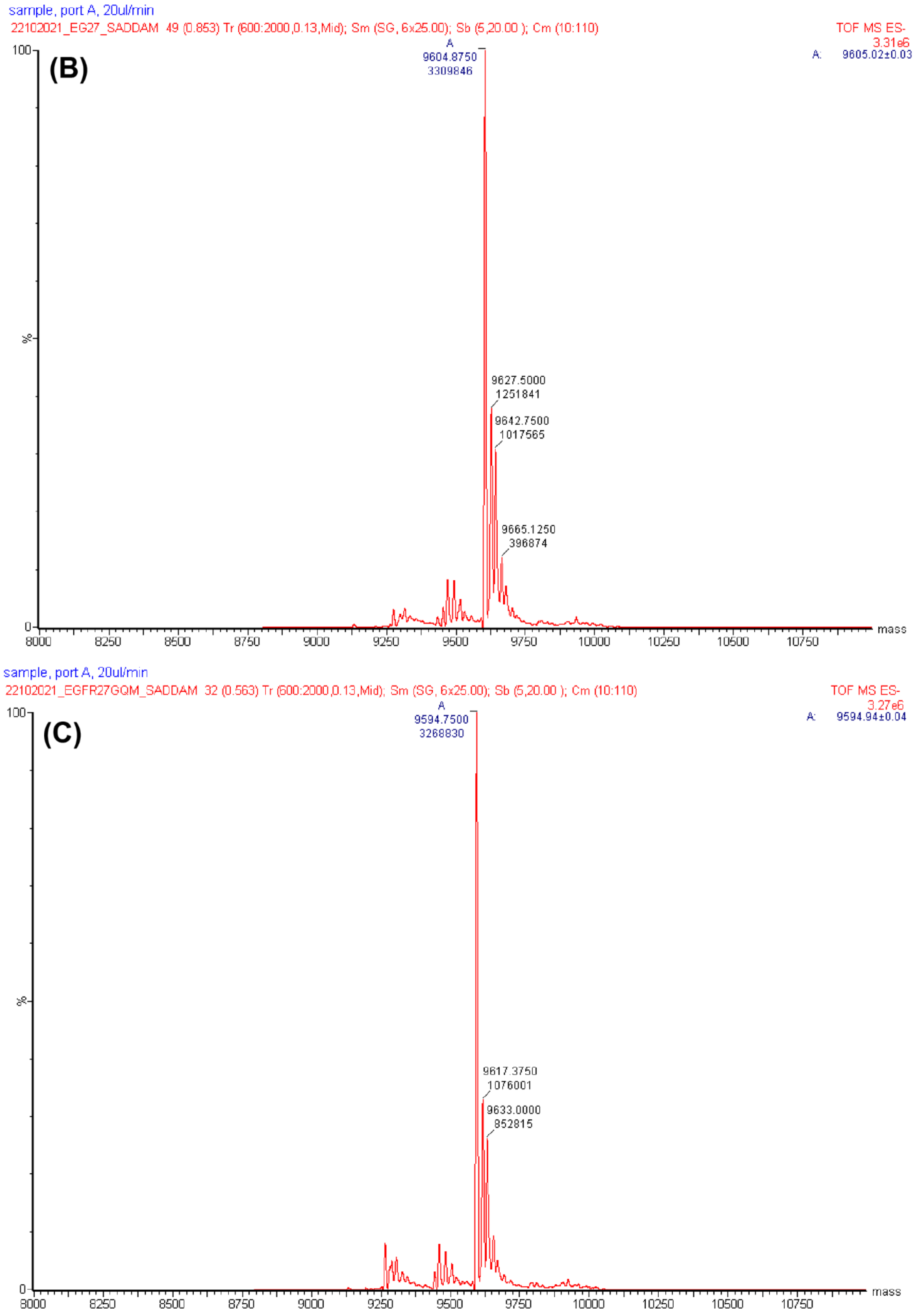


Figure 6. ESI-MS spectra of TFBF-modified EGFR ONs (**A**) **3**, (**B**) **4**, (**C**) **5**.

Table 3 Molar absorptivity and mass of modified DNA ONs.

DNA ON	$\mathcal{E}_{260}$ ^{a} [M^{-1} cm^{-1}]	Calculated mass	Observed mass
Telo2	256 x 10^3	8049.2	8047.2
3	295 x 10^3	9605.1	9604.8
4	295 x 10^3	9605.1	9604.9
5	297 x 10^3	9595.1	9594.8

[a] Molar absorption coefficient ($\mathcal{E}_{260}$) of the modified ONs was determined by using Oligo Analyzer 3.1. $\mathcal{E}_{260}$ of modified nucleoside **1** ($\mathcal{E}_{260}$ = 11.4 x 10^3 M^{-1} cm^{-1}) was used in the place of thymidine.

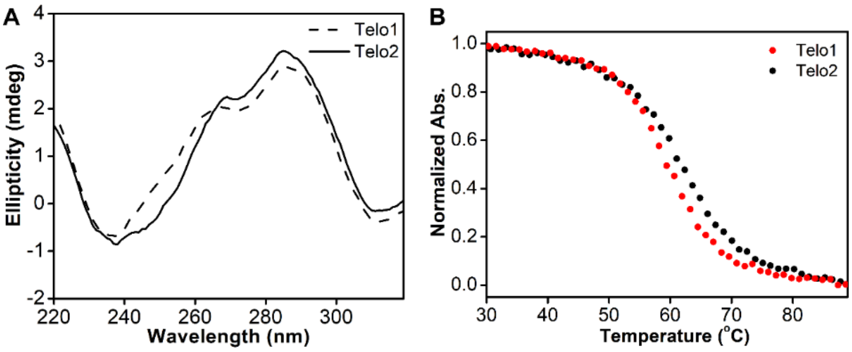


Figure 7. (**A**) CD spectra of control Telo1 and modified Telo2 ONs (5 µM). (**B**) UV-thermal melting profile (at 295 nm) of Telo1 and Telo2 ONs (1 µM).

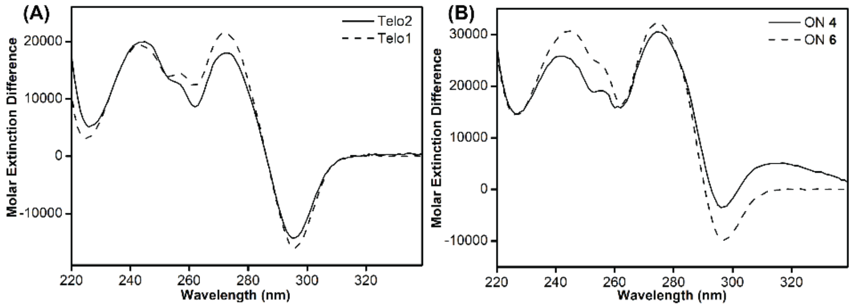


Figure 8. Thermal difference spectrum (TDS) for Telo2 (**A**) and EGFR ON **4** (**B**), and corresponding unmodified control ON sequences. ONs samples (5 µM) were annealed in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and UV spectrum of ONs was recorded at 25 °C and 90 °C. The TDS was obtained by subtracting the UV absorption profile of the folded form from the unfolded form.

2.2.3 Impact of TFBF-dU labeling on the formation of GQs

Native Telo1 and modified Telo2 ONs were annealed in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and the formation of the hybrid-type GQ structure and its stability was studied by CD and UV-thermal melting analysis. CD profiles of both control and modified ONs gave a positive band at ~285 nm, a prominent shoulder at ~269 nm and a negative band at ~240 nm characteristic of a hybrid-type GQ (Figure 7A).[75] The T_m values for the ONs were also found to be very similar (Figure 7B and Table 4). Further, a thermal difference spectrum (TDS) obtained by subtracting the UV absorption spectrum of the folded state from the unfolded state gave a profile corresponding to a GQ structure (Figure 8).[76,77] Together, these results suggest that the modification has negligible impact on the structure and stability of Telo GQ.

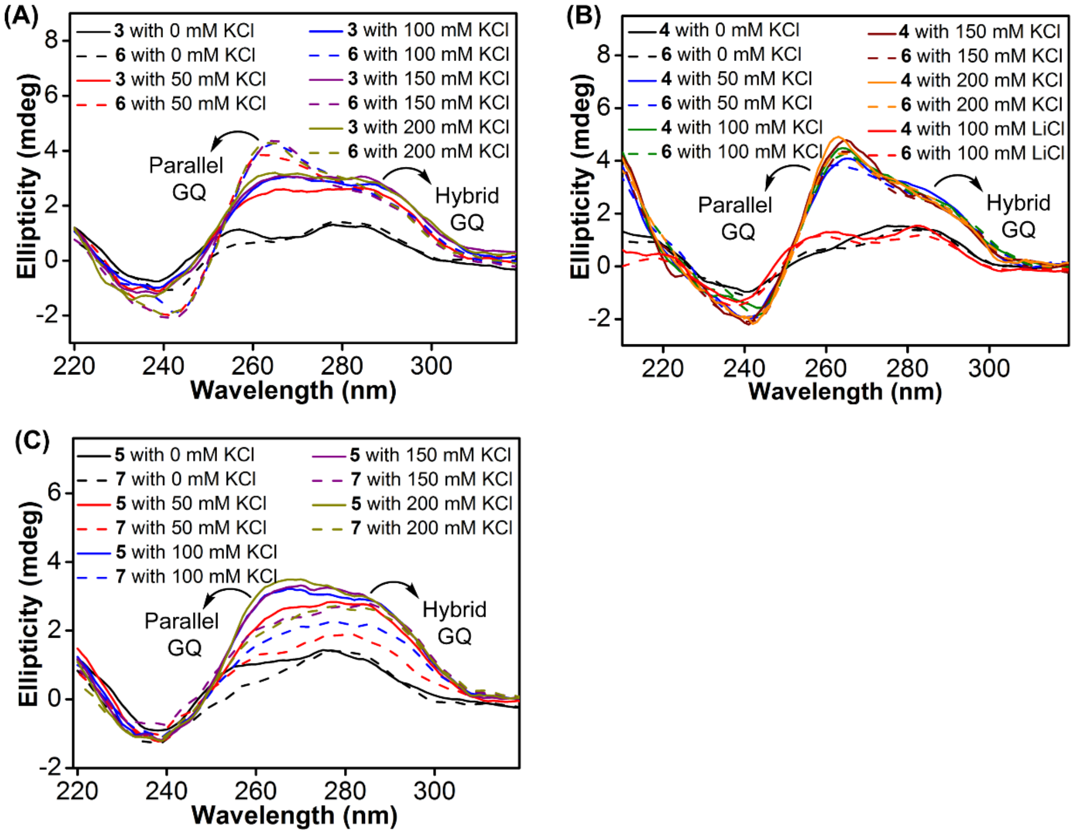


Figure 9. (**A**, **B** and **C**) CD spectra (5 µM) of modified EGFR ONs **3**–**5** (solid lines) and their control ONs **6** and **7** (dashed lines) in different ionic conditions.

Similarly, EGFR ONs **3**–**5** and control unmodified ONs **6** and **7** were annealed in 10 mM Tris.HCl buffer (pH 7.4) and the formation of GQ structures with an increase in K$^+$ ion concentration was monitored by CD spectroscopy. In the absence of KCl, EGFR ONs did not show a CD profile corresponding to a GQ structure (Figure 9). In the presence of K$^+$ ions (100 mM), control ON **6** exhibited a positive peak at 265 nm and a negative peak at 240 nm indicating the formation of a parallel GQ (Figure 9A and 9B, dashed lines). In addition, a prominent shoulder at 285 nm representing a hybrid-type GQ form was observed irrespective of K$^+$ ion concentration (Figure 9A and 9B).[38,78] Modified ON **3** gave significantly less intense peaks at +265 nm and -240 nm as compared to the native ON, suggesting that the modification at position T$_{11}$ hampered GQ formation (Figure 9A, red solid line). On the other hand, ON **4** displayed a CD profile (peak at +265 nm, -240 nm, and a shoulder at 285 nm) similar to control ON **6** (Figure 9B, red lines). This confirms that nucleoside **1** placed at T$_{27}$ in ON **4** has negligible impact on the GQ structure formation. Mutations in the stem of the hairpin affected

GQ formation of ONs **5** and **7** with a more pronounced effect on the parallel topology (Figure 9C). Further, we observed that labeled ON **4** and control ON **6** exhibited a similar profile in all K^+ ion concentrations tested (50 mM to 200 mM, Figure 9A). Although the parallel GQ structure seemed to be higher in amounts as compared to the hybrid form, CD profiles suggested that these structures coexist with no apparent change in the structural equilibrium as a function of KCl concentration. UV-thermal melting analysis indicate that a TFBF-dU modification at T_{11} and T_{27} positions only marginally affects the stability of GQs as compared to unmodified control ONs (Figure 10 and Table 4). However, CD data indicates that a modification at T_{11} (ON **3**) affects the GQ equilibrium, and hence for further analysis native ON **4** and mutated ON **5** containing the modification at T_{27} were used.

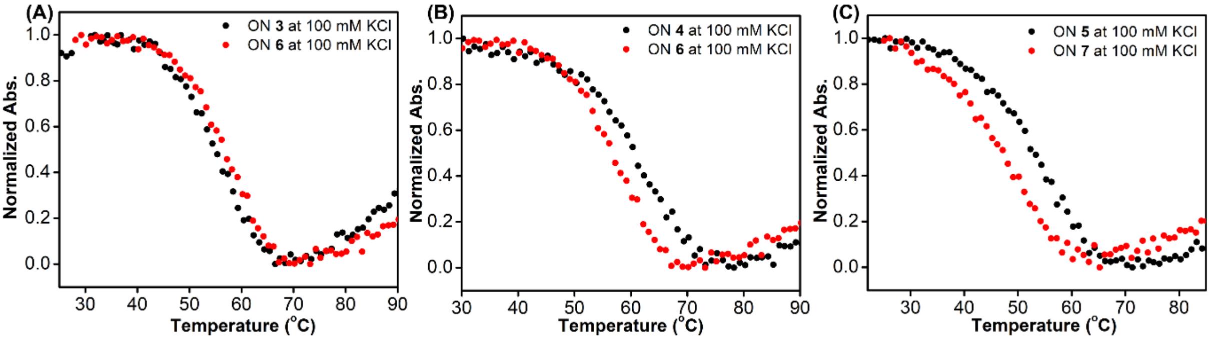


Figure 10. UV-thermal melting profile (at 295 nm) of modified EGFR ONs **3**–**5** and corresponding control unmodified ONs **6** and **7** (1 µM).

Table 4 T_m values of modified Telo2 and EGFR ONs **3**–**5** and control unmodified Telo1 and EGFR ONs **6** and **7.**

KCl	T_m (°C)						
	Telo1	Telo2	ON **3**	ON **4**	ON **5**	ON **6**	ON **7**
100 mM	60.8 ± 0.6	60.2 ± 0.6	53.0 ± 0.6	56.3 ± 1.0	51.7 ± 1.2	54.8 ± 1.2	48.1 ± 1.0

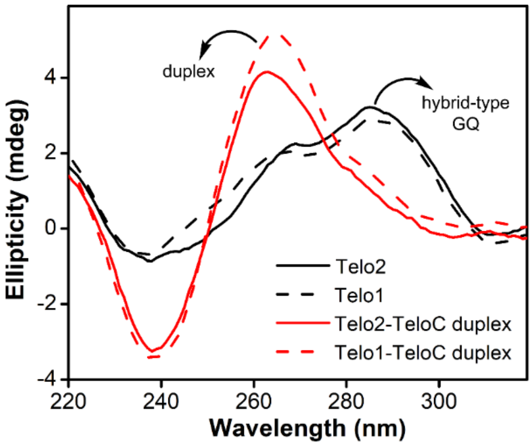


Figure 11. A comparison of CD spectra of duplex and GQ structures formed by the telomeric repeat ONs (5 µM).

2.2.4 Probing the GQ structure of the human telomeric repeat by fluorescence and ^{19}F NMR

TFBF-dU modified Telo2 ON predominantly forms a mixed parallel-antiparallel stranded hybrid-type 2 GQ structure like the control Telo1 ON (Figure 11).[73,74] Upon excitation at 330 nm, the GQ form of Telo2 exhibited a fluorescence band with an emission maximum at ~423 nm (Figure 12A). When Telo2 was hybridized to its complementary C-rich strand (TeloC), it formed a duplex structure (Figure 11), which exhibited a very weak fluorescence band. The observed fluorescence outcome in these structures is due to the difference in the microenvironment around the probe. It is to be noted that several probes display significant quenching in fluorescence due to stacking interaction with neighbouring bases and when located near a guanine base.[79,80] TFBF-dU placed at T_{12} position upon duplex formation would base pair and partially stack with adjacent bases namely, G_{11} and T_{13}. As a result of stacking interaction and closeness to the guanine base, the probe shows a quenched emission band. Based on the NMR structure of the hybrid 2 GQ form (PDB: 2JPZ),[74] the TFBF modification at C5-position would be projected into the groove away from the G-tetrad formed by the G_{11} residue (Figure 13). Possibly due to reduced stacking interaction the GQ form shows higher fluorescence intensity compared to the duplex structure.[80] Much like the fluorescent component, the fluorine label was also found to be conformation sensitive. TFBF-dU gave a distinct ^{19}F NMR signal for GQ and duplex structures (Figure 12B). The formation of these structures was also confirmed by imino protons signal appearing in GQ and duplex regions (Figure 13). Collective, these results indicate that our probe is structurally non-invasive and highly responsive to nucleic acid conformations, and hence, we embarked on using TFBF-dU in the analysis of EGFR GQs.

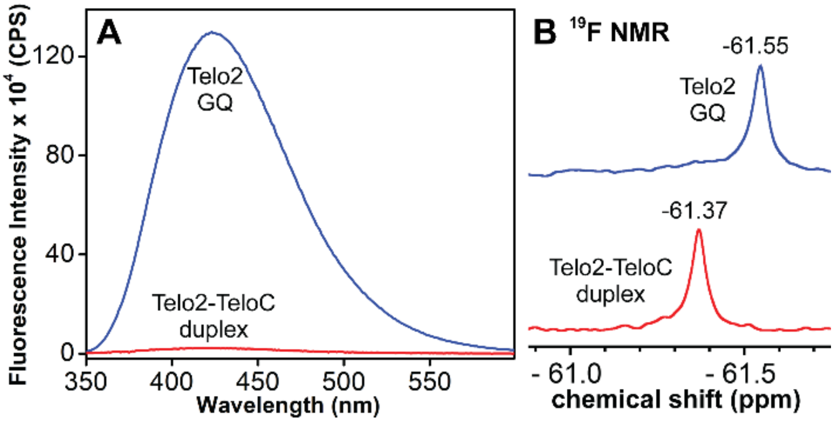


Figure 12. (**A**) Fluorescence and (**B**) ^{19}F NMR spectra of telomeric GQ and duplex structures.

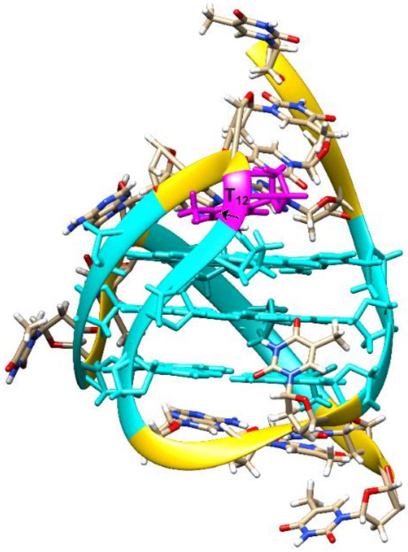


Figure 13. NMR structure of a hybrid GQ topology of the human telomeric repeat (PDB: 2JPZ).[74] Guanosines participating in the tetrad formation are shown in cyan. T_{12} residue is shown in magenta. Arrow shows the C5-postion of T_{12} where the TFBF heterocycle modification is placed in the modified Telo2 ON. TFBF modification does not affect the GQ structure. Hence, TFBF modification at C5-position is likely to be projected into the groove away from the G-tetrad. As a result of reduced stacking interaction, the GQ form shows higher fluorescence intensity. The figure was generated using UCSF Chimera version 1.15.

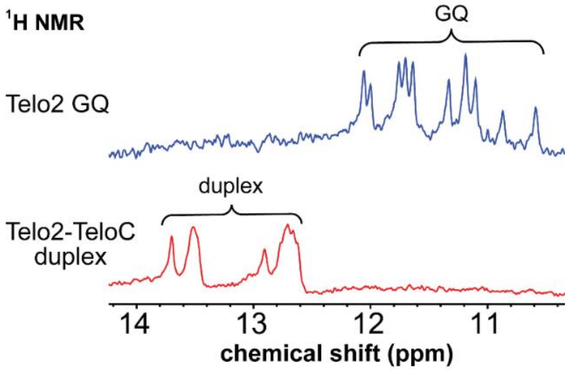


Figure 14. ^{1}H NMR spectra of the modified telomeric ON forming GQ and duplex structures.

2.2.5 Probing EGFR GQs by fluorescence

Samples of EGFR ON **4** annealed in Tris.HCl buffer (pH 7.4) containing 100 mM LiCl or varying concentrations of K^+ ions were excited at 330 nm and changes in steady-state emission were recorded. The unstructured ON in the presence of LiCl or in the absence of KCl (see CD profiles, Figure 9B) exhibited a very weak fluorescence band centered around 419 nm (Figure 15A). As the concentration of K^+ ions was increased (25–200 mM), a progressive increase in the fluorescence intensity with a minor shift in emission maximum was observed as a result of formation of GQs. The fluorescence intensity of ON **4** saturated at 150 mM of KCl (~8-fold increase as compared to non-GQ form). In contrast, CD profile was very similar at all K^+ ion concentrations tested suggesting that there is no apparent change in the structural equilibrium

(Figure 9B). To further probe the structures formed by EGFR ON **4**, we performed time-resolved fluorescence analysis (Figure 15B). In the absence of KCl, three decay components corresponding to lifetimes τ_1 = 1.07 ns, τ_2 = 2.99 ns, and τ_3 = 0.11 ns were observed (Table 5). This is consistent with the literature report suggesting the possibility of three hairpin structures co-existing at zero concentration of KCl.[38] At 25 mM KCl, we observed three different components with lifetimes τ_1 = 1.44 ns, τ_2 = 4.70 ns, and τ_3 = 0.14 ns. The individual lifetime components could correspond to random coil, parallel GQ and hybrid-type GQ structures in no particular order. The longest lifetime component (τ_2 = 4.70 ns) was found to be significantly more populated as compared to the other two fast decay components (τ_1 and τ_3). Interestingly, as we increased the concentration of KCl (25 mM to 200 mM), the population of the species corresponding to τ_2 increased from 63% to 80% with a concomitant decrease in the population of τ_1 and τ_3 (Table 5). Lifetime analysis also indicated a saturation point at 150 mM KCl similar to steady-state analysis. It is to be noted that the fluorescence of the free nucleoside analog was not significantly affected by changes in the K^+ ion concentration, indicating that the observed fluorescence is due to differences in the microenvironment of the nucleoside analog in different conformations adopted by the ON (Figure 16A).

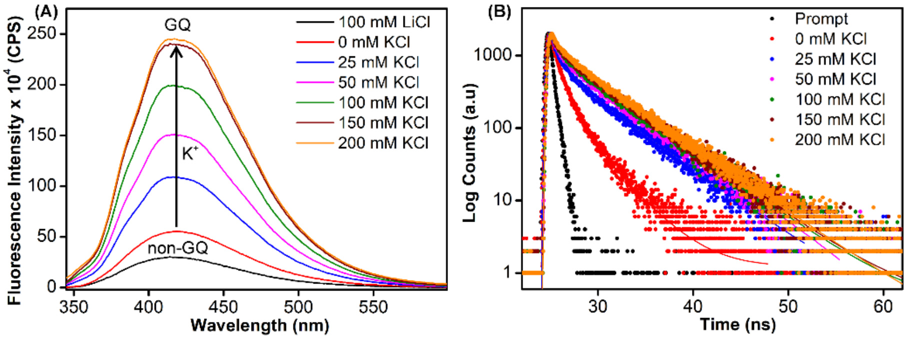


Figure 15. (**A**) Steady-state and (**B**) time-resolved fluorescence spectra of modified ON **4** at different KCl concentrations. For steady-state fluorescence, samples (1 µM) were excited at 330 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively. Instrument response (prompt) is shown in grey dots and decay curve fits are shown in solid lines.

Table 5 Fluorescence lifetimes of TFBF-modified ON **4**.

KCl (mM)	τ_1 (ns)	a_1[a]	τ_2 (ns)	a_2[a]	τ_3 (ns)	a_3[a]
0	1.07 ± 0.03	37%	2.99 ± 0.05	21%	0.11 ± 0.003	42%
25	1.44 ± 0.14	15%	4.70 ± 0.10	63%	0.14 ± 0.02	22%
50	1.29 ± 0.20	11%	4.60 ± 0.12	72%	0.11 ± 0.01	17%
100	1.35 ± 0.07	8%	4.59 ± 0.10	77%	0.13 ± 0.02	14%
150	1.27 ± 0.22	9%	4.51 ± 0.02	79%	0.12 ± 0.01	13%
200	1.38 ± 0.10	8%	4.56 ± 0.001	80%	0.13 ± 0.01	13%

[a]Relative amplitude (a_1, a_2, a_3) corresponding to lifetime components (τ_1, τ_2, τ_3).

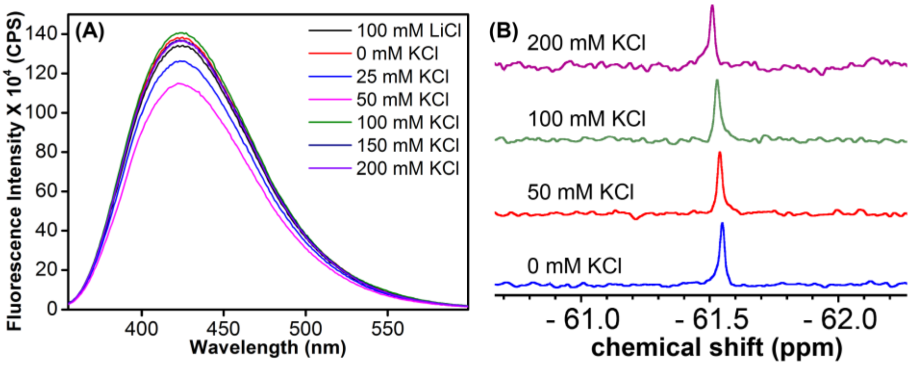
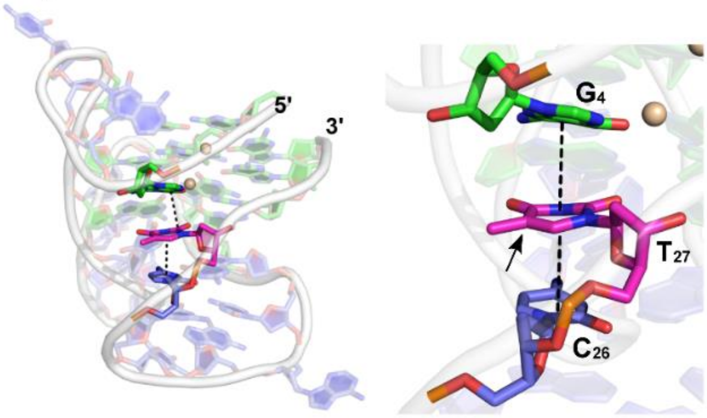


Figure 16. (**A**) Fluorescence spectra of TFBF-dU (2 µM), and (**B**) ^{19}F-NMR spectra of TFBF-dU (10 µM) at different KCl concentrations. In the fluorescence study, samples were excited at 320 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively.

Based on the fluorescence and CD data, and by comparing the structural models predicted for EGFR GQs in an earlier report,[38] we infer the following. CD profile of ON **4** in the presence of KCl (200 mM) shows signatures for a parallel and hybrid-type GQs with a ratio of 70:30.[38] The trend was consistent with the fluorescence data, and hence, we assigned the longest lifetime component (τ_2 = ~4.70 ns) to the parallel GQ topology and τ_1 and τ_3 to either random coil or hybrid-type GQ structure. The loop residues in a juxtaposed quadruplex and hairpin motif can orient coaxially or orthogonally depending on the GQ topology adopted by the sequence. While a coaxial orientation between the hybrid-type GQ and stem-loop results in a continuous base stacking across the two structures, orthogonal interaction between the parallel GQ and hairpin does not involve base stacking between the two components.[81,82] To gain further insights, models of hybrid and parallel GQ forms were built by computational analysis. The hairpin structure was found to be coaxially oriented with the hybrid-type GQ core (Figure 17A, left part). In this topology, T_{27} was sandwiched between the hairpin domain and a G-tetrad (Figure 17A, right part). Hence, TFBF-dU placed at this position may experience a similar environment, wherein it would be stacked between C_{26} of the hairpin and G_4 of the bottom tetrad. As a result of stacking interaction, the hybrid GQ structure of ON **4** exhibits a lower fluorescence intensity and a shorter lifetime. In the case of parallel topology, the hairpin structure was found to be orthogonal to the GQ core (Figure 17B, left part). In this orientation, T_{27} was found to partially stack with G_{17} that formed the top G-tetrad (Figure 17B, right part). TFBF-dU placed at T_{27} should experience reduced stacking interaction with neighbouring bases, and hence, the parallel form of ON **4** displays higher fluorescence and longer lifetime. For these reasons, the fluorescence intensity and contribution of τ_2 increases as a function of K^+ ion concentration due to progressive shift in the equilibrium towards the formation of the more emissive parallel GQ form.

A hybrid GQ



B parallel GQ

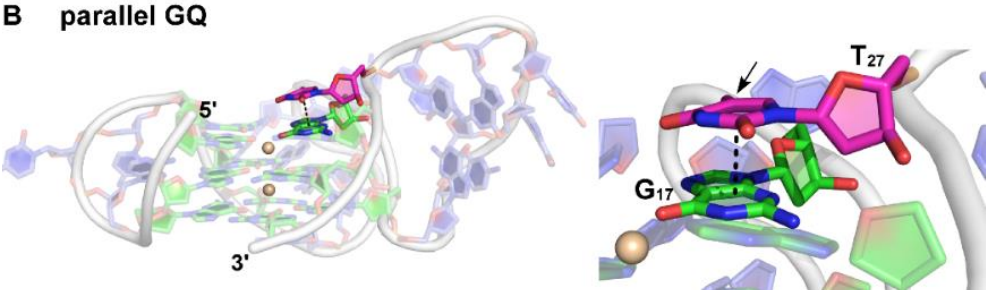


Figure 17. (**A**) Left: Axial view of the major cluster of the hybrid GQ structure. Right: Zoomed-in image representing the stacking interaction of T_{27} (shown in magenta) with adjacent bases. (**B**) Left: Axial view of the major cluster of the parallel GQ structure. Right: Zoomed-in image representing the stacking interaction of T_{27} (shown in magenta) with adjacent bases. Since the 5′ terminal dG was excluded in both architectures, the numbering of the nucleotides in the ON sequence has been renumbered. Dashed lines represent the stacking interaction. C5 position of T_{27} is shown in arrow where TFBF heterocycle modification is placed in ON **4**. Carbon atoms of G-tetrads are represented in green and all other nucleotides are represented in purple. Nitrogen atoms are represented in blue, oxygen in red, and phosphate in orange. Hydrogen atoms are omitted for clarity.

2.2.6 Probing EGFR GQ structural equilibrium by ^{19}F NMR

^{1}H NMR spectra of modified ON **4** and control ON **6** revealed broad peaks for imino protons in the GQ region (10 to 12 ppm) suggesting the presence of multiple GQ topologies in equilibrium. In addition, a peak at 12.8 ppm indicated the presence of a Watson-Crick paired hairpin structure (Figure 18B and 19). However, due to only minor changes in the signals, the effect of ionic condition on the conformational equilibrium could not be deduced from ^{1}H NMR analysis. Rewardingly, the incorporated nucleoside analog **1** gave distinct and resolved ^{19}F signals for different structures. A systematic analysis using ^{19}F signatures obtained under different ionic conditions allowed us to gain a deeper understanding of the GQ structural equilibrium. Notably, upon increasing the concentration of KCl, the peak at -60.55 ppm increased progressively at the expense of other peaks (Figure 18A). At 200 mM KCl, this peak was found to be predominant. CD and fluorescence data obtained at high concentrations of K^+

ions indicate the formation of a parallel GQ as the major form and a hybrid-type GQ as the minor form. So, the peak at -60.55 ppm was judiciously assigned to a parallel GQ topology and the broad peak (minor, -61.25 ppm) to hybrid GQ topology (Figure 18A). Further, integrating the ^{19}F NMR peaks at different KCl concentrations provided valuable information on the population dynamics of parallel and hybrid GQ topologies. At 25 mM of KCl, relative population of the parallel GQ was found to be 23%, and the hybrid topology and non-GQ structures accounted for the remaining population. Population of the parallel GQ significantly increased from 23% to 66%, with increase in KCl concentration, while the population of hybrid GQ and non-GQ structures decreased to 34%, a result similar to the fluorescence data. A control experiment with nucleoside probe **1** in the presence of different concentrations of KCl did not affect the ^{19}F NMR chemical shift similar to fluorescence (Figure 16B). Collectively, these results confirm that ^{19}F signatures exhibited by modified ON **4** is due to differences in the microenvironment of the probe in different structures.

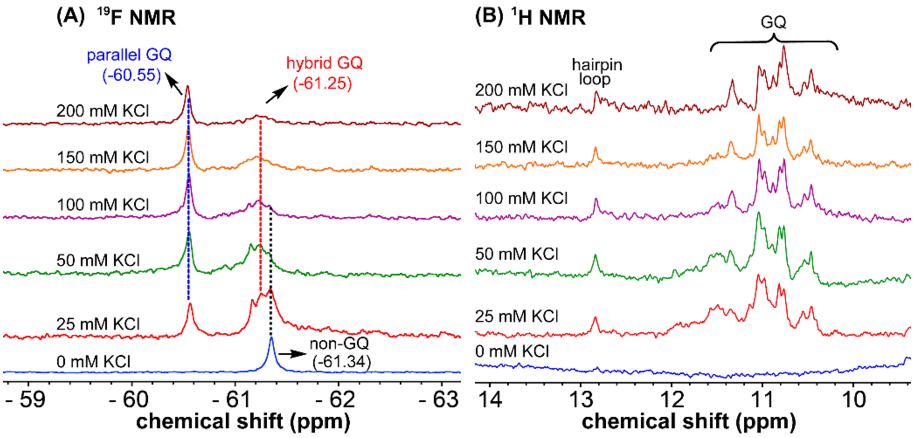


Figure 18. (**A**) ^{19}F NMR and (**B**) ^{1}H NMR spectra of TFBF-modified ON **4** at different KCl concentrations.

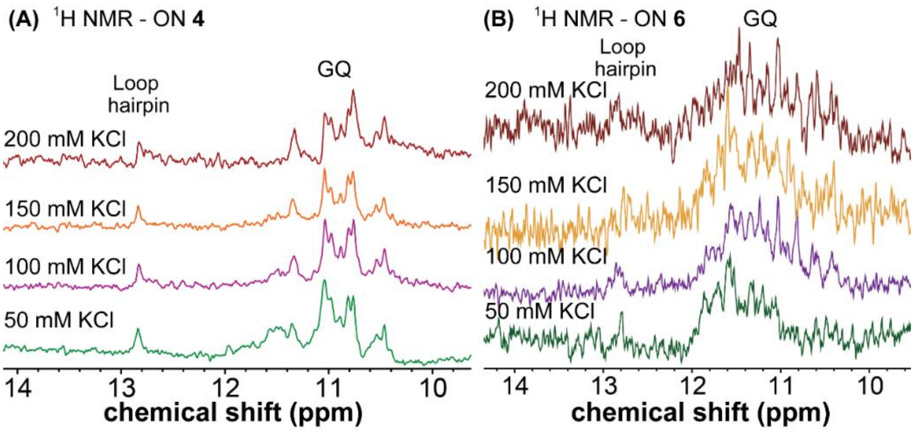


Figure 19. ^{1}H NMR of TFBF-dU modified ON **4** and unmodified ON **6** of the native sequence.

2.2.7 Juxtaposed hairpin structure influences the GQ structural equilibrium

To study the role of the hairpin junction, G_{20} and C_{21} nucleotides, involved in base paring, were mutated with Ts (ON **5**, Table 2). Mutated ON **5** exhibited contrasting fluorescence compared to native ON **4**. The random coil form of ON **5** in the absence of KCl, displayed a weak emission band centered at 419 nm (Figure 20A). Addition of KCl (25 mM) resulted in quenching in fluorescence with a slight blue shift in emission maximum as a result of formation of GQ structures (CD profiles also suggest the formation of GQs, Figure 9C). Further, addition of KCl did not affect the fluorescence profile of ON **5**. The ON exhibited three different decay components with lifetimes τ_1 = 0.95 ns, τ_2 = 0.11 ns and τ_3 = 2.67 ns in the presence of K^+ ions (Figure 20B, Table 6). Much like steady-state fluorescence, the population of these three components was more or less even and didn't change with increase in KCl concentration (Table 6). The highest lifetime component (τ_3 = 2.67 ns) was assigned to the parallel GQ topology as before, whereas shorter lifetimes could correspond to hybrid-type GQ and random coil structures (Table 6). Collectively, fluorescence at different K^+ ion concentrations indicates that mutations in the hairpin domain significantly reduced the population of more emissive parallel GQ topology and increased the amount of less emissive hybrid GQ topology.

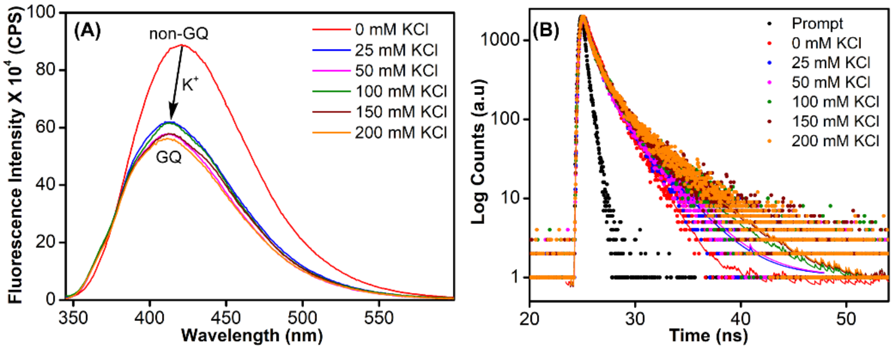


Figure 20. (**A**) Steady-state and (**B**) time-resolved fluorescence spectra of modified ON **5** at different KCl concentrations. For steady state fluorescence, samples (1 µM) were excited at 330 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively. Instrument response (prompt) is shown in grey dots and decay curve fits are shown in solid lines.

Table 6 Fluorescence lifetimes of TFBF-modified mutated ON **5**.

KCl (mM)	τ_1 (ns)	a_1 [a]	τ_2 (ns)	a_2 [a]	τ_3 (ns)	a_3 [a]
0	1.69 ± 0.02	58%	0.30 ± 0.01	42%	-	-
25	0.95 ± 0.02	37%	0.11 ± 0.01	34%	2.67 ± 0.05	29%
50	1.00 ± 0.09	37%	0.12 ± 0.01	34%	2.93 ± 0.17	29%
100	1.14 ± 0.001	37%	0.13 ± 0.001	36%	3.39 ± 0.11	27%
150	1.18 ± 0.03	38%	0.13 ± 0.001	35%	3.74 ± 0.003	27%
200	1.04 ± 0.04	35%	0.11 ± 0.002	34%	3.61 ± 0.11	31%

[a]Relative amplitude (a_1, a_2, a_3) corresponding to lifetime components (τ_1, τ_2, τ_3).

^{1}H NMR spectra of mutated ON **5** and its corresponding unmodified ON **7** in the presence of K^{+} ions revealed broad imino proton peaks between 10–12 ppm indicating the formation of more than one GQ structure (Figure 21B and 22). Further, absence of a peak at 12.8 ppm confirmed that the 3rd loop did not adopt a hairpin structure. ^{19}F NMR of **5** in the absence of KCl showed a single peak at -61.44 ppm corresponding to a non-GQ structure (Figure 21A). In presence of KCl, ON **5** folded into a mixture of parallel (-60.83 ppm) and hybrid GQ (-61.25 ppm) topologies. Even at high concentrations of K^{+} ions, the ON formed a random coil structure. CD profile revealed the presence of a mixture of GQs along with a random coil form (Figure 9C). As compared to native ON **4**, which formed majorly the parallel GQ conformation, mutated ON **5**, formed different structures that were more or less evenly populated (e.g., see peaks corresponding to parallel and hybrid-type at 100 mM KCl, Figure 23). These observations are also consistent with fluorescence intensity and lifetime data, alluding that both the labels behave similarly in reporting different conformations adopted by the ONs. Further, the existence of a random coil structure along with GQs clearly indicate that mutations in the hairpin domain reduced the overall stability of GQ structures. A ΔT_m of 4 °C between ON **4** and ON **5** also supports above results obtained from fluorescence and NMR studies (Table 4). Collectively, these results highlight the usefulness of our nucleoside probe in elucidating the role of juxtaposed hairpin motif on the formation and stability of different GQs of EGFR promoter region.

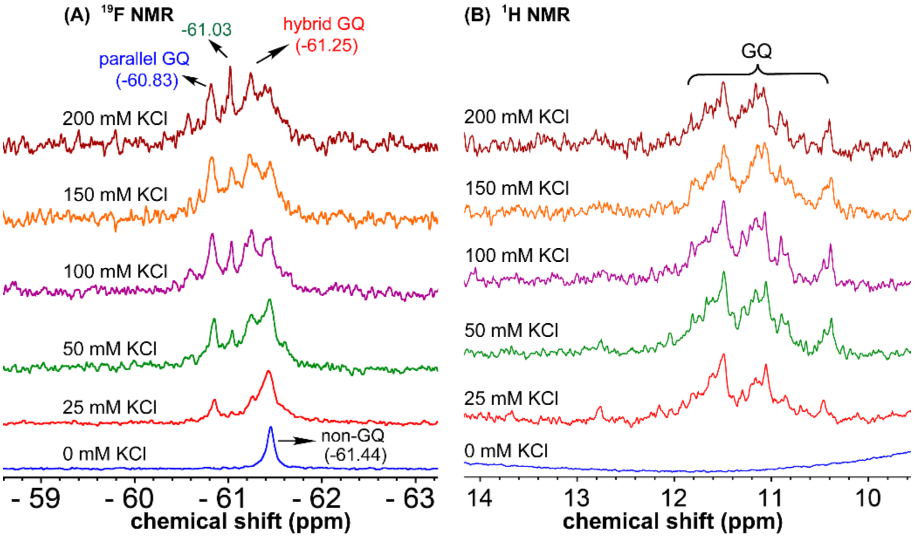


Figure 21. (**A**) ^{19}F NMR and (**B**) ^{1}H NMR spectra of mutated modified ON **5** at different KCl concentrations. Peak at -61.03 could not be assigned.

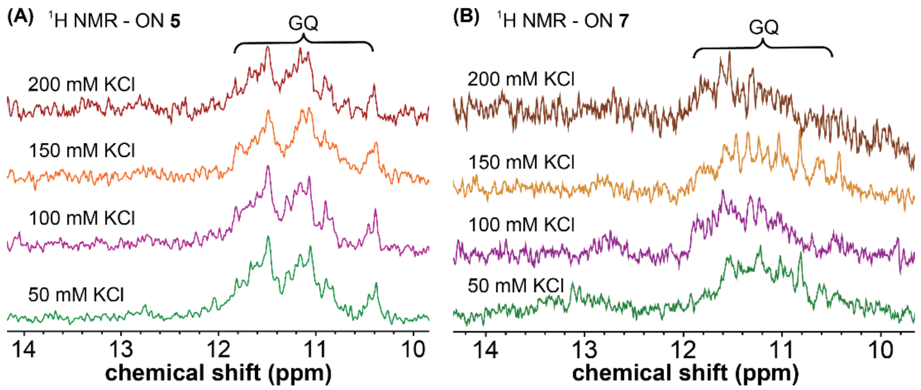


Figure 22. ^{1}H NMR of TFBF-dU modified ON **5** and unmodified ON **7** of the mutated sequence.

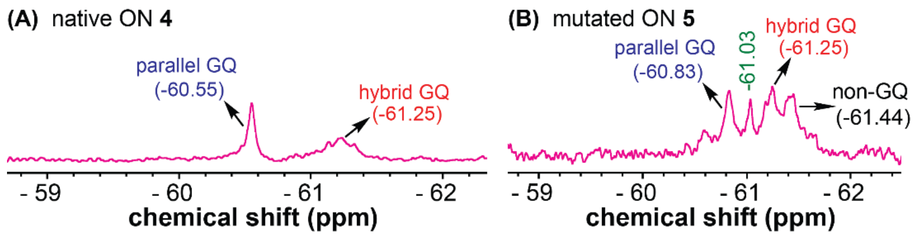


Figure 23. ^{19}F NMR spectra of the native ON **4** and mutated ON **5** at 100 mM KCl.

2.2.8 Probing ligand binding by fluorescence and ^{19}F NMR

Using the dual-labeled analog, we devised assays to estimate the ligand binding to GQ structures formed by EGFR G-rich segment. While the fluorescent label enabled the quantification of ligand binding to an ensemble of GQs, ^{19}F label provided information on the efficiency of ligand interaction with different co-existing GQ structures. Titration of modified ON **4** with increasing concentrations of commercially available GQ binders (TMPyP4, PDS, and BRACO-19, Figure 24A) resulted in a dose-dependent quenching in the fluorescence intensity with no apparent change in the emission maximum (Figure 24B and 25). At low concentrations of ligands, we observed a slight increase in fluorescence intensity. However, the quenching effect was significant as the ligand concertation was increased further, which gave reliable K_d values. The observed fluorescence quenching is likely due to the proximity of the probe to the bound polyaromatic ligands favouring nonradiative decay pathway.[83] TMPyP4 and PDS displayed a similar affinity for EGFR GQ with K_d values of 0.35 ± 0.02 µM and 0.37 ± 0.04 µM, respectively (Figure 24C). However, BRACO-19 displayed a slightly lower affinity with a K_d value of 0.51 ± 0.05 µM. Further, the tested ligands showed a cooperative binding with a hill coefficient of nearly 2. Since TMPyP4 binding to GQ structures exhibit a red-shifted absorption band,[84] the binding of this ligand to EGFR ON **4** was also evaluated by

UV-Vis absorption spectroscopy. Notably, UV titration experiments gave a comparable apparent K_d value (0.14 ± 0.01 μM) for TMPyP4 binding to GQs of ON **4** (Figure 26).

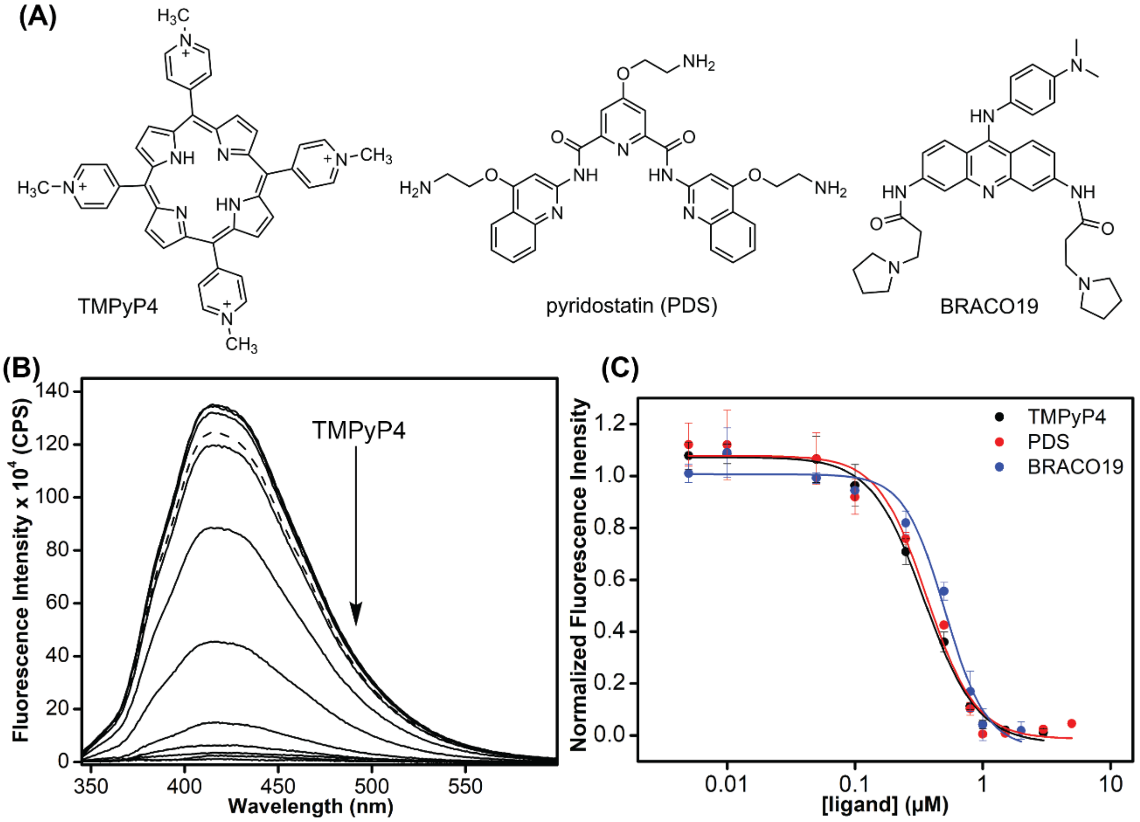


Figure 24. (**A**) Structures of ligands used in this study. (**B**) Fluorescence spectra of ON **4** (0.5 μM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl with increasing concentrations of TMPyP4 (solid lines). Fluorescence spectrum of ON **4** in the absence of ligand (dashed line). Samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively. (**C**) Curve fits for the binding of ligands (TMPyP4, PDS, and BRACO-19) to ON **4**.

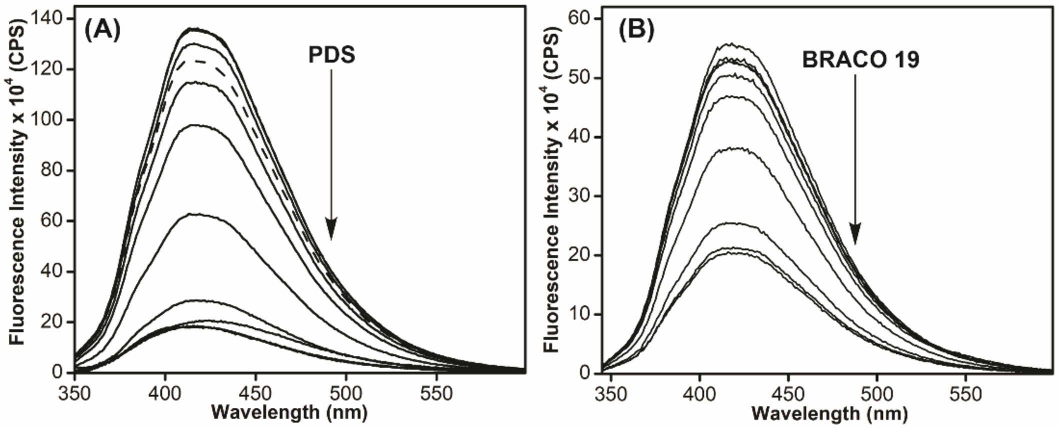


Figure 25. Fluorescence titration of ON **4** (0.5 μM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of (**A**) PDS, and (**B**) BRACO-19. Fluorescence spectra of the modified ON **4** in the absence of ligands represented by a dashed line. Samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively.

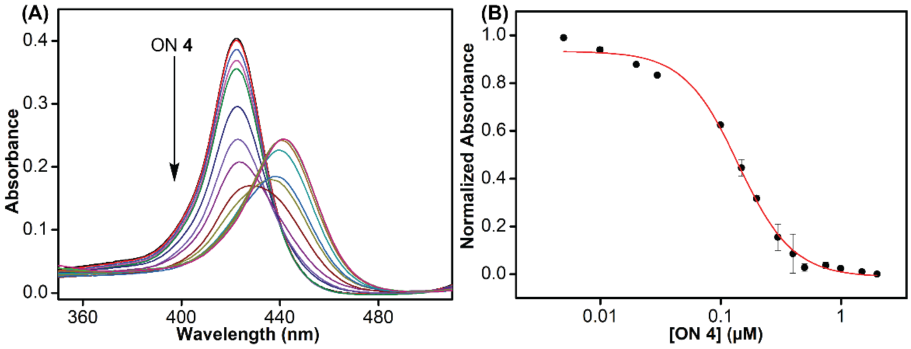


Figure 26. (**A**) UV-vis absorption spectra. Titration of TMPyP4 (2 μM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ON **4** (5 nM to 2 μM). (**B**) Curve fit plotted for normalized absorbance of TMPyP4 at 422 nm with increasing concentrations of ON **4**.

It is not straight forward to determine topologic-specific binding of ligands when a sequence forms multiple structures. Ligands could bind either to a specific topology or topologies and stabilizes the structures or could transform the structure, thereby redistributing the population.[85] The ability of our probe to display distinct ^{19}F signals for different GQs was put to use in detecting topology-specific binding of ligands. Modified ON **4**, which folds into a mixture of parallel (-60.56 ppm) and hybrid GQ (-61.20 ppm) topologies at 100 mM KCl was titrated with TMPyP4, PDS, and BRACO-19 (Figure 27). Upon increasing TMPyP4 concentration, peak intensity corresponding to the parallel GQ (-60.56 ppm) decreased, and further, it disappeared at a higher concentration of TMPyP4 (ON:ligand = 1:2, Figure 27A). A broad peak around -61.20 ppm increased with increasing concentrations of the ligand. A CD spectrum revealed the presence of multiple conformations (mainly parallel and hybrid forms, Figure 28A). Hence, it is speculated that the signal at -61.20 ppm could be associated to different GQ-ligand complexes. In contrast, PDS and BRACO-19 showed affinity for both parallel and hybrid GQ topologies. We observed that with an increase in PDS and BRACO-19 concentration, the peaks corresponding to parallel (-60.55 ppm) and hybrid (-61.21 ppm) GQs reduced, and new peaks corresponding to different GQ-ligand complexes appeared (Figure 27B and 27C), which was confirmed by the presence of CD signatures for both parallel and hybrid GQs (Figure 28B and 28C). Further, ligands are known to interact with GQ *via* different binding modes (external stacking, intercalation, and groove binding).[83] We speculate that multiple peaks observed upon PDS and BRACO-19 binding might be due to different modes of interaction of these ligands with GQs.[86,87] From our studies, we find that although ligands (TMPyP4, PDS, and BRACO-19) interact with EGFR GQs with similar K_d values (~0.35–0.5 μM), their specificity, mode of interaction and influence on structural equilibrium are different.

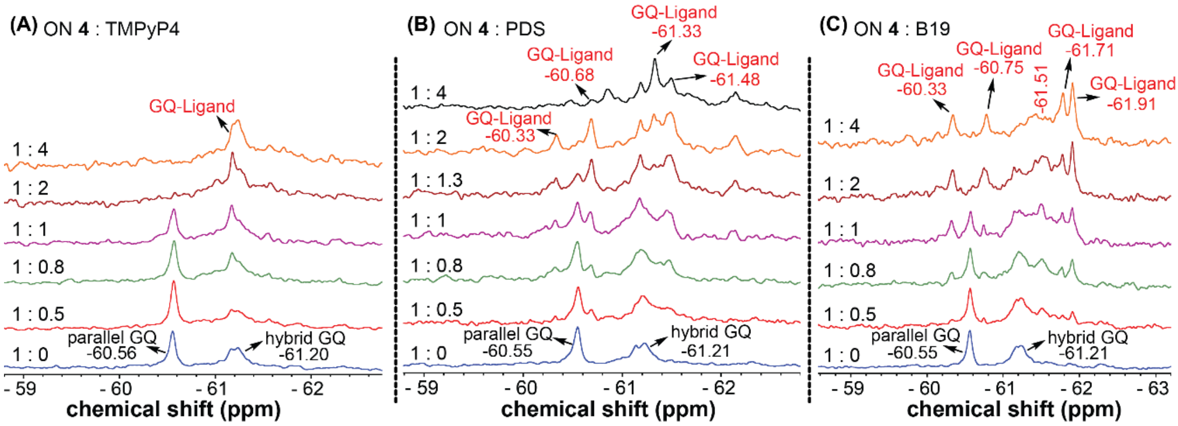


Figure 27. ^{19}F NMR of modified ON **4** (15 µM) in Tris.HCl buffer (pH 7.4) containing 100 mM KCl with increasing concentrations of (**A**) TMPyP4, (**B**) PDS, and (**B**) BRACO-19 (B19).

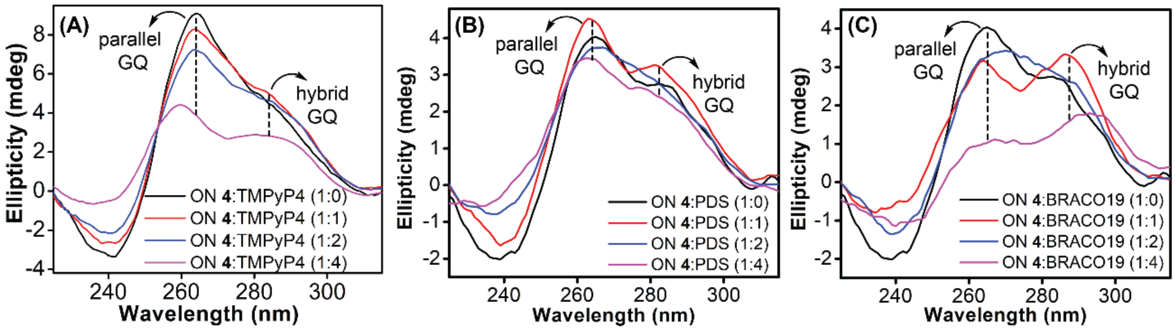


Figure 28. (**A**) CD spectra of modified ON **4** (10 µM) upon addition of increasing concentrations of TMPyP4. CD of modified ON **4** (5 µM) at different concentrations of (**B**) PDS, and (**C**) BRACO-19.

2.2.9 Probing GQ structures in a model cellular environment by ^{19}F NMR

Next, we studied the utility of the ^{19}F label in detecting the formation of GQ structures in a model mimicking a cellular environment. Lysate and egg extract from frog oocytes serve as good *ex vivo* models to conduct NMR experiments to study the structure of biopolymers.[88,89] To gain a progressive understanding of GQ structures formed by EGFR ON **4** in cell-free and cellular environments, ^{19}F NMR experiments were performed in intraoocyte ionic conditions, oocyte clear lysate and egg extract. The clear lysate was prepared by mechanically crushing the oocytes and the suspension thus obtained was incubated at 95 °C for 10 min to denature the proteins. The lysate was obtained by centrifugation. To obtain an egg extract, the oocytes were crushed and without any further manipulations the mixture was centrifuged. The inter-phase egg extract is considered to provide a molecular crowded environment and also known to maintain most of the biological activity of a cell.[88] ^{19}F NMR of ON **4** recorded in a buffer mimicking intraoocyte ionic conditions (25 mM HEPES pH = 7.5, 110 mM KCl, 10.5 mM NaCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA) gave a spectral profile reminiscent of GQ structures formed in Tris buffer containing 100 mM KCl (Figure 29). However, in intraoocyte ionic conditions the population of hybrid GQ (-61.23 ppm) was found to be discernibly higher compared to the parallel form (-60.55 ppm), which was the major conformer in Tris buffer containing 100 mM KCl (compare with Figure 18A). Interestingly, in both lysate and egg extract the signal corresponding to the hybrid GQ population dominated as the signal for parallel GQ structure almost disappeared (Figure 29). Based on these data it appears that a hybrid topology of the EGFR ON is supported in a cellular environment. The observed signal broadening in the egg extract is not surprising as inhomogeneity and crowding can reduce the rate of tumbling and augment the relaxation process.[88-90] To ascertain if the ^{19}F signals obtained in lysate are associated with the GQ structures formed by the intact ON, the lysate sample was subjected to HPLC and mass analyses. A comparison of HPLC chromatograms of lysate, lystate plus ON **4** (after recording NMR), ON **4** and modified nucleoside analog clearly indicate that ON **4** is stable in the lysate and did not degrade during the NMR acquisition process (Figure 30). Further, mass analysis of ON **4** isolated from the lysate sample confirmed the identity of the ON (Figure 31). It is important to note that ^{1}H NMR spectrum of ON **4** gives a basic information on GQ formation till lysate level, but fails to provide information in the egg extract due to severe line broadening. This is a commonly observed effect and a limitation of ^{1}H NMR use in the cell analysis.

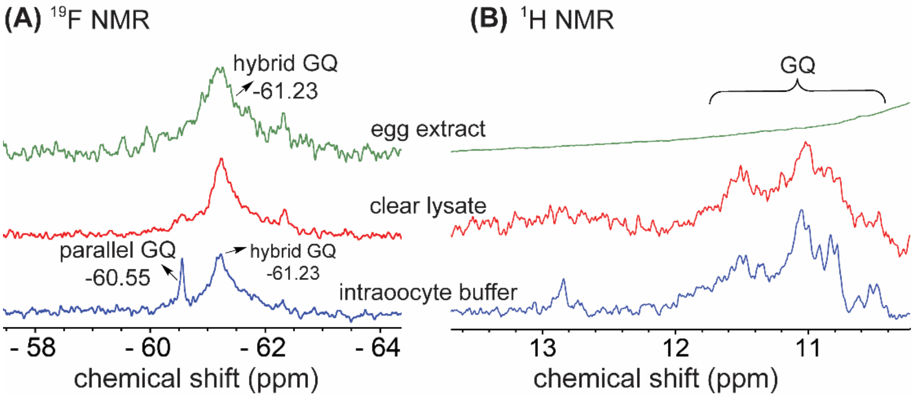


Figure 29. ^{19}F and ^{1}H NMR spectra of ON **4** (50 µM) in intraoocyte buffer (blue), lysate (red) and egg extract (green). See section for 13 details.

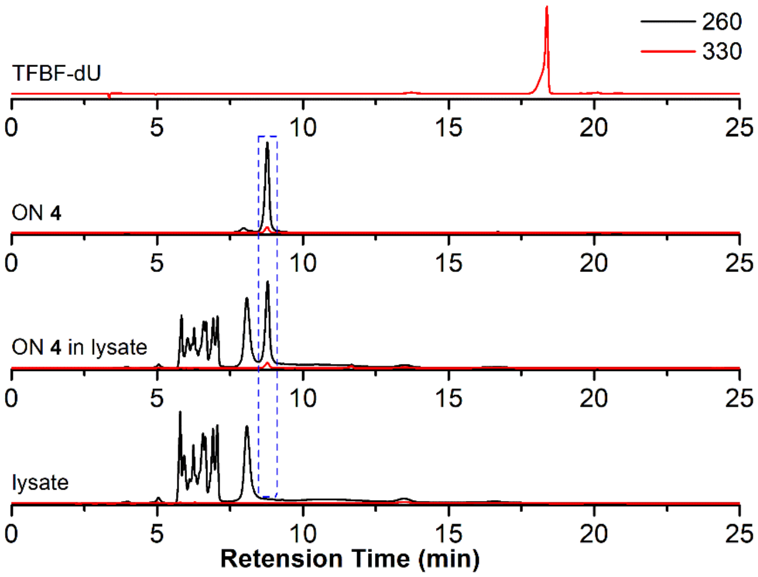


Figure 30. Comparison of HPLC chromatograms of lysate, lysate containing ON **4** (after recording NMR), ON **4** and modified nucleoside analog (TFBF-dU). There was no detectable degradation of ON **4** in lysate.

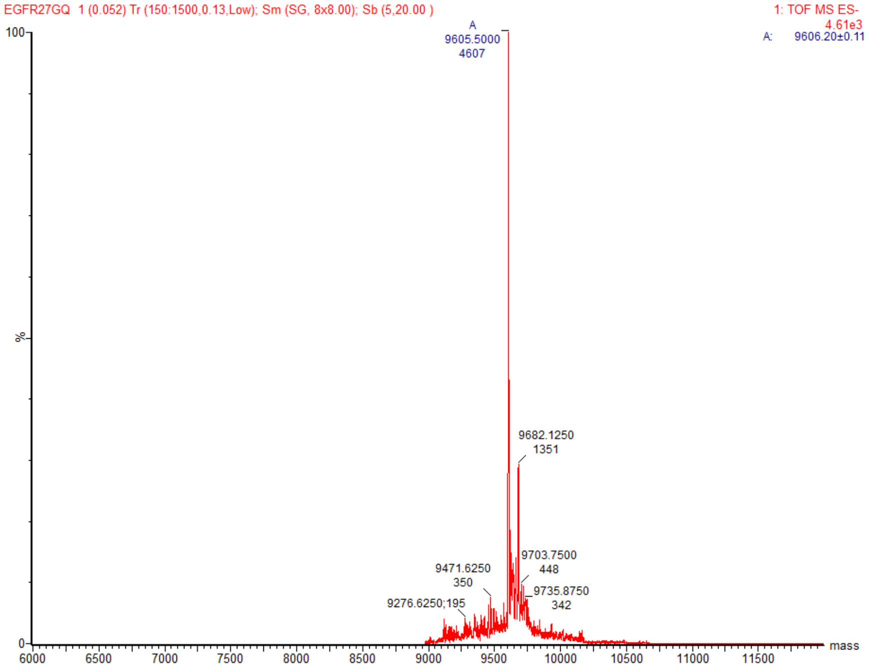


Figure 31. ESI-MS spectra of modified ON **4** extracted from lysate sample after NMR analysis. (calculated mass = 9605.1, observed mass = 9605.5).

2.2.10 GQ formation stalls the processivity of DNA polymerase

Noncanonical DNA structures can cause genomic instability by stalling the DNA replication process.[91] To study the effect of GQs, we designed three templates (**T$_1$**–**T$_3$**) composed of a wild-type EGFR G-rich sequence, a mutated EGFR sequence or a random non-GQ forming sequence downstream of the polymerization start site (Table 7). A 5'-FAM-labeled primer **P$_1$** was annealed with templates **T$_1$**–**T$_3$** in Tris.HCl buffer containing 100 mM KCl, and the replication reaction was performed using *Taq* DNA polymerase at 37 °C. At different time intervals, individual reaction samples were quenched by adding denaturing gel loading buffer and flashed cooled on a dry ice bath. Reaction products were resolved by PAGE under denaturing conditions and imaged using a fluorescence scanner. Reactions with **T_3** majorly produced the full-length product in 1 min as a result of absence of a GQ-forming motif in the template (Figure 32A, blue bars and Figure 33C, gel image). On the other hand, **T_1** containing the wild-type GQ-forming sequence stalled the polymerase activity and produced largely truncated primer-extended products just above the primer band along with minor amounts of the full-length product (Figure 32A black bars and Figure 33A, gel image). At a reaction time of 5 min, it produced nearly 40% of the full-length product as compared to 60% product in the presence of a control template **T_3**. A mutated EGFR template **T_2** also yielded stalled products similar to template **T_1** (Figure 32A, red bars and Figure 33B, gel image). While these results indicate that GQ formation hampers the processivity of *Taq* polymerase, mutations in the loop hairpin of the GQ has only a minor impact on the polymerization process. This is because mutations in the hairpin per se only destabilized the GQs by ~4 °C (Table 4) and do not fully hamper the formation of respective GQs of EGFR as proved by our CD, fluorescence and ^{19}F NMR studies (*vide supra*).

Table 7 DNA ON sequences used in Taq polymerase stop assay.

ON	5'------------sequence--------- 3'
P_1	FAM-GGAGCTCAGCCTTCACTGC
T_1	GGGGACCGGGTCCAGAGGGGCAGTGCTGGGCGGCGCAGTGAAGGCTGAGCTCC
T_2[a]	GGGGACCGGGTCCAGAGGG**TT**AGTGCTGGGCGGCGCAGTGAAGGCTGAGCTCC
T_3	TCCTAACCCTAACTCTAACTCTAACGGCGCAGTGAAGGCTGAGCTCC

[a]**T** represents mutation points.

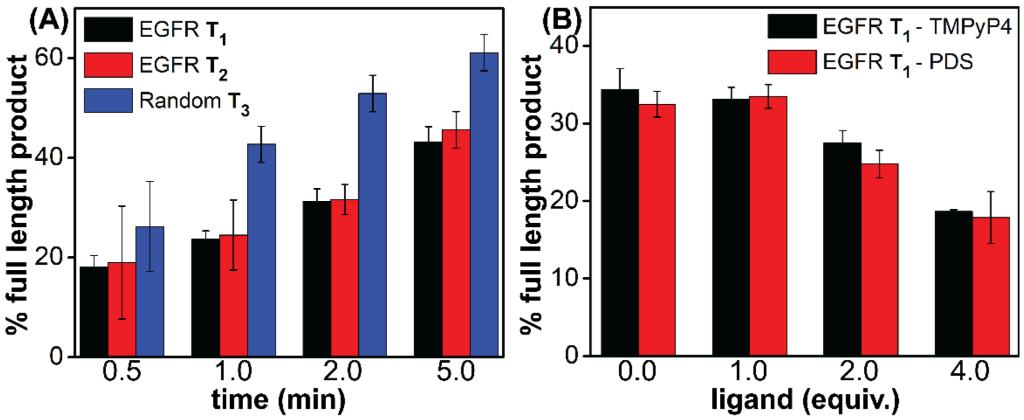


Figure 32. Percentage of the full-length product obtained from *Taq* polymerase reactions. (**A**) Reactions performed using templates **T_1–T_3** at different time points. (**B**) Reactions performed using **T_1** with increasing concentrations of ligands (TMPyP4 and PDS) at 2 min.

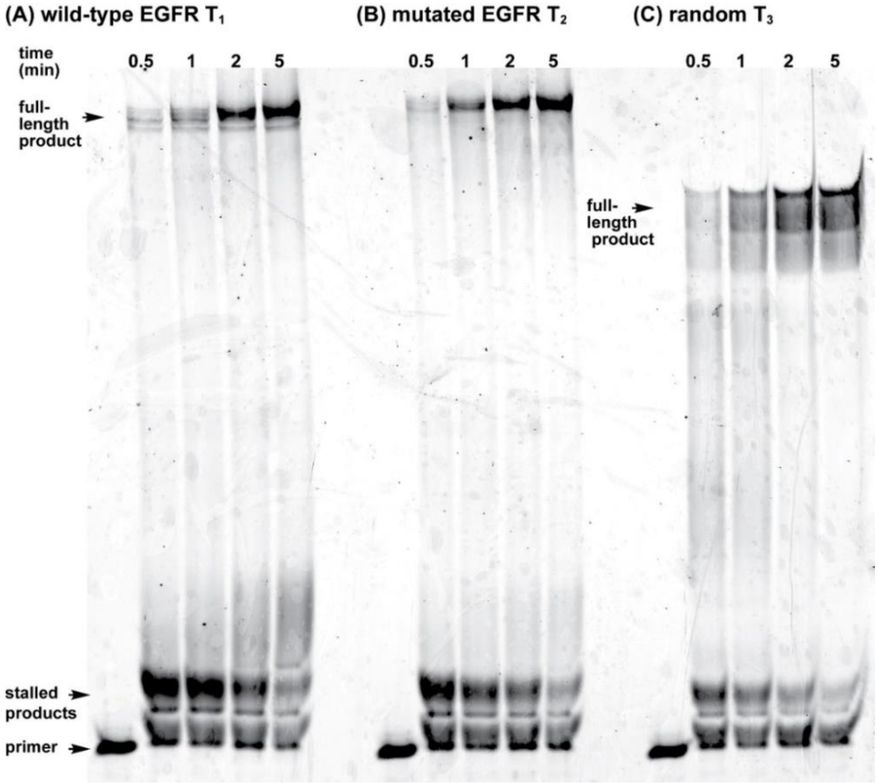


Figure 33. PAGE analysis of the replication reactions using (**A**) a wild-type EGFR G-rich template **T$_1$**, (**B**) a mutated EGFR G-rich template **T$_2$**, and (**C**) a random non-GQ forming template **T$_3$**.

Next, to evaluate the effect of GQ-ligand interaction on the polymerase activity, stop assay was performed in the presence of ligands TMPyP4 and PDS. Since the primer extension with a wild-type EGFR template produced reasonable amounts of the full-length product in 2 min (~35%), reactions with increasing concentration of ligands were performed for 2 min at 37 °C (Figure 34). Upon increasing the concentration of both the ligands, the intensity of the full-length product reduced significantly with a concomitant increase in the stalled products. At 4 equivalents of the ligand, a nearly two-fold decrease in the efficiency of formation of the full-length was observed (Figure 32B and Figure 34). The observed extent of inhibition of the polymerase activity as a result of ligand binding to the GQs is reasonable in case of EGFR because regulated EGRF expression is very important for the normal cell growth, differentiation and proliferation. Hence, instead of totally abrogating the expression of EGRF gene, it is more useful if the overexpression of this gene is suppressed so that its normal cellular functions are not affected. In addition, EGRF expression can be more specifically controlled by designing ligands that target hybrid GQ-hairpin conformations adopted by the promoter region.

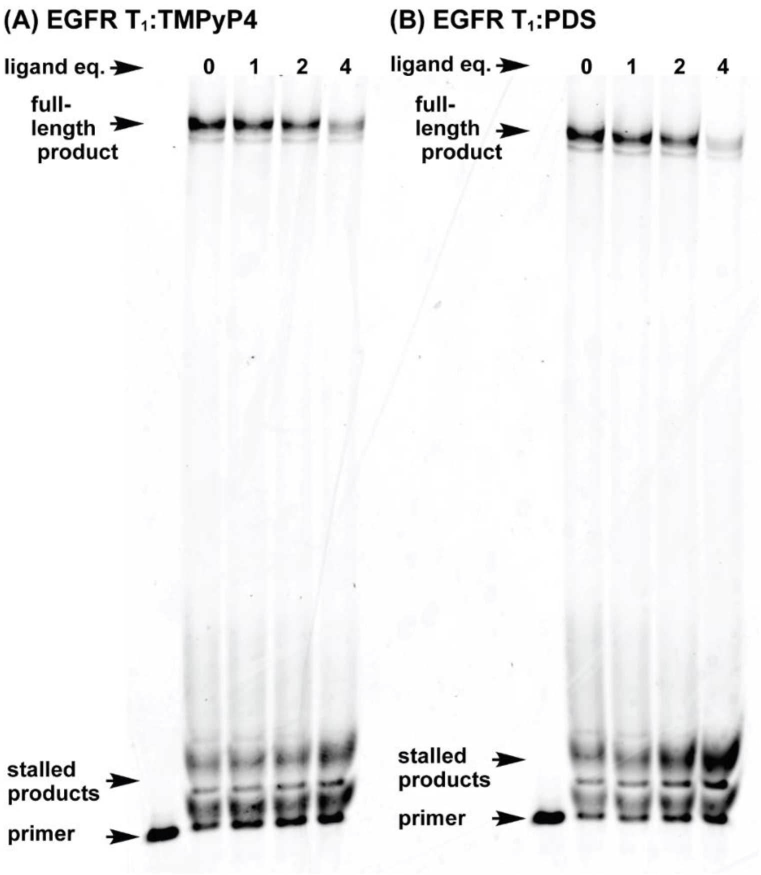


Figure 34. PAGE analysis of the replication reactions using a wild-type EGFR **T1** with increasing concentrations of the ligands (**A**) TMPyP4, (**B**) PDS.

2.3 Conclusions

Given the high number of putative GQ forming sequences in the human genome, non-canonical GQ-hairpin forms exhibited by EGFR G-rich promoter DNA sequence are perceived as unique hotspots for achieving ligand-binding selectivity. To validate this notion, we utilized a new dual-purpose nucleoside analog probe (TFBF-dU) that enabled a systematic analysis of different GQ topologies formed by native and mutated EGFR promoter ONs by fluorescence and ^{19}F NMR techniques. Distinct spectral signatures exhibited by the labels allowed the (i) detection of different GQ topologies, (ii) quantification of GQ population equilibrium, (iii) estimation of ligand binding to different GQs. Notably, ^{19}F label provided insights on the ligand-induced conformational changes in GQs, thereby suggesting that our probe could be utilized to detect topology-specific binders. Further, a comparative analysis of the readouts from the probe within native and mutated ONs clearly revealed the influence of the hairpin domain in not only affecting the stability of GQs but also redistributing the population of hybrid-type and parallel GQ topologies. Notably, experiments in frog egg lysate and extract suggest that TFBF-dU could serve as a useful probe in detecting nucleic acid structures in a cellular environment. Finally, progressive stalling of the DNA polymerase activity in the absence and then in the presence of ligands, suggest that stabilization of EGFR GQs could be used as a viable approach to repress the oncogenic activity of EGFR. Taken together, the utility of TFBF-dU as a GQ probe, which could be potentially used in cellular milieu, and deeper insights gained on the structural and functional aspects of EGFR GQs should facilitate the development of binders that simultaneous target both GQ and hairpin domains for enhanced selectivity and druggability.

2.4 Experimental Section

2.4.1 Materials. 4-trifluoromethyl phenol, bromoacetaldeyde diethyl acetal, polyphosphoric acid, tributyltin chloride, n-butyllithium, 4,4′-dimethoxytrityl chloride, 5-iodo-2′-deoxyuridine, *N*,*N*,*N′*,*N′*-tetramethylethylenediamine, bis(triphenylphosphine)-palladium(II) dichloride, *N*,*N*-diisopropylethylamine, 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite and other reagent were procured from Sigma-Aldrich. *N*-acetyl-protected dC, *N*-Benzoyl-protected dA, *N*,*N*-dimethylformamide-protected dG, and dT phosphoramidite substrates needed for solid-phase DNA synthesis were purchased from ChemGenes. Solid supports required for DNA synthesis were purchased from Glen Research. All other reagents required for solid phase ON synthesis were obtained from Sigma-Aldrich. Control DNA ONs Telo1, TeloC, **6** and **7** were purchased from Integrated DNA Technology, and purified using 18% denaturing polyacrylamide gel electrophoresis (PAGE). All other reagents (BioUltra grade) for the preparation of buffer solutions were purchased from Sigma-Aldrich. Autoclaved water was used for preparing all buffer solutions, and in all biophysical studies.

2.4.2 Instruments. NMR spectra of small molecules were acquired on a Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. Mass analysis was carried out using ESI-MS Waters Synapt G2-Si Mass Spectrometry instrument. Modified DNA oligos were synthesized on Applied Biosystems DNA/RNA synthesizer (ABI-394). RP-HPLC analysis was performed using Agilent Technologies 1260 Infinity HPLC. Absorption spectra were recorded on a UV-2600 Shimadzu spectrophotometer. Fluorescence of the ONs samples were recorded using a Fluoromax-4 spectrophotometer (Horiba Scientific). The time-resolved fluorescence of the ONs was performed using a HORIBA Delta Flex Time-Correlated Single Photon Counting (TCSPC) system using a 340 nm laser source. The fluorescence decay profile was deconvoluted by using EZ software, and decay was fitted with χ^2 values close to unity. UV-thermal melting analysis of the ONs was carried out on Cary 300 Bio UV-Vis spectrophotometer. CD measurements was done on a JASCO J-815 CD spectrometer. NMR spectra of the ONs were acquired on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT) and processed using Bruker TopSpin Software.

2.4.3 Synthesis of 5-(5-trifluoromethyl benzofuran-2-yl)-2'-deoxyuridine 1 and corresponding phosphoramidite 2

1-(2,2-diethoxyethoxy)-4-(trifluoromethyl)benzene (1a). 4-trifluoromethyl phenol (5.0 g, 30.8 mmol, 1.0 equiv.) and Cs_2CO_3 (14.10 g, 43.3 mmol, 1.4 equiv.) were taken in a round bottom flask and kept under high vacuum for 30 min. Anhydrous DMF (75 mL) was added under nitrogen atmosphere. The solution was cooled to 0 °C and bromoacetaldeyde diethyl acetal (6.95 mL, 46.3 mmol, 1.5 equiv.) was added dropwise. The resulting suspension was stirred at 65 °C for 24 h. Reaction was stopped by adding 200 mL of water, extracted with ethyl acetate (2 x 150 ml). The organic layer was washed with brine solution, dried over Na_2SO_4 and evaporated using a rotary evaporator under reduced pressure. The crude product was purified by silica gel column chromatography to offered a colorless liquid (8.12 g, 95%). TLC (10% EtOAc in hexane); $R_f = 0.50$; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 7.53 (d, $J = 8.4$ Hz, 2H), 6.98 (d, $J = 8.4$ Hz, 2H), 4.84 (t, $J = 5.2$ Hz, 1H), 4.04 (d, $J = 5.2$ Hz, 2H), 3.81–3.73 (m, 2H), 3.68–3.60 (m, 2H), 1.25 (t, $J = 7.2$ Hz, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 161.2, 127.0, (q, 3.6 Hz), 124.5 (q, 270 Hz), 123.3 (q, 32.6 Hz), 114.7, 100.5, 68.8, 62.9, 15.4; ^{19}F NMR (376.6 MHz, $CDCl_3$): δ (ppm) -61.54. HRMS: m/z Calculated for $C_{13}H_{18}F_3O_3$ $[M+H]^+$ = 279.1208, found=279.0965.

5-Trifluoromethyl benzofuran (1b). Compound **1a** (8.0 g, 28.8 mmol, 1.0 equiv.) and polyphosphoric acid (19.44 g, 57.5 mmol, 2.0 equiv.) were dissolved in benzene (200 mL) under nitrogen atmosphere and stirred at 100 °C for 6 h. After completion of the reaction, the reaction mixture was cooled to 0 °C and benzene was decanted. Water was then added to the above solid mass and stirred for 30 min. The water layer was washed with $CHCl_3$ (2 x 75 mL). The organic extract was mixed with the above decanted benzene extract, dried over Na_2SO_4 and evaporated using a rotary evaporator under reduced pressure at 5 °C as the desired product is volatile. The crude product was purified by silica gel column chromatography to offered a pale yellow liquid (5.10 g, 96%). TLC (5% EtOAc in hexane); $R_f = 0.86$; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 7.91 (m, 1H), 7.73 (d, $J = 2.4$ Hz, 1H), 7.6 (dd, $J_1 = 8.6$ Hz, $J_2 = 0.6$ Hz, 1H), 7.56 (dd, J_1 = 8.6 Hz, J_2 = 1.8 Hz, 1H), 6.85 (dd, J_1 = 2.4 Hz, J_2 = 0.8 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 156.4, 146.8, 127.6, 125.7 (q, 31.9 Hz), 124.8 (q, 270 Hz), 121.6 (q, 3.6 Hz), 119.1 (q, 4.1 Hz), 112.0, 107.0; ^{19}F NMR (376.6 MHz, $CDCl_3$): δ (ppm) -60.85. GCMS: m/z calculated for $C_9H_5F_3O$ $[M]^+$ = 186, found = 186.

Tributyl (5-trifluoromethyl benzofuran-2-yl) stannane (1c). 5-Trifluoromethyl benzofuran **1b** (4.5 g, 24.2 mmol, 1 equiv.) and TMEDA (4.3 mL, 29.0 mmol, 1.2 equiv.) were dissolved in dry THF (90 mL) and was cooled to -78 °C under nitrogen atmosphere for 15 min. To the above reaction mixture, *n*-BuLi [14.5 mL (2 M in cyclohexane), 29.0 mmol, 1.2 equiv.] was added dropwise and allowed to stir for 1.5 h at -78 °C. nBu_3SnCl (9.4 g, 29.0 mmol, 1.2 equiv.) was then added dropwise to the reaction mixture at -78 °C. The reaction was further stirred for another 1 h. Reaction mixture was brought to room temperature slowly and was quenched by adding saturated aqueous ammonium chloride solution (100 mL). The product was extracted with ethyl acetate (200 mL). The organic layer was dried over Na_2SO_4 and evaporated using a rotary evaporator under reduced pressure. The crude product was purified by silica gel column chromatography (100% hexane) to afford clear oily compound **1c** (10.0 g, 87%). TLC (5% EtOAc in hexane); R_f = 0.90; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 7.83 (br, 1H), 7.56 (d, J = 8.8 Hz, 1H), 7.48 (dd, J_1 = 8.6 Hz, J_2 = 1.8 Hz, 1H), 6.96–6.94 (m, 1H), 1.63–1.57 (m, 6H), 1.38–1.33 (m, 6H), 1.20–1.16 (m, 6H), 0.90 (t, J = 7.2 Hz, 9H); ^{13}C NMR (150 MHz, $CDCl_3$): δ (ppm) 168.4, 160.1, 128.3, 125.0 (q, 269.7 Hz), 124.8 (q, 29.7 Hz), 120.6, 118.2, 118.1, 111.4, 29.1, 27.3, 13,8, 10.4; ^{19}F NMR (376.6 MHz, $CDCl_3$): δ (ppm) -60.61; ^{119}Sn NMR (149.2 MHz, $CDCl_3$): δ (ppm) -55.80.

5-(5-trifluoromethyl benzofuran-2-yl)-2'-deoxyuridine (1). 5-Iodo-2'-deoxyuridine (1.0 g, 2.8 mmol, 1 equiv.) was dissolved in dry dioxane (30 mL) and degassed for 15 min by bubbling nitrogen gas. Compound **1c** (2.7 g, 5.6 mmol, 2 equiv.) and *bis*(triphenylphosphine)-palladium (II) dichloride (0.01 g, 0.1 mmol, 0.05 equiv.) were added and the reaction mixture was heated at 90 °C for 2 h under N_2 atmosphere. The reaction mixture was filtered through a Celite pad, the pad was washed with dioxane (2 x 20 mL) and the filtrate was evaporated. The crude product was purified by silica gel column chromatography (5% MeOH in CH_2Cl_2 containing 1% NEt_3) to afford product **1** as a white solid (0.5 g, 43%). TLC (15% MeOH in CH_2Cl_2); R_f = 0.61; ^{1}H NMR (400 MHz, d_6-DMSO): δ (ppm) 11.81 (br, 1H), 8.88 (s, 1H), 8.06 (br, 1H), 7.77 (d, J = 8.8, 1H), 7.63 (dd, J_1 = 8.8 Hz, J_2 = 1.6 Hz 1H), 7.46 (s, 1H), 6.23 (t, J = 6.2 Hz, 1H), 5.31–5.28 (m, 2H), 4.37– 4.32 (m, 1H), 3.90–3.88 (m, 1H), 3.78–3.66 (m, 2H), 2.27–2.23 (m, 2H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 160.7, 155.0, 151.6, 149.8, 138.5, 129.7, 125.1 (q, 270 Hz), 124.6 (q, 31 Hz), 121.6 (q, 3.3 Hz), 119.0 (q, 4 Hz), 112.0, 104.5, 104.2, 87.9, 85.6, 70.3, 61.1, 40.9; ^{19}F NMR (376.6 MHz, d_6-DMSO): δ (ppm) -59.25; HRMS: m/z calculated for $C_{18}H_{16}F_3N_2O_6$ [M+H]$^+$ = 413.0960, found = 413.0971; λ_{max} (H_2O) = 265, 271 and 319 nm, ε_{265} = 12.10 x 10^3 M^{-1}cm^{-1}, ε_{271} = 12.45 x 10^3 M^{-1}cm^{-1}, ε_{319} = 17.08 x 10^3 M^{-1}cm^{-1}, ε_{260} = 11.40 x 10^3 M^{-1}cm^{-1}.

5-Trifluoromethyl benzofuran modified 5'-O-DMT-2'-deoxyuridine (1d). A mixture of **1** (0.4 g, 1.0 mmol, 1 equiv.), DMAP (0.06 g, 0.5 mmol, 0.5 equiv.) and DMT-Cl (0.5 g, 1.5 mmol, 1.5 equiv.) in a round bottomed flask was kept under high vacuum for 30 min. Dry pyridine (12 mL) was added to the above mixture and was stirred at room temperature for 12 h under N_2 atmosphere. After completion of the reaction, pyridine was evaporated under vacuum and the crude product was purified by silica gel column chromatography (1% MeOH in CH_2Cl_2 containing 1% NEt_3) to afford product **1d** as a light yellow foam. (0.35 g, 51%). TLC (5% MeOH in CH_2Cl_2); R_f = 0.31; ^{1}H NMR (400 MHz, d_6-DMSO): δ (ppm) 11.93 (br, 1H), 8.41 (s, 1H), 7.98 (br, 1H), 7.46–7.43 (m, 3H), 7.32–7.29 (m, 5H), 7.22 (t, *J* = 7.8 Hz, 2H), 7.10 (t, *J* = 7.4 Hz, 1H), 6.77–6.73 (m, 4H), 6.44 (d, *J* = 8.8 Hz, 1H), 6.22 (t, *J* = 6.4, 1H), 5.35 (d, *J* = 4.4 Hz, 1H), 4.26–4.22 (m, 1H), 4.0–3.97 (m, 1H), 3.58 (s, 3H), 3.56 (s, 3H), 3.39 (dd, J_1 = 10.6 Hz, J_2 = 2.2 Hz, 1H), 3.17 (dd, J_1 = 10.6 Hz, J_2 = 4.2 Hz, 1H), 2.36–2.29 (m, 2H); ^{13}C NMR (100 MHz, d_6-DMSO): δ (ppm) 160.2, 158.0, 157.9, 154.2, 150.7, 149.2, 144.7, 136.7, 135.6, 135.4, 129.7, 129.6, 129.1, 127.8, 127.7, 126.6, 124.7 (q, 270 Hz), 123.9 (q, 31 Hz), 120.5 (q, 3.6 Hz), 118.4 (q, 4 Hz), 113.1, 113.0, 110.9, 104.2, 104.0, 86.0, 85.8, 85.5, 70.0, 63.3, 54.82, 54.79, 40.6; ^{19}F NMR (376.6 MHz, d_6-DMSO): δ (ppm) -59.35; HRMS: m/z calculated for $C_{39}H_{34}F_3N_2O_8$ [M+H]$^+$ = 715.2278, found = 715.2235.

5-trifluoromethyl benzofuran modified 5'-O-DMT-2'-deoxyuridine phosphoramidite (2). To a mixture of DMT-protected nucleoside **1d** (0.32 g, 0.4 mmol, 1 equiv.) in dry CH_2Cl_2 (7 mL) was added DIPEA (0.39 mL, 2.2 mmol, 5 equiv.) and stirred at 0 °C for 15 min under N_2 atmosphere. To this solution 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.15 mL, 0.67 mmol, 1.5 equiv.) was added dropwise at 0 °C and the mixture was stirred for 2.5 h at room temperature. Reaction mixture was diluted with ethyl acetate (20 mL) and washed with 5% sodium bicarbonate solution. The organic layer was dried over sodium sulfate, evaporated to dryness and crude product was purified by silica gel column chromatography (~20% acetone in hexane containing 1% NEt_3) to get pure product **2** as a white foam. (0.23 g, 56%). TLC (40% acetone in hexane containing 1% NEt_3); R_f = 0.43 and 0.51 for two diastereomers; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.69 (s, 1H), 7.7 (br, 1H), 7.54–7.52 (m, 2H), 7.46 (br, 1H), 7.40–7.36 (m, 4H), 7.20 (t, *J* = 7.6 Hz, 2H), 7.11–7.04 (m, 2H), 6.71–6.67 (m, 4H), 6.45–6.41 (m, 1H), 5.76 (apparent d, *J* = 8.8 Hz, 1H), 4.63–4.57 (m, 1H), 4.23–4.18 (m, 1H), 3.74 (dd, J_1 = 10.6 Hz, J_2 = 2.6 Hz, 1H), 3.62 (brs, 6H), 3.60–3.51 (m, 4H), 3.22 (dd, J_1 = 11.0 Hz, J_2 = 2.6 Hz, 1H), 2.62–2.56 (m, 1H), 2.46–2.33 (m, 3H), 1.16–1.14 (m, 12H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 160.2, 158.7, 158.6, 154.7, 149.5, 149.3, 144.6, 136.1, 136.0, 135.6, 130.3, 130.2, 129.0, 128.4, 128.0, 127.1, 125.4 (q, 32 Hz), 124.7 (q, 270 Hz), 121.1 (q, 3.5 Hz), 118.3 (q, 3.7 Hz), 117.5, 113.3, 113.2, 111.1, 106.4, 105.6, 86.8, 86.2, 86.2, 85.8, 73.0, 72.8, 62.6, 58.4, 58.2, 55.3, 43.5, 43.4, 41.0, 40.9, 24.8, 24.7, 24.6, 20.3, 20.2; ^{31}P NMR (162 MHz, $CDCl_3$): δ (ppm) 149.79, 149.18; ^{19}F NMR (376.6 MHz, $CDCl_3$): δ (ppm) -60.93; HRMS: m/z calculated for $C_{48}H_{51}F_3N_4O_9P$ $[M+H]^+$ = 915.3346, found = 915.3328.

2.4.4 Photophysical studies of 5-(5-trifluoromethyl-benzofuran-2-yl)-2'-deoxyuridine (1) in different solvents

UV absorption and steady-state fluorescence. UV absorption spectra of the modified nucleoside **1** (25 μM) were recorded in solvents of different polarity (water, methanol, dioxane) and viscosity (water, ethylene glycol, glycerol) in a quartz cuvette (Hellma, path length 1 cm). All samples contained 2.5% DMSO to solubilize the nucleoside and measurements were carried out in triplicate. Steady-state fluorescence of nucleoside **1** (5 μM) was performed in different solvents by exciting the samples at their respective lowest energy absorption maximum (Table 1). For Figures **2A** and **2B** excitation and emission slit widths were kept at 2 and 4 nm, and 2 and 3 nm, respectively. Fluorescence experiments were performed in triplicate with samples containing 0.5% DMSO in a micro fluorescence cuvette (Hellma, path length 1 cm) using Fluorolog-3 spectrophotometer (Horiba Jobin Yvon).

Time-resolved fluorescence. Excited-state decay profile of modified nucleoside **1** (5 μM) in different solvents was recorded on a TCSPC instrument (Fluorolog-3, Horiba Jobin Yvon). Samples were excited by using a 339 nm LED source (IBH, UK, NanoLED-339L) and fluorescence decay at respective emission maximum was collected in duplicate samples. Lifetimes were calculated by fitting the decay profile using IBH DAS6 software with χ^2 (goodness of fit) values nearly one for all decay fittings.

Quantum yield determination. Quantum yield of modified nucleoside **1** was determined relative to 2-aminopurine as a standard in different solvents using the following equation.

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

Here, s represents standard, x is nucleoside **1**, A is absorbance at the excitation wavelength, F is the area under the emission curve, *n* is refractive index of the solvent, and Φ_F is quantum yield. Quantum yield of the standard in water is 0.68.

^{19}F NMR of nucleoside 1 in different solvents. ^{19}F NMR spectra of the modified nucleoside **1** (150 µM) were recorded in solvents of different polarity (water, methanol, dioxane) and viscosity (water, ethylene glycol). All samples contained 15% *d6*-DMSO, and each spectrum was referenced relative to an external standard [trifluorotoluene (TFT) -63.72 ppm]. Also, fluorine NMR of the TFBF-2'-deoxyurindine **1** (15 µM) and fluorobenzofuran-modified 2'-deoxyurindine (15 µM) in D_2O were recorded to compare the fluorine peak intensity.

2.4.5 Solid-phase DNA ON synthesis TFBF-dU modified DNA ONs Telo2 and **3**–**5** were synthesized in 1 µmole scale (1000 Å CPG solid support) on ABI-394 DNA/RNA synthesizer by standard solid phase synthesis protocol using phosphoramidite **2**. ON sequences with final trityl deprotection step were synthesized, and the solid supports were treated with 30% aqueous ammonium hydroxide solution for 24 h at 55 °C. Each sample was cooled on an ice bath and centrifuged. Then supernatant was transferred to a 1.5 mL Eppendorf tube, evaporated to dryness and the ON was purified by denaturing polyacrylamide gel electrophoresis (PAGE) (18% gel). Gel was irradiated with UV-light to identify the desired band corresponding to the modified ON, which was isolated and transferred to a poly-prep column (Bio-Rad). The gel pieces were crushed and soaked in aqueous ammonium acetate (0.5 M, 3 mL) for 12 h to extract the ON. Oligo was desalted using Sep-Pak C-18 cartridges (waters). The purity and integrity of modified ONs were confirmed by RP-HPLC and ESI-MS analysis.

2.4.6 Mass analysis of modified ON

MALDI TOF analysis. The mass of TFBF-dU modified Telo2 was obtained using Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer. A solution containing 1.5 µL of Telo2 (250 µM), 2 µL of an internal DNA standard (100 µM), 4 µL of an 8:2 solution of 3-hydroxypicolinic acid and ammonium citrate buffer (100 mM, pH 9) was desalted by adding an ionexchange resin (Dowex 50W-X8, 100-200 mesh). 2 µL of the above solution was spotted on a MALDI plate and air dried. The MALDI spectrum was referenced relative to the mass of an internal DNA standard. Internal DNA standard sequence 5' TAATACGACTCACTATAG 3', m/z of +1 and +2 ions are 5466.6 and 2733.3.

ESI-MS analysis. Mass of the EGFR modified ONs were determined by ESI-MS analysis in negative mode by injecting DNA ONs (~300 pmol) dissolved in 50% acetonitrile in an aqueous solution of 10 mM triethylamine and 100 mM hexafluoro-2-propanol.

2.4.7 CD analysis. Samples of control Telo1 and modified Telo2 (10 µM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl was heated at 90 °C for 3 min. To construct duplex structures, Telo1/Telo2 (10 µM) was hybridized with a complementary TeloC sequence (1:1 equiv.) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM LiCl at 90 °C for 3 min. Similarly, EGFR DNA ONs (10 µM) samples were prepared in 10 mM Tris.HCl buffer (pH 7.4) containing different KCl concentrations or 100 mM LiCl and annealed by heating at 90 °C for 3 min. ONs samples were cooled slowly to RT and incubated overnight at RT. Samples were then diluted to a final concentration of 5 µM in 10 mM Tris.HCl buffer (pH 7.4) containing different KCl concentrations or 100 mM LiCl. CD measurements of ON samples were recorded from 310 nm to 220 nm on a JASCO J-815 CD spectrometer at 25 °C using 1 nm bandwidth. All experiments were done in duplicate with an average of three scans for each sample. The spectrum of buffer in absence of ON was subtracted from all ON sample spectra.

2.4.8 UV-thermal melting analysis. Samples of DNA ONs were annealed in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl as mentioned above. The UV-thermal melting experiment using ONs (1 µM) was performed on a Cary 300Bio UV/Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C min^{-1} interval and changes in absorbance at 295 nm were measured at every 1 °C interval. T_m values were determined from the forward and revere cycles.

2.4.9 Fluorescence and NMR studies of modified Telo2 ON.

Fluorescence. Single-stranded modified Telo2 ON was annealed in 10 mM Tris.HCl containing 100 mM KCl and its duplex was constructed by hybridizing with a complimentary TeloC sequence (1:1) in 10 mM Tris.HCl containing 100 mM LiCl. Samples were heated at 90 °C for 3 min and slowly cooled to RT. Fluorescence of DNA ON samples (1 µM) was recorded by exciting the samples at 330 nm with excitation and emission slit widths of 6 nm and 6 nm, respectively. Experiments were done in triplicate in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on a Fluoromax-4 spectrofluorometer (Horiba Scientific).

NMR. The sample of Telo2 (25 µM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and 20% D_2O and 1:1 mixture of Telo2 and its complementary ON TeloC in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM LiCl and 20% D_2O was annealed at 90 °C for 3 min. Samples were slowly cooled to RT and transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F and ^{1}H NMR spectra were acquired at a frequency of 564.9 MHz and 600 MHz, respectively, on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT). All ^{19}F NMR spectrum were calibrated relative to an external standard, trifluorotoluene (TFT = −63.72 ppm). Spectral parameters for ^{19}F NMR: excitation pulse: 12 µs; spectral width: 29.90 ppm; transmitter frequency offset: -60.00 ppm; acquisition time: 0.1 s; relaxation delay: 1.0 s; number of scans: 500. Using these parameters, spectra were obtained in 10 min. Each spectrum was processed with an exponential window function using lb = 15 Hz. ^{1}H NMR spectra were obtained with water suppression using excitation sculpting with gradients. Number of scans were in the range of 700.

2.4.10 Fluorescence study of modified EGFR ONs.

Steady-state fluorescence: Samples of TFBF-dU modified ON **4** and **5** (1 µM) were prepared in 10 mM Tris.HCl (pH 7.4) buffer containing different KCl concentrations or 100 mM LiCl by using the above procedure. Additionally, control solutions of free nucleoside (TFBF-dU) (2 µM) in 10 mM Tris.HCl (pH 7.4) buffer containing different KCl concentrations or 100 mM LiCl were prepared. Fluorescence measurements were performed by exciting ONs samples at 330 nm and TFBF-dU samples at 320 nm, respectively. Excitation and emission slit widths for respective samples are provided in the figures caption. Experiments were performed in triplicate in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on a Fluoromax-4 spectrofluorometer (Horiba Scientific).

Time-resolved fluorescence analysis of modified ONs. Samples of TFBF-dU modified ONs **4** and **5** (4 µM) in 10 mM Tris.HCl (pH 7.4) buffer were prepared and used for excited-state decay kinetics measurements. ONs samples were excited using a 340 nm LED source and the fluorescence decay was collected at respective emission maximum with an increase in KCl concentrations. The concentration of the KCl was increased by adding different aliquots of 3

M KCl into the ON sample and incubated for one hour before measurements. Fluorescence decay of the ON samples were deconvoluted using EZ software and fitted by an exponential decay with χ^2 value close to unity.

2.4.11 Computational analysis

In order to generate the model for the two architectures of the EGFR G-rich sequences, a combination of 3D-NuS[92] webserver and manual editing has been employed. For generating the hybrid architecture, the loop sequences were initially added to the 3D-NuS web server, where the hybrid GQ structure of class Q17 was generated. The 5′ terminal dG was excluded in both architectures. The web server-generated structure was extracted in PDB format. The hairpin domain of the structure lacked any base pairing interaction. Therefore, the sequence of the hairpin domain was separately modeled with the help of the DNA folding form of the Mfold web server,[93] RNAalifold web server, and 3DNA[94]. The hairpin domain from the 3D-NuS PDB structure was removed, and the newly generated hairpin was placed with the help of PyMOL software. Further, the coordinates of the GQ and the newly generated hairpin domain were combined manually to generate a new PDB file. Similarly, for the parallel GQ, the primary architecture was generated from the 3D-NuS web server, after which the hairpin domain was placed separately. The hairpin domain, as well as dT26 (dT27 of ON **4**), were manually placed with the help of PyMOL, and residue numbers were corrected and then saved in the PDB format.

Both the PDB structures generated were added into the Tleap module of AmberTools 21, and K^+ counter ions were added to generate the coordinate and topology files. The central K^+ ions were also added in both cases. The OL15[95] force field was used for DNA and TIP3P water model for water and counter ions. The structures were subjected to a 100000-step minimization using the implicit solvent model in AMBER 18.[96] The final structures after the minimization was again loaded to the tleap module. These structures were enclosed in a rectangular water box of 10 Å, and the coordinate and topology files were generated. The systems were then subjected to 10000 steps of minimization by the steepest descent method with a restraint of 2.0 kcal/mol $Å^2$ on the DNA and central K^+ ions. The minimization followed 100 ps of heating and 100 ps of density equilibration with restraints of 50 kcal/mol $Å^2$ and 2.0 kcal/mol $Å^2$. The systems were then equilibrated for 800 ps in NPT ensemble, and unrestrained MD simulation was performed using NPT ensemble for 400 ns in GPU accelerated version of PMEMD[97-99] in AMBER 18. SHAKE algorithm was applied to subject the hydrogen atoms to bond length constraints. Langevin[100] thermostat with a collision frequency of 2 ps^{-1} was used to maintain the temperature at 300 K, and Berendsen[101] barostat with a relaxation time of 2 ps was used to maintain pressure. The trajectories were visualized using UCSF Chimera[102], and images were rendered using PyMOL. The analysis was carried out using the CPPTRAJ[103] module of AmberTools.

The 400 ns trajectories were clustered into 5 ensembles using the hierarchical agglomerative clustering algorithm. The hybrid GQ has one major cluster which existed for ~60 % of the simulation time (Figure 17A), and the parallel GQ has one major cluster which existed for ~50 % (Figure 17B) of the simulation time. Since the 5′ terminal dG was excluded in both architectures, the numbering of the nucleotides in the ON sequence has been renumbered. In the hybrid GQ, the hairpin domain maintains the two base pairs (dG19:dC25 and dC20:dG24) intact during the entire simulation, while the same is lost towards the end of the simulation in the parallel GQ. dT26 (dT27 in ON **4** and **6**) stacks with both dC25 and dG3 maintaining an average distance of 3.80 ± 0.24 Å and 4.10 ± 0.32 Å respectively in the hybrid GQ. In the case of parallel GQ, dT26 partially stacks over dG16, maintaining an average distance of 4.49 ± 1.17 Å. The center of mass of the heavy atoms of the bases (excluding sugar) was considered for the distance calculation.

Note: Computational analysis has done by Prof. P. I. Pradeepkuma and his student Sruthi Sudhakar.

2.4.12 ^{19}F and ^{1}H NMR analysis of modified EGFR DNA ONs at different KCl

Samples of the modified DNA ONs **4** and **5** (25 µM) were prepared in 10 mM of Tris.HCl buffer (pH 7.4) containing no KCl or 100 mM LiCl and 20% D_2O and annealed by heating at 90 °C for 3 min. Samples were cooled slowly to RT and incubated overnight at RT. The sample was transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F and ^{1}H NMR spectra were recorded at a frequency of 564.9 MHz and 600 MHz, respectively, on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT). Furthermore, aliquot of 3 M KCl was added into ON sample, and incubated for one hour after each addition and ^{19}F and ^{1}H NMR spectra were acquired with increase in KCl concentrations at 25 °C. Additionally, samples of control DNA ONs **6** and **7** (25 µM) were prepared as mentioned above and ^{1}H NMR spectra were recorded at different KCl concentrations. All ^{19}F NMR spectra were referenced relative to an external standard, trifluorotoluene (TFT = −63.72 ppm). Spectral parameters for ^{19}F NMR are same as mentioned in section 2.4.9. The ^{19}F NMR spectra were obtained in 40 min with 1500 scans. Spectra were processed with an exponential window function using lb = 10 Hz. ^{1}H NMR spectra were recorded with water suppression using excitation sculpting with gradients. Number of scans were in the range of 700.

2.4.13 GQ-ligand interaction by fluorescence

A series of samples of modified ON **4** (0.5 µM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ligands (5 nM to 5 µM) were prepared. Samples were incubated at 4 °C for 1 h. Fluorescence emission spectra were recorded by exciting the samples at 330 nM with excitation and emission wavelength slit widths of 6 nm and 7 nm, respectively. The fluorescence of samples was recorded in triplicate reading at 25 °C. Further, the fluorescence of the blank sample containing control ON **6** and respective ligand concentration was subtracted from the individual sample reading. The apparent K_d values of ligands were determined by plotting the normalized fluorescence intensity (F_N) *vs* ligand concentration. The graph was prepared using OriginPro 8.5 software.[58]

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

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

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

2.4.14 GQ-ligand interaction by UV absorption

The samples of TMPyP4 (2 µM) were prepared in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of pre-annealed ON **4** (5 nM to 2 µM). Samples were incubated at 4 °C for 1 h and UV absorption spectra were acquired at 25 °C. Titration was performed until the wavelength and intensity of the absorption band of TMPyP4 remained unchanged.[84] UV experiment was performed in duplicate. Binding constant (K_d) was obtained from the plot of normalized absorbance (A_N) at 422 nm vs concentration of ON **4**. The graph was prepared using OriginPro 8.5 software.

$$A_N = \frac{A_i - A_s}{A_0 - A_s}$$

A_i is the absorbance intensity at each titration point. A_0 and A_s are the absorbance intensity in the absence of ON **4** and at saturation point, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$A_N = A_0 + (A_s - A_0)\left(\frac{[\mathrm{ON}]^n}{[K_d]^n + [\mathrm{ON}]^n}\right)$$

2.4.15 GQ-ligand interaction by ^{19}F NMR

Modified ON **4** (15 µM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and 20% D_2O was prepared and ^{19}F NMR spectra (ns = 4000) of the sample was recorded with increasing concentration of ligand (0–60 µM). Spectral parameters are the same as mentioned in Section 2.4.9 and the data was processed with an exponential window function using lb = 20 Hz.

2.4.16 Preparation of EGFR GQ (ON 4) sample for ^{19}F NMR analysis in intraoocyte buffer, lysate and egg extract.

Intraoocyte buffer. Modified ON **4** (50 µM) was annealed in an intraoocyte buffer (25 mM HEPES pH = 7.5, 110 mM KCl, 10.5 mM NaCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA)[58] containing 20% D_2O at 90 °C for 3 min. The sample was cooled slowly to RT and incubated for 1 h at RT. The ^{19}F (number of scans = 1000) and ^{1}H NMR (number of scans = 1500) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned in section 2.4.9 and the ^{19}F NMR data was processed with an exponential window function using lb = 20 Hz.

Oocytes were surgically removed from anesthetized adult female *Xenopus laevis* in accordance with a protocol approved by the Institutional Animal Ethics Committee (IAEC), IISER Bhopal.

Clear lysate.[88] Healthy *Xenopus laevis* stage V/VI oocytes (~275) were selected and suspended in a Petri dish containing Ori-Ca^{2+} buffer (5 mM HEPES pH = 7.5, 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. The oocytes were washed with ice-cold intraoocyte buffer (3 x 10 mL) and resuspended in the same buffer. Oocytes were transferred in an Eppendorf tube, allowed them to settle down and the supernatant was removed carefully without disturbing the settled oocytes. Oocytes were rinsed with intraoocyte buffer (200 µL) containing 20% D_2O and then removed (this step was performed twice). Finally, 200 µL of intraoocyte buffer containing 20% D_2O was added to the Eppendorf tube containing oocytes, which were then mechanically crushed. The suspension was centrifuged at 20000g for 20 min at 4 °C, the interphase layer was carefully transferred into another Eppendorf and heated at 95 °C for 10 min. The solution was centrifuged at 20000g for 10 min at 4 °C and the clear lysate (285 μL) was transferred to another Eppendorf tube. 1 mM of preannealed ON **4** (15 μL) in an intraoocyte buffer supplemented with 20% D_2O was added to the clear lysate (final ON concentration = 50 μM). The sample was incubated at 4 °C for 30 min and transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F (number of scans = 2000) and ^{1}H NMR (number of scans = 3072) spectra were acquired at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR plot was processed with an exponential window function using lb = 20 Hz.

Egg extract.[89] Oocytes (~900) were kept in a Petri dish containing Ori Ca^{2+} buffer (pH 7.5) for 15 min on an ice bath. The oocytes were washed with ice-cold intraoocyte buffer (3 x 20 mL) and transferred to an Eppendorf tube. The buffer just above the oocytes was removed carefully and oocytes were washed with intraoocyte buffer (400 μL) containing 20% D_2O (two times). The oocytes were centrifuged at 400g for 1 min at 4 °C and the supernatant buffer was removed. The oocytes were resuspended in the intraoocyte buffer (100 μL) containing 30% D_2O and centrifuged at 12000g for 5 min at 4 °C. The eggs were mechanically crushed and the suspension was centrifuged at 12000g for 30 min at 4 °C to obtain the interphase layer. This crude egg extract thus obtained was directly used in the NMR analysis. 1 mM of the preannealed ON **4** (15 μL) in an intraoocyte buffer containing 20% D_2O was added to the above crude egg extract (285 μL) and incubated for 30 min at 4 °C. The ^{19}F (number of scans = 3500) and ^{1}H NMR (number of scans = 4000) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR spectrum was processed with an exponential window function using lb = 20 Hz.

Note: Mr. Satyajit Mishra and Dr. Jeet Kalia has performed the surgery of female *Xenopus laevis* to obtaine the oocytes. Also, they helped in preparing *clear lysate* and *egg extract*.

2.4.17 *Taq* polymerase stop assay

5’-FAM-labeled primer (5 µM) and template DNA (**T_1–T_3**) (6 µM) were annealed in 10 mM Tris.HCl buffer containing 100 mM KCl by heating at 90 °C for 5 min. Samples were cooled slowly to RT and incubated on an ice batch for 1 h. The primer-DNA duplexes were diluted to a final concentration of 1 µM in 10 mM Tris.HCl buffer containing 100 mM KCl. Replication reaction was performed on a 20 µL reaction volume containing duplex (50 nM), dNTPs (500 µM), KCl (100 mM), 1X DNA polymerase buffer.[104] The reaction mixture was incubated at 37

°C for 20 min and the reaction was initiated by adding 0.5 µL of Taq DNA polymerase (5 U/µL, *New England Biolabs*, Catlog. M0273S). The reaction was quenched at different time points by adding 10 µL of gel loading buffer (80% formamide by volume, 10 mM NaOH, 0.005% bromophenol blue (w/v)) and flashed cooled on a dry ice bath. The reaction mixture was concentrated on a speed-vac concentrator and analysed on a 15% denaturing polyacrylamide gel. The gel was imaged by using a Typhoon gel scanner at FAM wavelength.

2.5 References

1. A. Ullrich, L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield and P. H. Seeburg, *Nature*, 1984, **309**, 418–425.
2. P. Blume-Jensen and T. Hunter, *Nature*, 2001, **411**, 355–365.
3. S. Sigismund, D. Avanzato and L. Lanzetti, *Mol. Oncol*., 2018, **12**, 3–20.
4. N. Normanno, A. D. Luca, C. Bianco, L. Strizzi, M. Mancino, M. R. Maiello, A. Carotenuto, G. D. Feo, F. Caponigro and D. S. Salomon, *Gene*, 2006, **366**, 2–16.
5. J.-Y. Han, K. Park, S.-W. Kim, D. H. Lee, H. Y. Kim, H. T. Kim, M. J. Ahn, T. Yun, J. S. Ahn, C. Suh, J.-S. Lee, S. J. Yoon, J. H. Han, J. W. Lee, S. J. Jo and J. S. Lee, *J. Clin. Oncol*., 2012, **30**, 1122–1128.
6. J. R. Goffin and K. Zbuk, *Clin. Therapeutic*, 2013, **35**, 1282–1303.
7. A. J. Cooper, L. V. Sequist and J. J. Lin, *Nat. Rev. Clin. Oncol.*, 2022, **19**, 499–514.
8. A. N. Hata, M. J. Niederst, H. L. Archibald, M. Gomez-Caraballo, F. M. Siddiqui, H. E. Mulvey, Y. E. Maruvka, F. Ji, H. E. C. Bhang, V. K. Radhakrishna, G. Siravegna, H. Hu, S. Raoof, E. Lockerman, A. Kalsy, D. Lee, C. L. Keating, D. A. Ruddy, L. J. Damon, A. S. Crystal, C. Costa, Z. Piotrowska, A. Bardelli, A. J. Iafrate, R. I. Sadreyev, F. Stegmeier, G. Getz, L. V. Sequist, A. C. Faber and J. A. Engelman, *Nat. Med*., 2016, **22**, 262–269.
9. J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893–5909.
10. A. Jain, G. Wang and K. M. Vasquez, *Biochimie.*, 2008, **90**, 1117–1130.
11. H. A. Assi, M. Garavís, C. González and M. J. Damha, *Nucleic Acids Res*., 2018, **46**, 8038–8056.
12. G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867–5892.
13. M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770–780.
14. S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 275–283.
15. D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.
16. D. Varshney, J. Spiegel, K. Zyner, D. Tannahill and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 459–474.
17. H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2021, **49**, 7839–7855.
18. C. L. Grand, T. J. Powell, R. B. Nagle, D. J. Bearss, D. Tye, M. Gleason-Guzman and L. H. Hurley, *Proc. Natl. Acad. Sci. USA.*, 2004, **101**, 6140–6145.
19. A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. USA.*, 2002, **99**, 11593–11598.
20. J.-H. Tan, T.-M. Ou, J.-Q. Hou, Y.-J. Lu, S.-L. Huang, H.-B. Luo, J.-Y. Wu, Z.-S. Huang, K.-Y. Wong and L.-Q. Gu, *J. Med. Chem.*, 2009, **52**, 2825–2835.

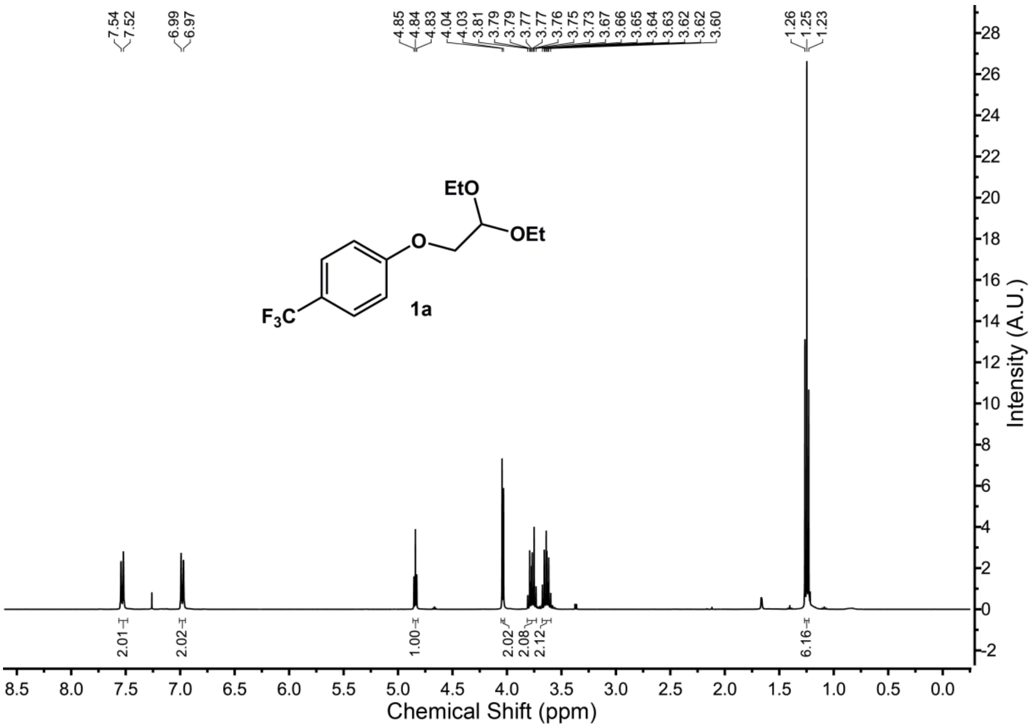
21. K. I. E. McLuckie, Z. A. E. Waller, D. A. Sanders, D. Alves, R. Rodriguez, J. Dash, G. J. McKenzie, A. R. Venkitaraman and S. Balasubramanian, *J. Am. Chem. Soc*., 2011, **133**, 2658–2663.
22. M. Micco, G. W. Collie, A. G. Dale, S. A. Ohnmacht, I. Pazitna, M. Gunaratnam, A. P. Reszka and S. Neidle, *J. Med. Chem.* 2013, **56**, 2959–2974.
23. M. Debnath, S. Ghosh, D. Panda, I. Bessi, H. Schwalbe, K. Bhattacharyya and J. Dash, *Chem. Sci.*, 2016, **7**, 3279–3285.
24. V. Dhamodharan and P. I. Pradeepkumar, *ACS Chem. Biol*., 2019, **14**, 2102–2114.
25. S. Asamitsu, S. Obata, Z. Yu, T. Bando and H. Sugiyama, *Molecules*, 2019, **24**, 429–457.
26. B. R. Vummidi, J. Alzeer and N. W. Luedtke, *ChemBioChem*, 2013, **14**, 540–558.
27. S. Wu, L. Wang, N. Zhang, Y. Liu, W. Zheng, A. Chang, F. Wang, S. Li and D. Shangguan, *Chem. Eur. J.*, 2016, **22**, 6037–6047.
28. V. Grande, C.-A. Shen, M. Deiana, M. Dudek, J. Olesiak-Banska, K. Matczyszyn and F. Würthner, *Chem. Sci.*, 2018, **9**, 8375–8382.
29. Y. Ma, K. Iida, S. Sasaki, T. Hirokawa, B. Heddi, A. T. Phan and K. Nagasawa, *Molecules*, 2019, **24**, 263–277.
30. J.-H. Yuan, W. Shao, S.-B. Chen, Z.-S. Huang and J.-H. Tan, *Biochem. Biophys. Res. Commun.*, 2020, **531**, 18–24.
31. T. Vo, S. Oxenford, R. Angell, C. Marchetti, S. A. Ohnmacht, W. D. Wilson and S. Neidle, *ACS Med. Chem. Lett.*, 2020, **11**, 991−999.
32. S. Kumar, S. P. P. Pany, S. Sudhakar, S. B. Singh, C. S. Todankar and P. I. Pradeepkumar, *Biochemistry*, 2022, **61**, 2546−2559.
33. K. W. Lim, P. Jenjaroenpun, Z. J. Low, Z. J. Khong, Y. S. Ng, V. A. Kuznetsov and A. T. Phan, *Nucleic Acids Res*., 2015, **43**, 5630–5646.
34. Y. M. Vianney and K. Weisz, *Nucleic Acids Res*., 2022, **50**, 11948–11964.
35. T. Q. N. Nguyen, K. W. Lim and A. T. Phan, *Sci. rep*., 2017, **7**, 11969–11975.
36. S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando and H. Sugiyama, *Chem. Eur. J*., 2018, **24**, 4428–4435.
37. M. Yang, S. Carter, S. Parmar, D. D. Bume, D. R. Calabrese, X. Liang, K. Yazdani, M. Xu, Z. Liu, C. J. Thiele and J. S. Schneekloth Jr, *Nucleic Acids Res.,* 2021, **49**, 7856–7869.
38. M. L. Greco, A. Kotar, R. Rigo, C. Cristofari, J. Plavec and C. Sissi, *Nucleic Acids Res.*, 2017, **45**, 10132–10142.
39. J. T. Grün and H. Schwalbe, *Biopolymers*, 2022, **113**, e23477–e23491.
40. P. A. Rachwal, T. Brown and K. R. Fox, *Biochemistry,* 2007, **46**, 3036-3044.
41. H. You, J. Wu, F. Shao and J. Yan, *J. Am. Chem. Soc*., 2015, **137**, 2424–2427.
42. J. Marquevielle, C. Robert, O. Lagrabette, M. Wahid, A. Bourdoncle, L. E. Xodo, J.-L. Mergny and G. F. Salgado, *Nucleic Acids Res.*, 2020, **48**, 9336–9345.
43. R. C. Monsen, L. DeLeeuw, W. L. Dean, R. D. Gray, T. M. Sabo, S. Chakravarthy, J. B. Chaires and J. O. Trent, *Nucleic Acids Res*., 2020, **48**, 5720–5734.
44. M. Cheng, Y. Cheng, J. Hao, G. Jia, J. Zhou, J.-L. Mergny and C. Li, *Nucleic Acids Res*., 2018, **46**, 9264–9275.
45. L. Lacroix, A. Séosse and J.-L. Mergny, *Nucleic Acids Res*., 2011, **39**, e21–e31.
46. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc*., 2006, **128**, 9963–9970.
47. G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
48. F. Doria, M. Nadai, M. Zuffo, R. Perrone, M. Freccero and S. N. Richter, *Chem. Commun*., 2017, **53**, 2268–2271.

49. Y. V. Suseela, N. Narayanaswamy, S. Pratihar and T. Govindaraju, *Chem. Soc. Rev.*, 2018, **47**, 1098–1131.
50. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182–186.
51. A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh Jr, D. Sen and P. M. Lansdrop, *Nucleic Acids Res*., 2014, **42**, 860–869.
52. A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J.-M. Perrier-Cornet, J. M. Y. Wong and D. Monchaud, *J. Am. Chem. Soc*., 2015, **137**, 8521–8525.
53. A. Shivalingam, M. A. Izquierdo, A. L. Marois, A. Vyšniauskas, K. Suhling, M. K. Kuimova and R. Vilar, *Nat. Commun.*, 2015, **6**, 8178–8187.
54. Z. Yu, J. D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu and H. Mao, *J. Am. Chem. Soc.*, 2009, **131**, 1876–1882.
55. H. You, X. Zeng, Y. Xu, C. J. Lim, A. K. Efremov, A. T. Phan, and J. Yan, *Nucleic Acids Res.*, 2014, **42**, 8789–8795.
56. L. Ying, J. J. Green, H. Li, D. Klenerman and S. Balasubramanian, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 14629–14634.
57. S. L. Noer, S. Preus, D. Gudnason, M. Aznauryan, J.-L. Mergny and V. Birkedal, *Nucleic Acids Res.*, 2016, **44**, 464–471.
58. S. Manna, D. Sarkar and S. G. Srivatsan, *J. Am. Chem. Soc*., 2018, **140**, 12622–12633.
59. Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579–2619.
60. Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem.* **2017**, *9*, 1043–1055.
61. Bood, M.; Sarangamath, S.; Wranne, M. S.; Grøtli, M.; Wilhelmsson, L. M. *Beilstein J. Org. Chem.* **2018**, *14*, 114–129.
62. Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G. *Bioconjugate Chem.* **2013**, *24*, 1367-1377.
63. Dumas, A.; Luedtke, N. W. Highly fluorescent guanosine mimics for folding and energy transfer studies. *Nucleic Acids Res*. **2011**, *39*, 6825–6834.
64. Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A. *ACS Chem. Biol.* **2015**, *10*, 1311–1318.
65. Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149–e158.
66. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.
67. Himmelstoß, M.; Erharter, K.; Renard, E.; Ennifar, E.; Kreutz, C.; Micura, R. *Chem. Sci*. **2020**, *11*, 11322–11330.
68. Li, Q.; Chen, J.; Trajkovski, M.; Zhou, Y.; Fan, C.; Lu, K.; Tang, P.; Su, X.; Plavec, J.; Xi, Z.; Zhou, C. *J. Am. Chem. Soc.* **2020**, *142*, 4739–4748.
69. (a) Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. *Nucleic Acids Res.* **2017**, *45*, 5501–5511. (b) Raikar, U. S.; Renuka, C. G.; Nadaf, Y. F.; Mulimani, B. G.; Karguppikar, A. M.; Soudagar, M. K. *Spectrochimica Acta Part A* **2006**, *65*, 673–677.
70. Sinkeldam, R. W.; Wheat, A. J.; Boyaci, H.; Tor, Y. Emissive nucleosides as molecular rotors. *ChemPhysChem*. **2011**, *12*, 567–570.
71. Giam, C. S.; Lyle, J. L. Medium effects on the fluorine-19 magnetic resonance spectra of fluoropyridines. *J. Am. Chem. Soc*. **1973**, *95*, 3235–3239.
72. Sapper, H.; Lohmann, W. Stacking interaction of nucleobases: NMR investigations III. Molecular aspects of the solvent dependence. *Biophys. Struct. Mechanism* **1978**, *4*, 327–335.
73. A. T. Phan, K. N. Luu and D. J. Patel, *Nucleic Acids Res*., 2006, **34**, 5715–5719.
74. J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.

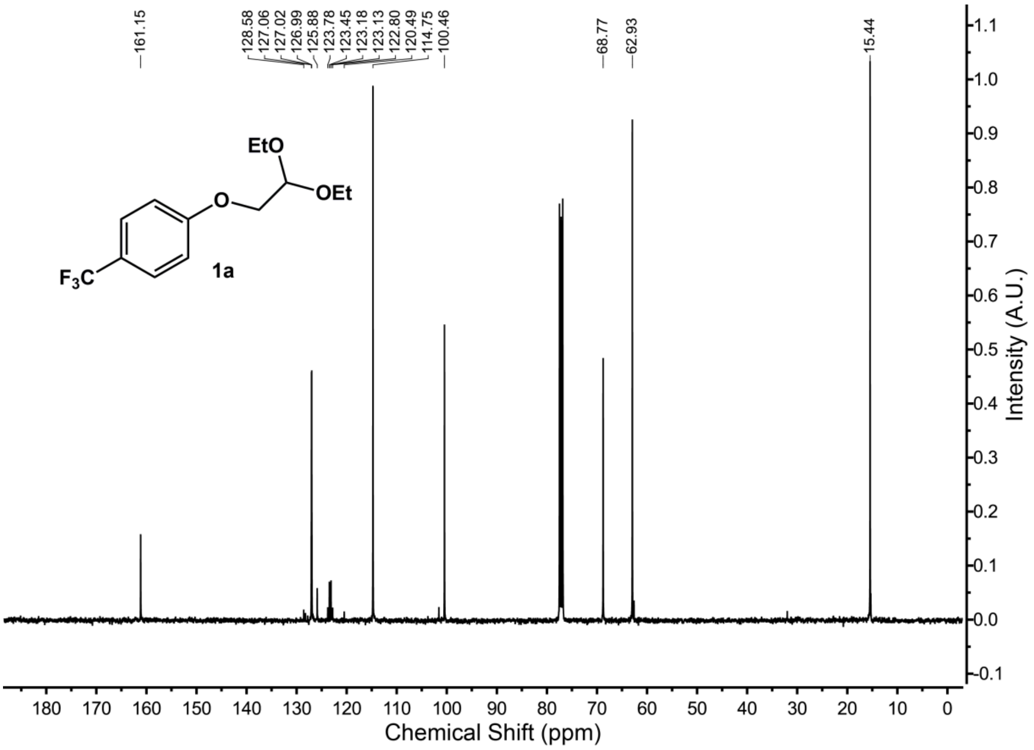
75. A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, 34, 2723–2735.
76. J.-L. Mergny, J. Li, L. Lacroix, S. Amrane and J. B. Chaires, *Nucleic Acids Res*., 2005, **33**, e138–e144.
77. L. Evans, A. Kotar, M. Valentini, A. Filloux, S. Jamshidi, J. Plavec, K. M. Rahman and R. Vilar, *RSC Chem. Biol*., 2023, **4**, 94–100.
78. J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713–1725.
79. S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375-81
80. S. Doose, H. Neuweiler and M. Sauer, *ChemPhysChem*, 2009, **10**, 1389–1398.
81. K. W. Lim and A. T. Phan, *Angew. Chem. Int. Ed.*, 2013, **52**, 8566–8569.
82. K. W. Lim, Z. J. Khong and A. T. Phan, *Biochemistry*, 2014, **53**, 247–257.
83. D. D. Le, M. D. Antonio, L. K. M. Chan and S. Balasubramanian, *Chem. Commun*., 2015, **51**, 8048–8050.
84. C. Wei, G. Jia, J. Yuan, Z. Feng and C. Li, *Biochemistry*, 2006, **45**, 6681-6691.
85. Y. Ma, K. Iida and K. Nagasawa, *Biochem. Biophys. Res. Commun*., 2020, **531**, 3–17.
86. B. Machireddy, H.-J. Sullivan and C. Wu, *Molecules*, 2019, **24**, 1010–1035.
87. L.-Y. Liu, T.-Z. Ma, Y.-L. Zeng, W. Liu, and Z.-W. Mao, *J. Am. Chem. Soc*., 2022, **144**, 11878−11887.
88. R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res*., 2011, **39**, 5768–5775.
89. H.-L. Bao, T. Ishizuka, T. Sakamoto, K. Fujimoto, T. Uechi, N. Kenmochi and Y. Xu, *Nucleic Acids Res*., 2017, **45**, 5501−5511.
90. Y. Ye, X. Liu, Z. Zhang, Q. Wu, B. Jiang, L. Jiang, X. Zhang, M. Liu, G. J. Pielak and C. Li, *Chem. Eur. J.*, 2013, **19**, 12705–12710.
91. L. K. Lerner and J. E. Sale, *Genes*, 2019, **10**, 95–119.
92. L. P. P. Patro, A. Kumar, N. Kolimi and T. Rathinavelan, *J. Mol. Biol.*, 2017, **429**, 2438–2448.
93. M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.
94. X.-J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–1227.
95. R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.
96. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi and M. K. Gilson, AMBER 2018; 2018. *Univ. California*, *San Fr.* **2018**.
97. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory. Comput.*, 2013, **9**, 3878–3888.
98. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
99. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
100. P. Turq and F. Lantelme, *J. Chem. Phys.*, 1977, **66**, 3039–3044.
101. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
102. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
103. D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
104. G. Wu and H. Han, *Methods Mol. Biol.*, 2019, **2035**, 223–231.

2.6 Appendix-I: NMR and mass data of synthesized compounds

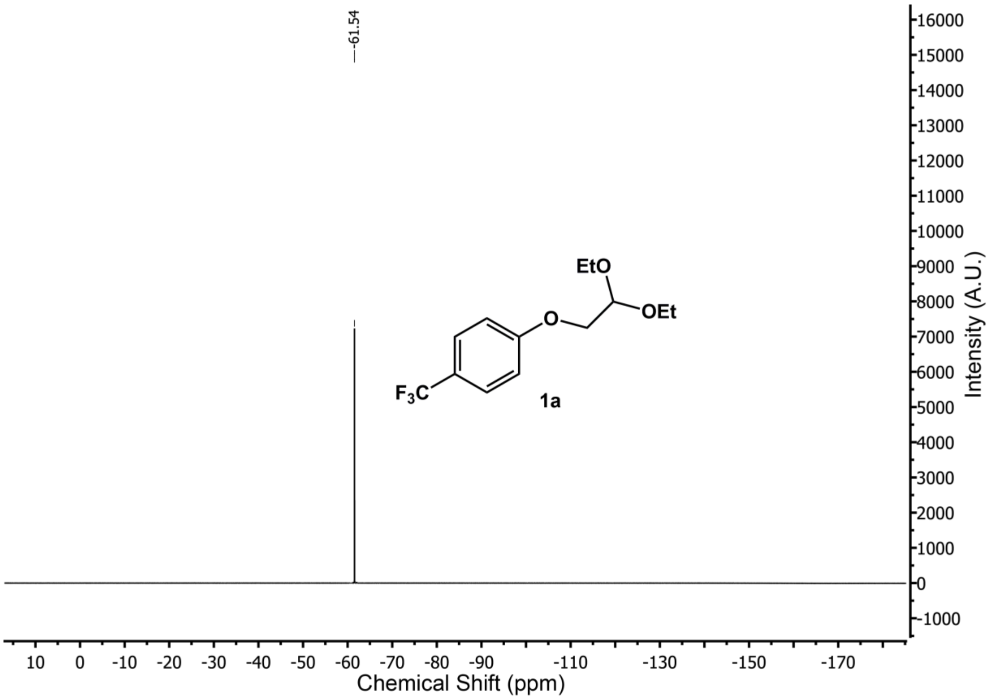
^{1}H NMR of **1a** (400 MHz, $CDCl_3$)



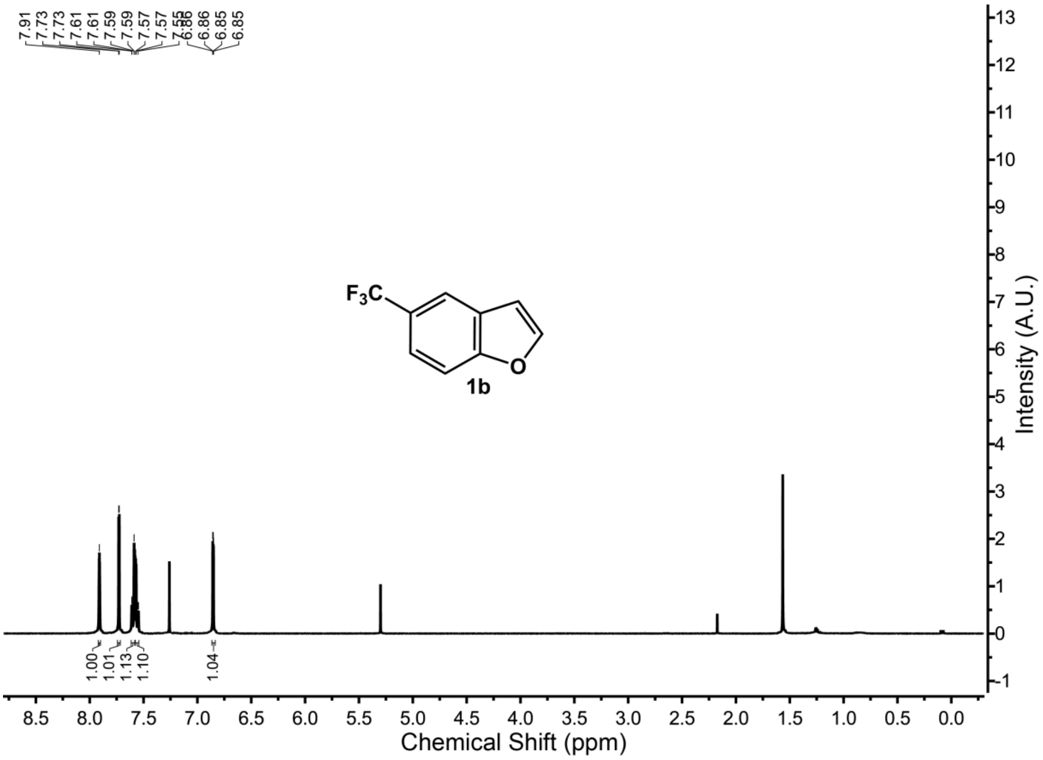
^{13}C NMR of **1a** (100 MHz, $CDCl_3$)



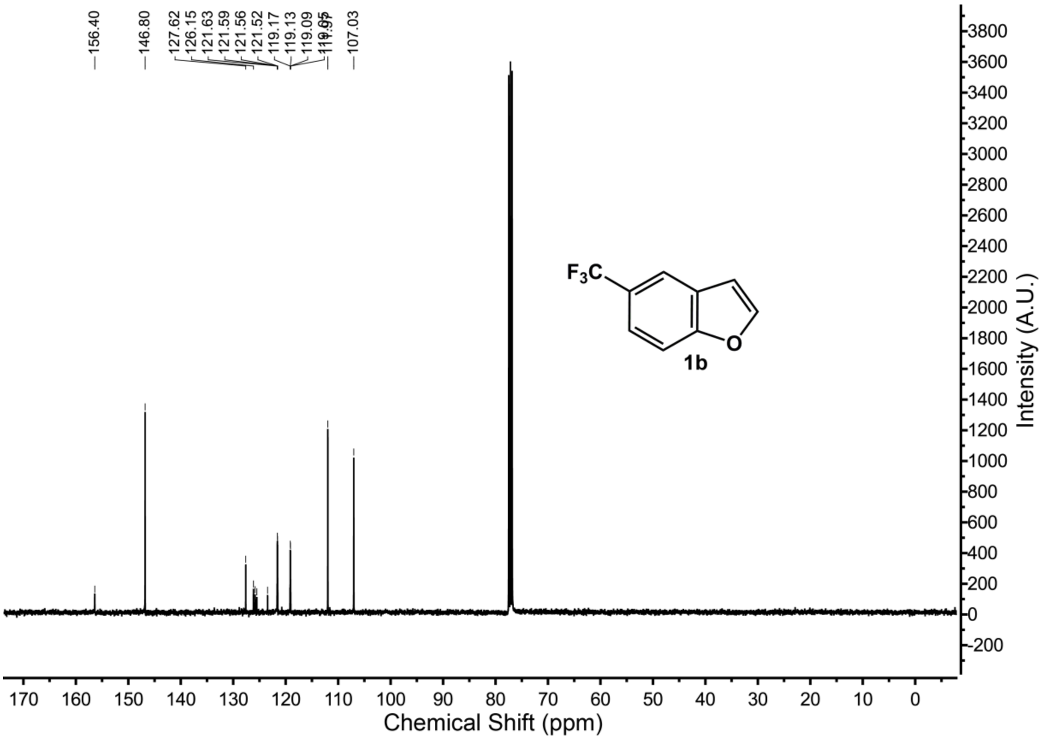
^{19}F NMR of **1a** (376.6 MHz, CDCl$_3$)



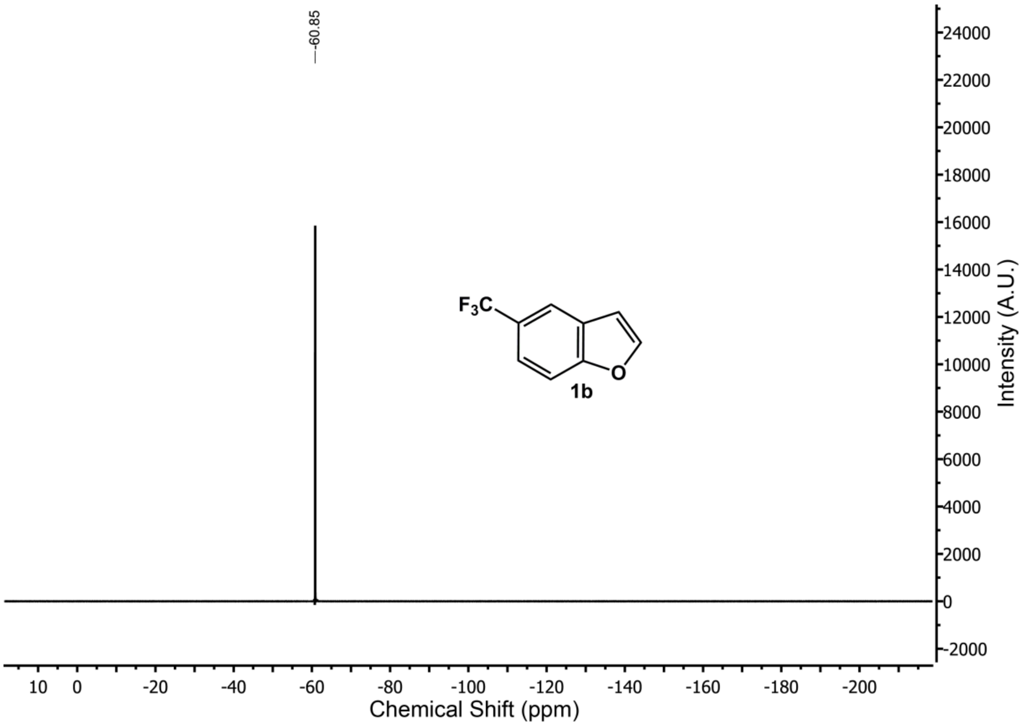
^{1}H NMR of **1b** (400 MHz, CDCl$_3$) (trace amount of DCM is present)



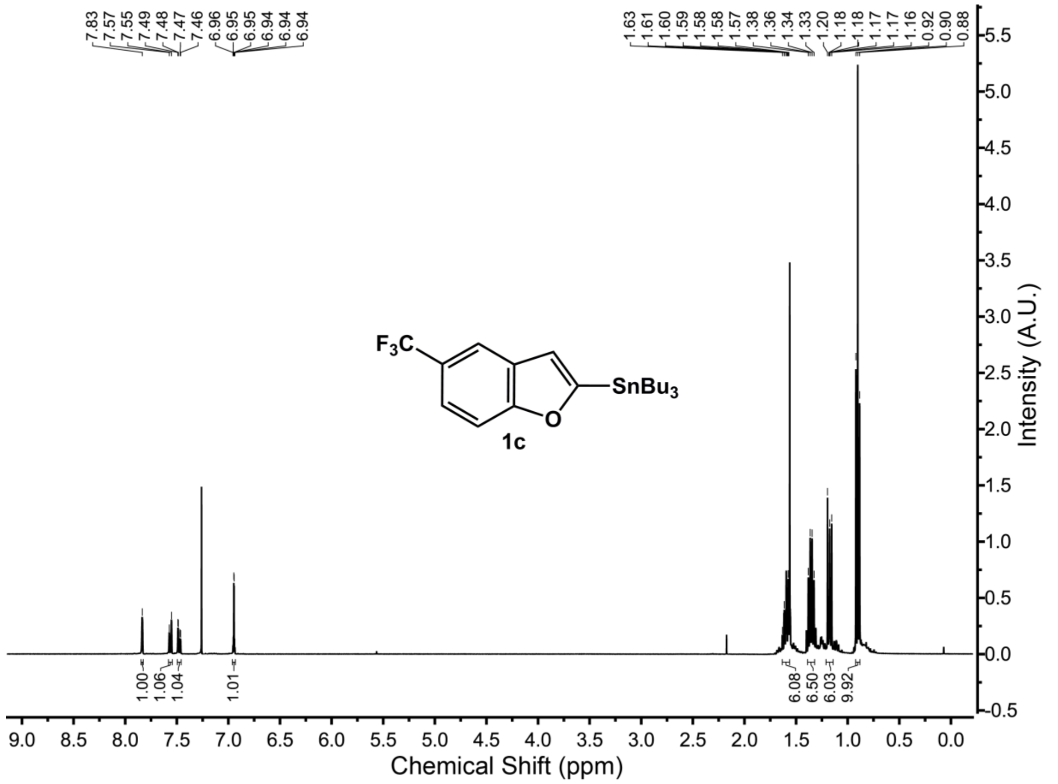
^{13}C NMR of **1b** (100 MHz, CDCl$_3$)



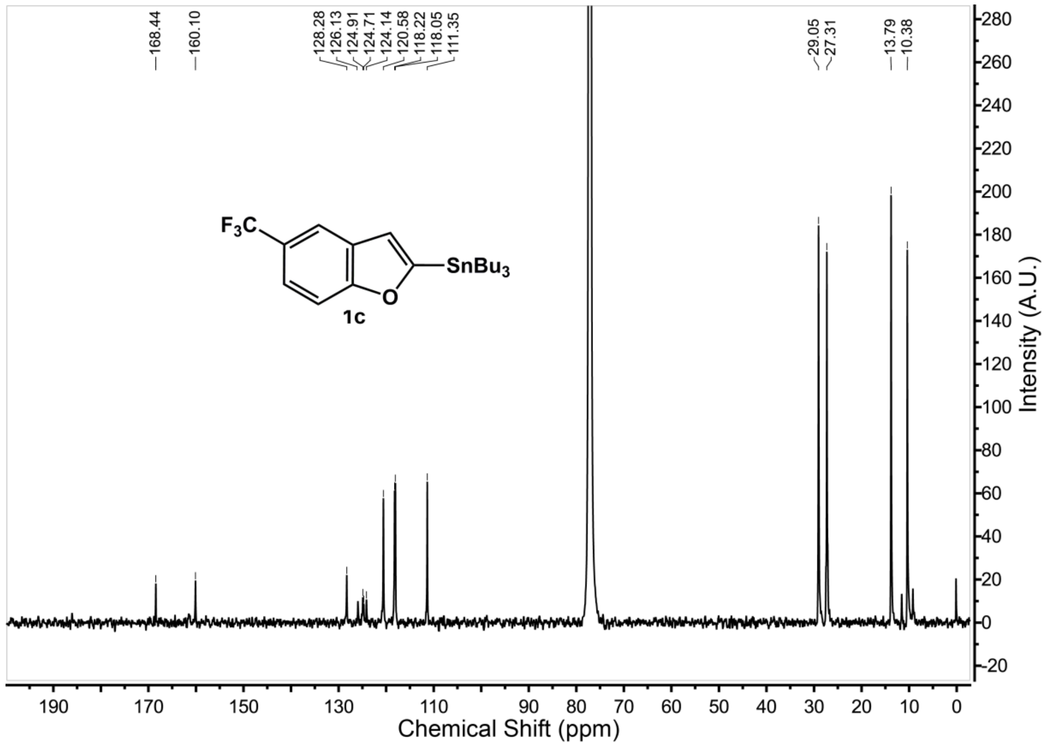
^{19}F NMR of **1b** (376.6 MHz, CDCl$_3$)



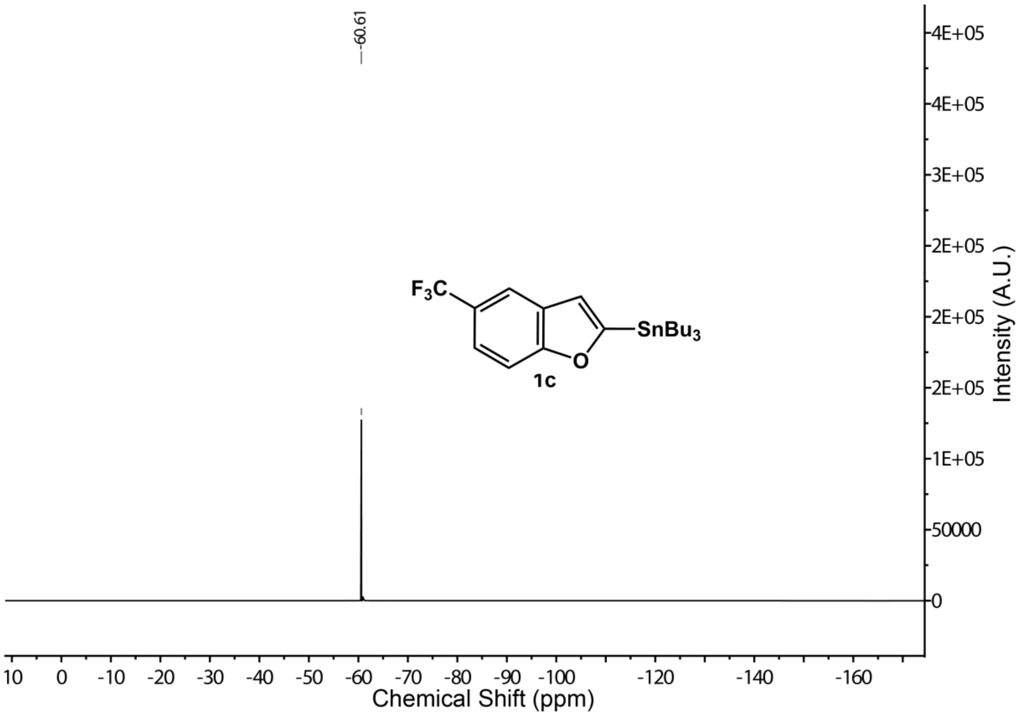
^{1}H NMR of **1c** (400 MHz, CDCl$_3$)



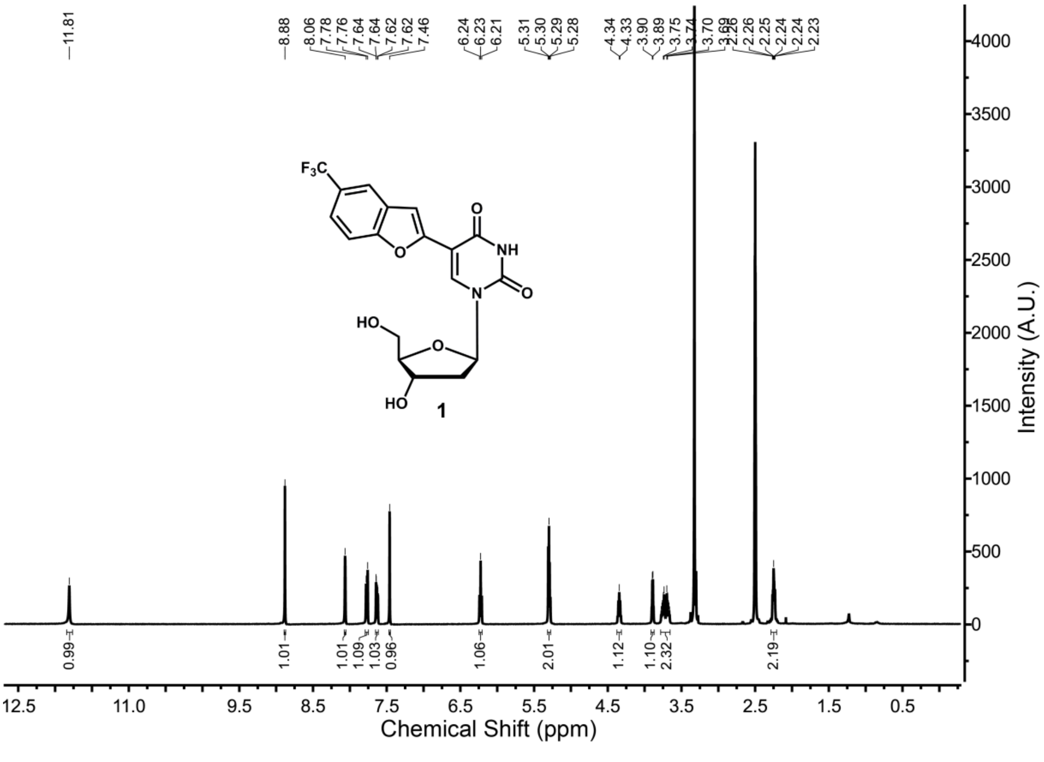
^{13}C NMR of **1c** (150 MHz, CDCl$_3$)



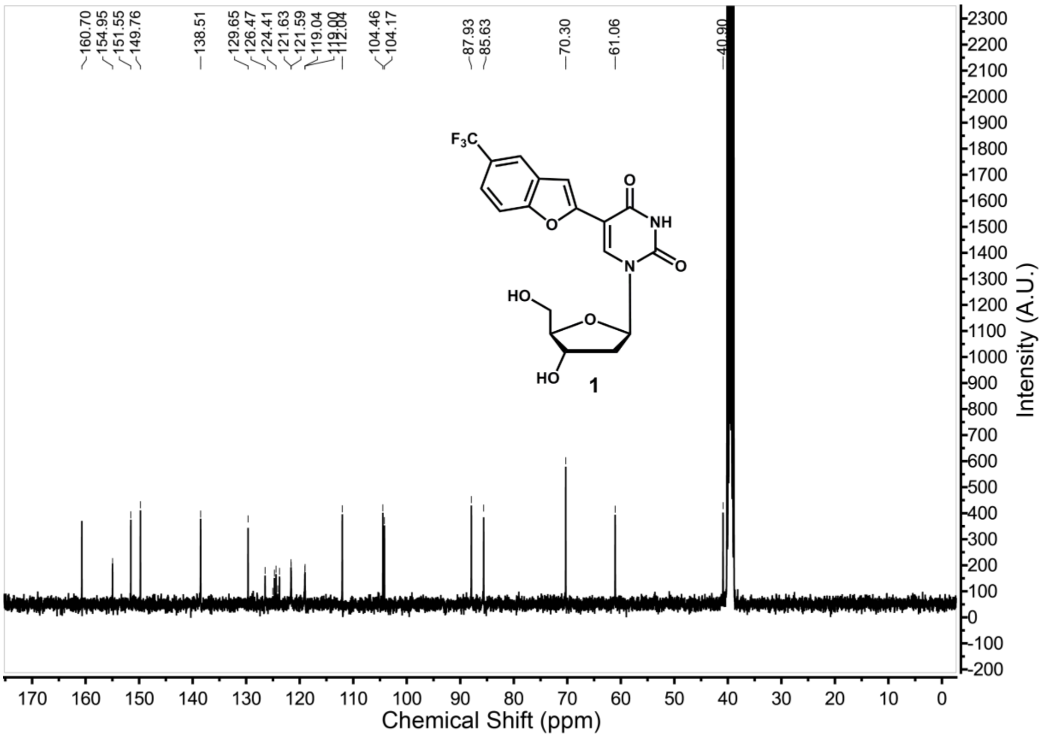
^{19}F NMR of **1c** (376.6 MHz, $CDCl_3$)



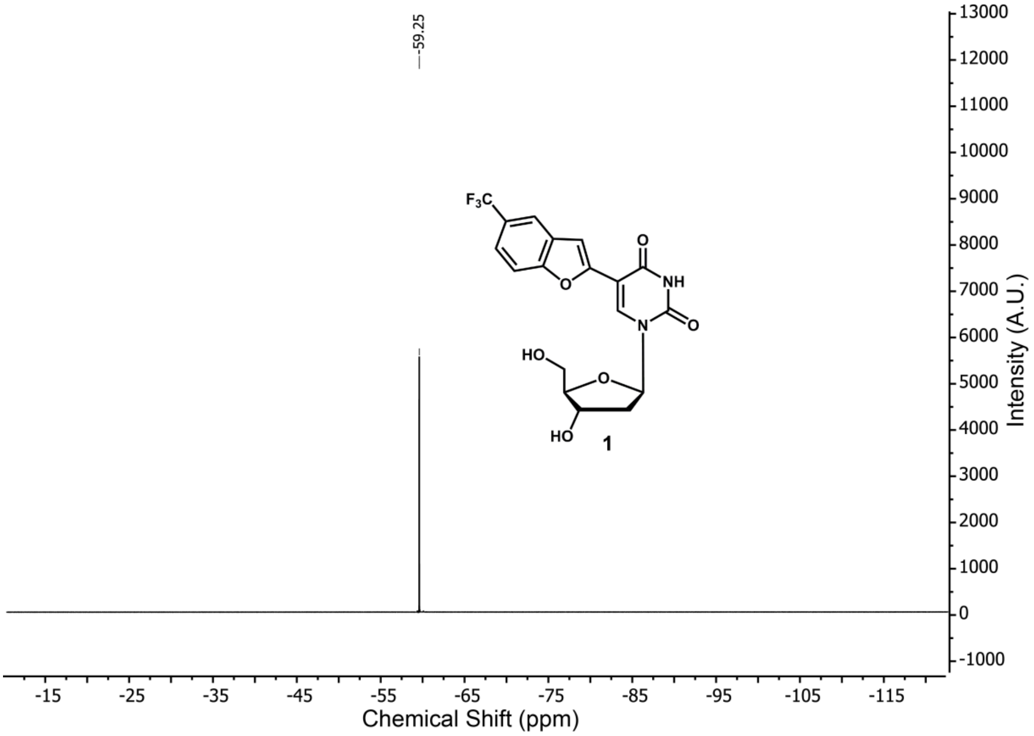
^{1}H NMR of **1** (400 MHz, *d$_6$*-DMSO)



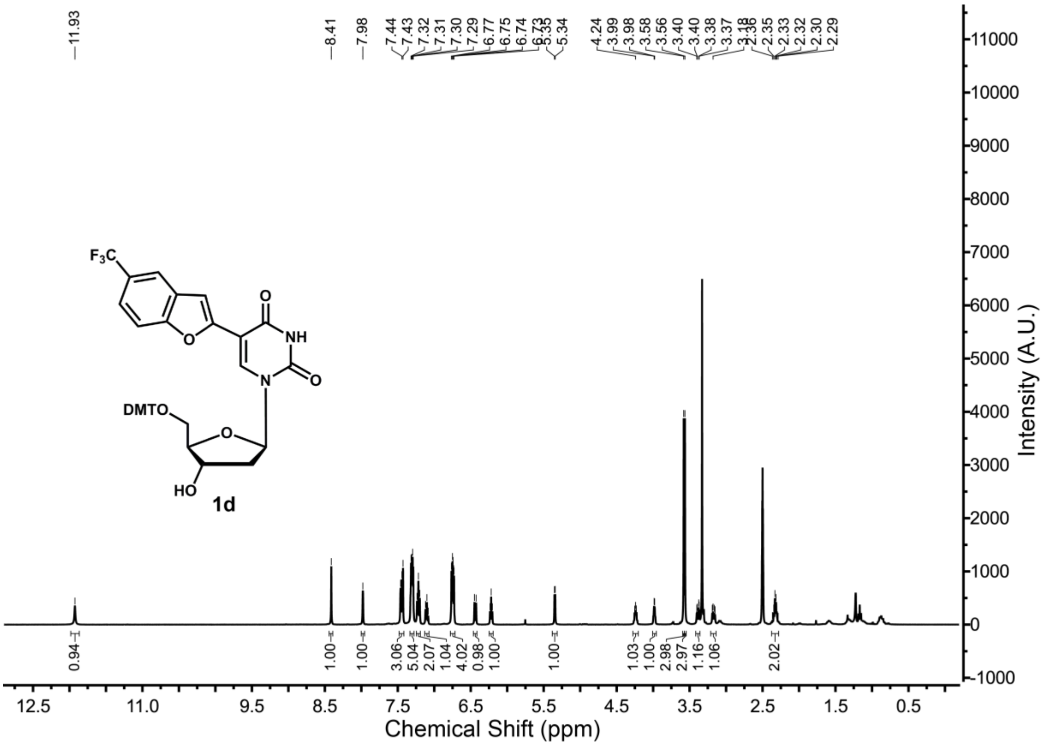
^{13}C NMR of **1** (100 MHz, d_6-DMSO)



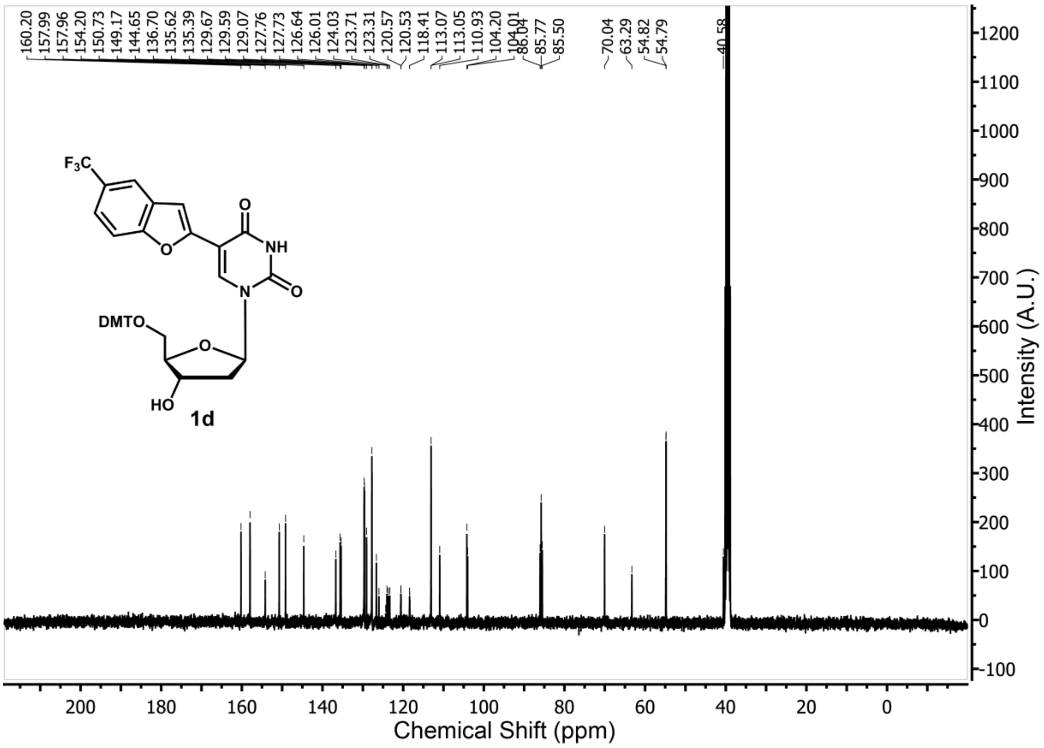
^{19}F NMR of **1** (376.6 MHz, d_6-DMSO)



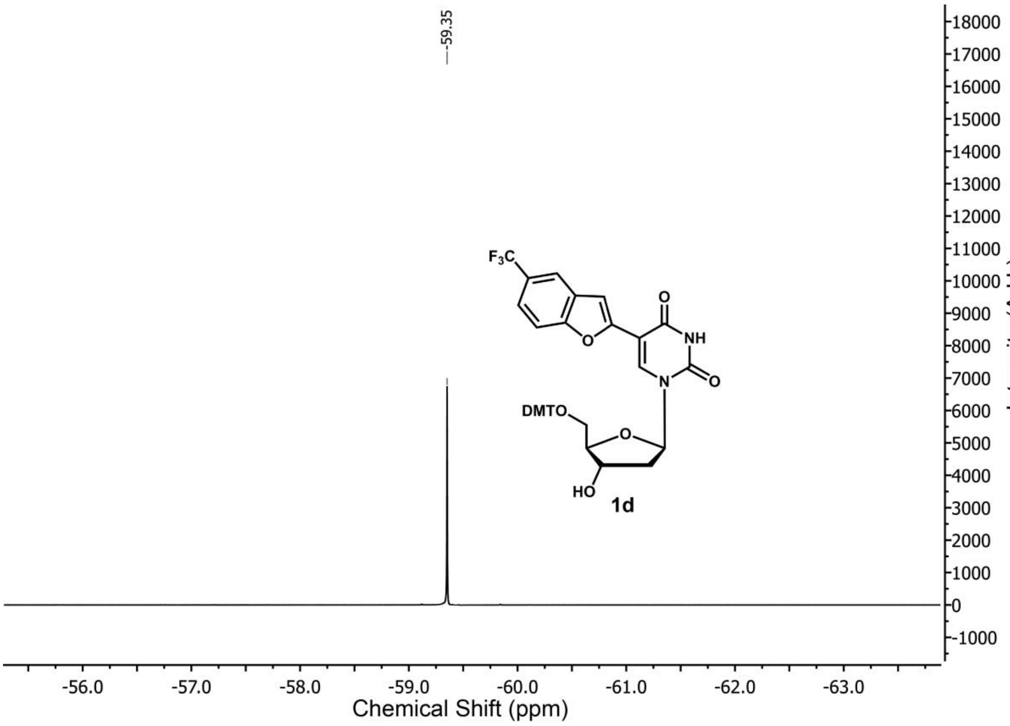
^{1}H NMR of **1d** (400 MHz, *d$_6$*-DMSO)



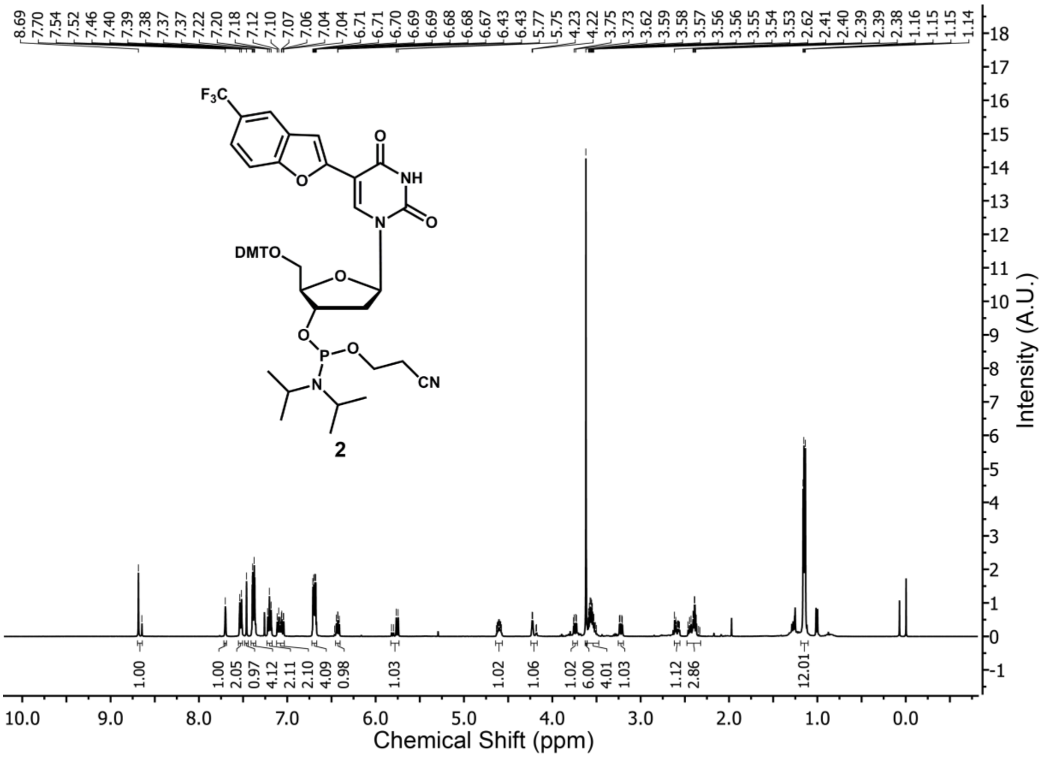
^{13}C NMR of **1d** (100 MHz, *d$_6$*-DMSO)



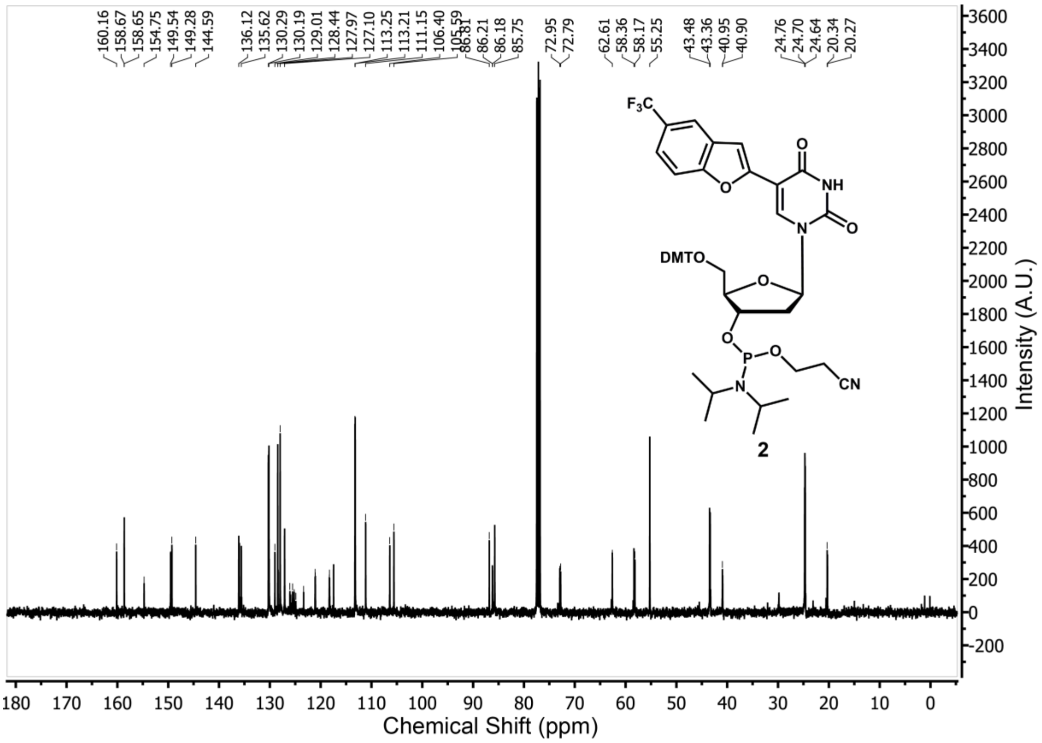
^{19}F NMR of **1d** (376.6 MHz, d_6-DMSO)



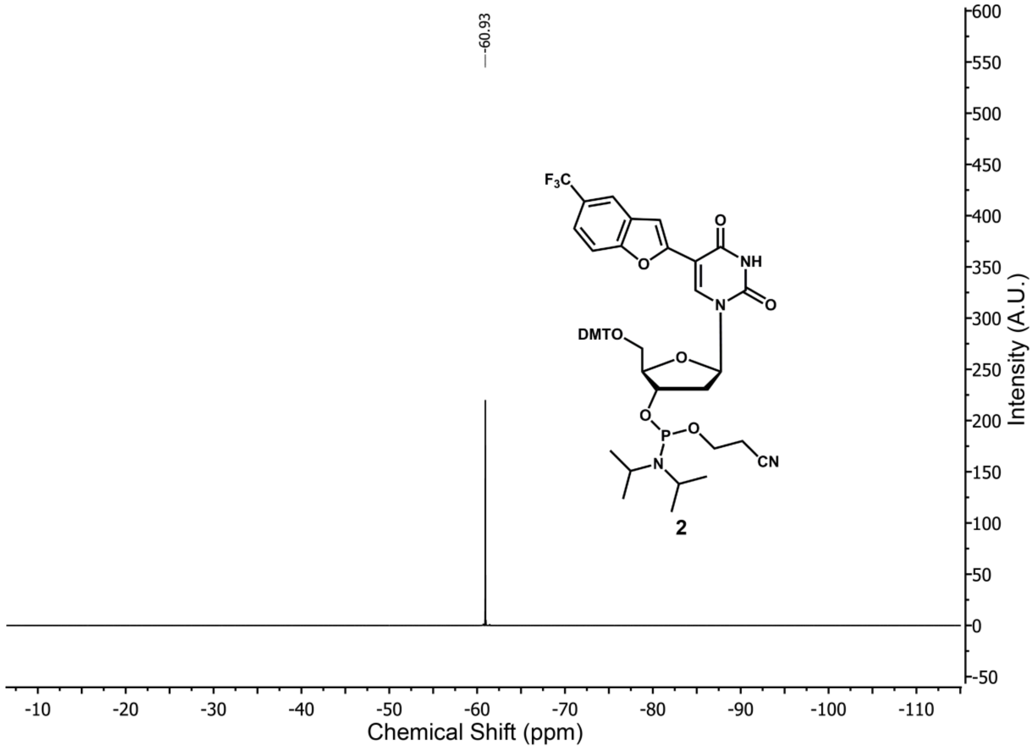
^{1}H NMR of **2** (400 MHz, CDCl$_3$)



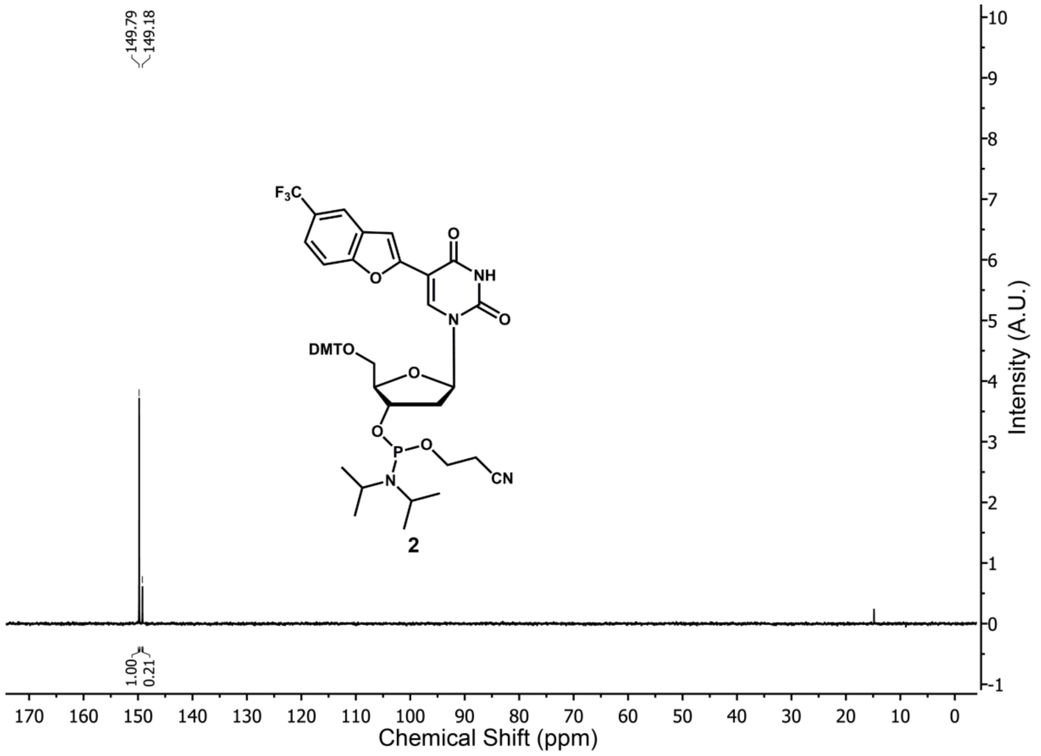
^{13}C NMR of **2** (100 MHz, CDCl$_3$)



^{19}F NMR of **2** (376.6 MHz, CDCl$_3$)

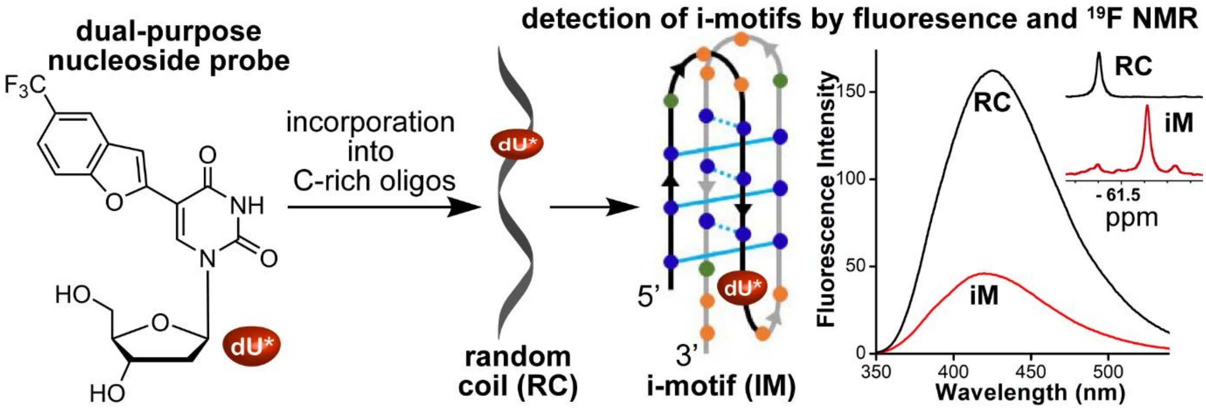


^{31}P NMR of **2** (162 MHz, $CDCl_3$)



Chapter 3:

Nucleoside probe reports the formation of human telomeric i-motif structures



Chapter 3 is a reprint of part of the data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.
The thesis author is the main author and researcher for this work.

3.1 Introduction

Nucleic acids, apart from classical double helical structures, adopt a number of different secondary structures, which enable them to perform several essential functions of the cell. Among the various secondary structures, four-stranded structures formed by guanine-rich sequences (G-quadruplexes) are the best studied.[1–4] G-quadruplex (GQ) forming sequences are widespread within the genomes and transcriptomes of humans and human pathogens.[5–7] Biochemical studies indicate that this class of structural element plays important roles in the conservation of chromosomes and in the regulation of tumor and neurodegenerative disease-related genes.[8–10] Complementary cytosine-rich sequences adopt another type of four-stranded structure called i-motif (iM) wherein two parallel stranded duplexes composed of hemiprotonated C-C$^+$ base pairs intercalate in an antiparallel orientation.[11,12] As one of the paired cytosines needs to be protonated, iM’s role in cellular processes under physiological pH was less anticipated. However, recent studies suggest that some C-rich sequences depending on the environmental conditions can form iMs at near physiological pH.[13,14] For example, negative superhelicity, molecular crowding, and certain metal ions and modifications support iMs at neutral pH.[15–19] Further, iM-binding small molecule ligands and proteins binding to C-rich domains regulate the promoter activity of some oncogenes.[20–26] Recent in-cell proton NMR spectroscopy of preformed iMs obtained from different human promoter regions, when transfected into cells, indicates that these structures are stabilized in live cells.[27,28] The seminal evidence for the existence of iMs in cells was obtained using a structure-specific antibody (iMab).[29] Notably, studies using iMab and a GQ-specific antibody (BG4)[30] indicate that the relative population of these structures vary with cell cycle, wherein iM formation is associated with transcriptional activation phase and GQ formation is observed during the replication phase. These results suggest a segregated biological role of GQs and iMs.

iM structure, stability and recognition are usually evaluated by CD, UV-thermal melting, fluorescence and to some extent by NMR and X-ray crystallography techniques.[13,14,31] Notably, fluorescence method have proven to be a powerful tool to monitor the nucleic acid conformational changes in real time.[32–35] In this direction, few fluorescent analogs have employed to investigate pH dependent formation of iM structure or iM to duplex structure exchange. Similarly, fluorine labeled nucleoside analogs have used to explore the nucleic acid structure transitions and interactions using the ^{19}F NMR spectroscopy.[36–40] Although much studies have revealed about dynamics and thermodynamics of iM structure *in vitro* conditions, the limited information is available on the effect of different conditions (pH, temperatures, metal ions) on the iM conformation equilibrium. This information could provide better understanding about the biological importance of iM structure or their application in nanotechnology.

Here, we have described the establishment of a biophysical platform using a microenvironment-sensitive dual-app nucleoside analog (Figure 1A, TFBF-dU **1**) to investigate the conformation equilibrium of human telomeric iM structures under different conditions. The nucleoside analog serves as a responsive fluorescent and ^{19}F NMR probe. When incorporated into the C-rich ON, it is minimally perturbing and photophysically reports the formation of iMs as a function of pH, thereby enabling the determination of transition pH_T value (50% folded state). Importantly, the ^{19}F label of the nucleoside exhibits distinct and reasonably resolved chemical shifts for various structures (random coil and different iM conformations), which allows systematic analysis of the structural dynamics of C-rich telomeric repeat under different conditions (pH and temperatures). In addition, the nucleoside probe is useful in monitoring silver ion-induced iM structure formation by fluorescence and ^{19}F NMR techniques.

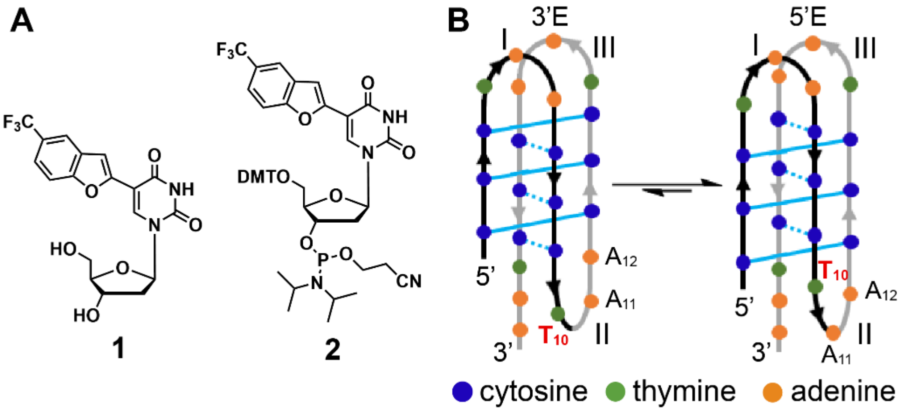


Figure 1. (**A**) Structure of the modified nucleoside **1** and it's corresponding phosphoramidite **2** and (**B**) 3′E (minor) and 5′E (major) iM conformations of native telomeric C-rich ON.

3.2 Results

3.2.1 Incorporation of the probe into iM forming sequence

To estimate the proficiency of the dual-purpose nucleoside analog to report the formation of different iM structures, we chose the human telomeric repeat C-rich segment. The telomeric DNA repeat $(CCCTAA)_n$ serves as a simple system because it forms one major iM structure composed of 6 intercalated C-C$^+$ base pairs under acidic conditions and a random coil (RC) at physiological pH.[31] Using this sequence, we systematically studied the formation and stability of different iM structures as a function of pH and temperature in intracellular ionic conditions. TFBF-modified telomeric DNA repeat ON **3** was synthesized by incorporating the phosphoramidite substrate **2** using solid-phase ON synthesis (Table 1). We purposely chose to replace a T nucleoside in one of the loop positions with the modified nucleoside so that it does not affect the formation and stability of C-C$^+$ intercalated structures. The native telomeric repeat ON forms two iM structures called 5'E and 3'E, which differ in the intercalation topology of the C-C$^+$ pair (Figure 1B).[31,41] In the major 5'E structure, the second loop formed by $T_{10}A_{11}A_{12}$ is more rigid than other loops as T_{10} stacks on the first C-C$^+$ base pair on the single-loop side of iM, and A_{11} stacks over T_{10} (Figure 2). The interaction of T_{10} with adjacent bases should be different in 5'E and 3'E forms (figure 1B), and hence, we replaced T_{10} with 2'-deoxyuridine analog **1** to produce ON **3**. Modified ON was purified by PAGE under denaturing conditions and the purity and integrity was confirmed by RP-HPLC and mass analysis (Figure 3 and 4, Table 2).

Table 1. TFBF-modified and respective control unmodified ON sequence.

ON[a]	5'------------sequence--------- 3'
3	CCC TAA CCC **T***AA CCC TAA CCC TAA
4	CCC TAA CCC TAA CCC TAA CCC TAA

[a]**T*** represents site of modification

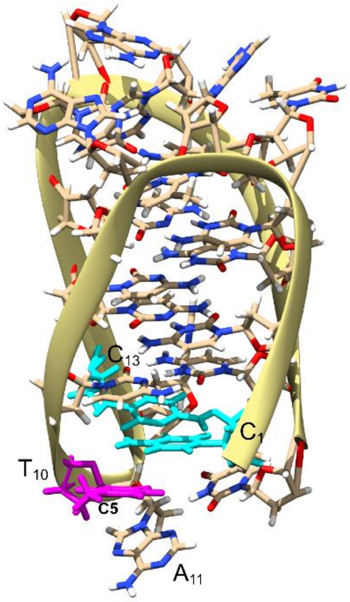


Figure 2. NMR structure of 5'E topology (PDB: 1ELN)[31] showing the conformation of T_{10} loop residue (magenta). It is solvent exposed and partially stacks on the base pair formed by cytosine C_1 and C_{13} (cyan). Figure was generated using UCSF Chimera.

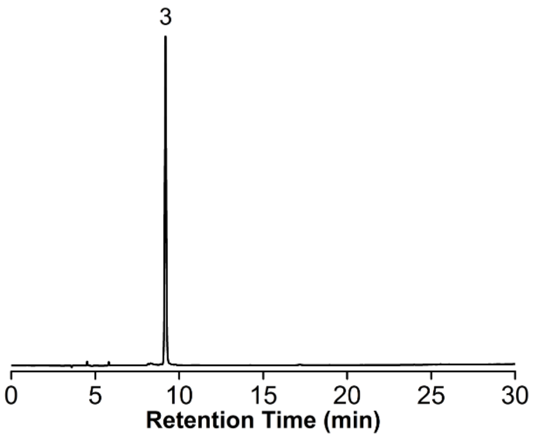


Figure 3. Reversed phase-HPLC chromatograms of PAGE purified ON **3** analyzed at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).

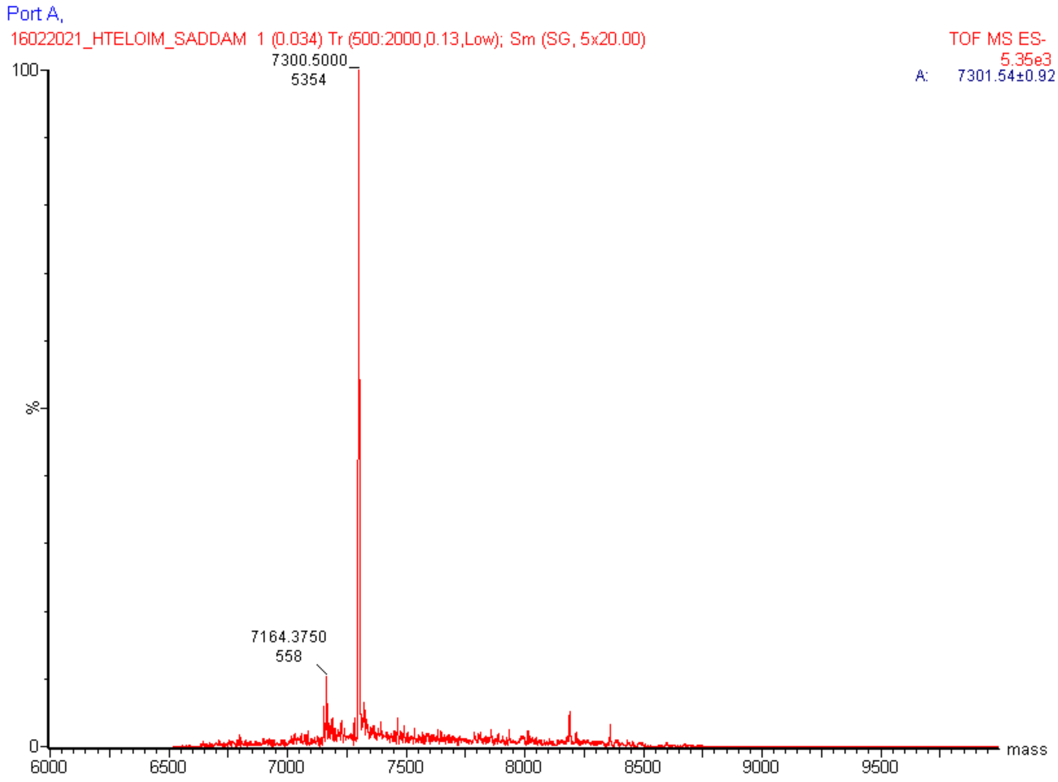


Figure 4. ESI-MS spectra of modified ON **3**.

Table 2. Molar absorptivity and mass of modified DNA ON.

DNA ON	$\varepsilon_{260}{}^{a}$ [M^{-1} cm^{-1}]	Calculated mass	Observed mass
3	223 x 10^3	7300.7	7300.5

[a] Molar absorptivity (ε_{260}) of modified ON was determined by using Oligo Analyzer 3.1. The molar absorptivity of modified nucleoside **1** (ε_{260} = 11.4 x 10^3 M^{-1} cm^{-1}) was used in the place of thymidine.

CD analysis of modified ON **3** and native ON **4** as a function of pH in buffers mimicking intracellular ionic conditions[42] reveals the transition from RC to iM form with a positive peak at 286 nm and a negative peak at 263 nm (Figure 5A).[43] The pH$_T$ values for modified ON and control unmodified ONs **3** and **4**, respectively, as determined from a plot of normalized ellipticity at 286 nm *versus* pH are very similar and consistent with the literature reports (Figure 5B, Table 3).[44,45] The melting temperatures of iM structure of modified ON **3** and unmodified ON **4** are nearly same at different pH (Figure 6A and 6B, Tables 4). Hysteresis between melting and annealing processes is observed for telomeric ON due to slow folding kinetics, which is more apparent near the pH$_T$ value. This observation suggests that the telomeric C-rich ON could adopt more than one iM form differing in folding kinetics and stabilities.[46,47] Collectively, it is clear that the modification has negligible impact on the iM structure formed by the telomeric C-rich ON.

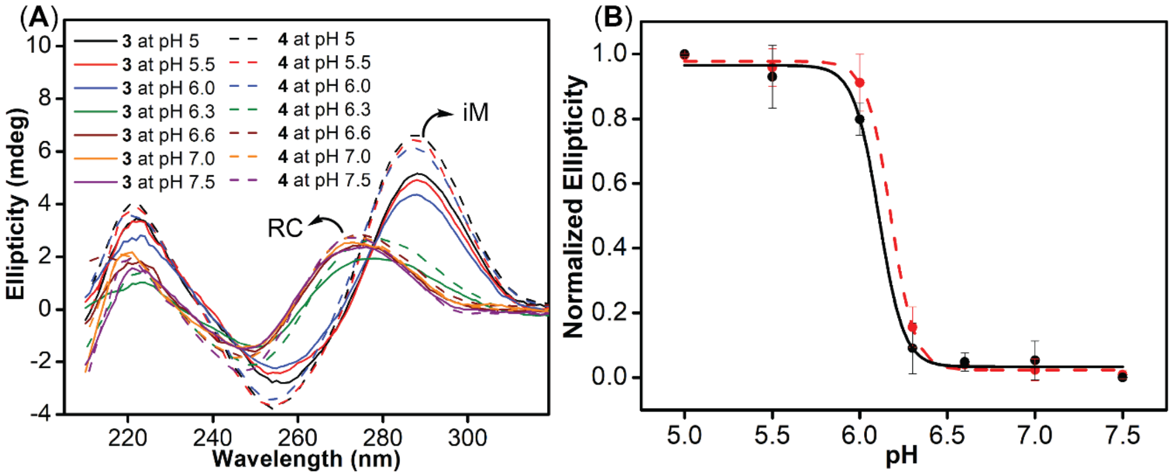


Figure 5. (**A**) CD spectra (5 µM) of modified telomeric ON **3** and control ON **4** at different pH. (**B**) Curve fitted to the plot of normalized ellipticity at 286 nm *versus* pH. Solid line is for modified ON and dashed line is for control ON.

Table 3. Transition pH (pH$_T$) values for modified and control unmodified ONs by CD and fluorescence.

ON	CD	fluorescence
modified **3**	6.1 ± 0.03	6.3 ± 0.03
control **4**	6.2 ± 0.01	-

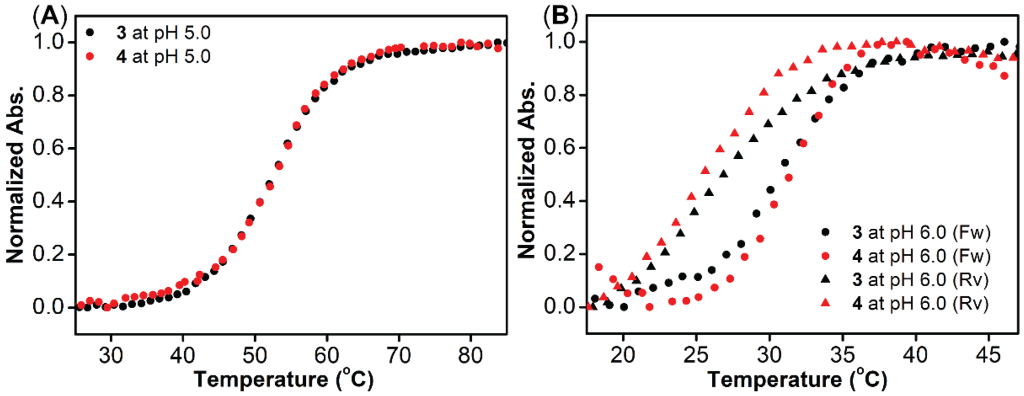


Figure 6. (**A** and **B**) UV-thermal melting profile of TFBF-modified telomeric DNA ON **3** and unmodified DNA ON **4** (1 µM) at pH 5.0 and 6.0, respectively. Changes in absorbance were measured at 260 nm.

Table 4. T_m values of TFBF-modified telomeric DNA ON **3** and control unmodified ON **4** at different pH.

ONs	pH 5.0	pH 5.5	pH 5.7		pH 6.0	
	T_m (°C)	T_m (°C)	T_m (°C) (Fw)	T_m (°C) (Rv)	T_m (°C) (Fw)	T_m (°C) (Rv)
3	54.5 ± 0.2	46.0 ± 0.1	38.8 ± 0.6	35.9 ± 0.1	31.1 ± 0.1	24.2 ± 0.6
4	53.3 ± 0.5	45.5 ± 0.4	38.3 ± 0.1	35.3 ± 0.6	30.6 ± 0.5	24.4 ± 0.6

3.2.2 Nucleoside probe reports iM structures of the telomeric repeat ON by fluorescence and ^{19}F NMR

Samples of C-rich telomeric DNA ON **3** annealed in intracellular ionic conditions at different pH were excited at 330 nm and changes in emission profile were recorded. ON **3** forms a RC structure in the pH range of 7–8. This structure displayed an intense emission band centered around 425 nm (Figure 7A). As the pH was reduced from 8 to 5, a significant reduction in the fluorescence intensity with a minor change in the emission maximum was observed. The emission intensity saturated at pH 6.0. Normalized fluorescence intensity at 425 nm plotted as a function of pH gave a pH_T value of 6.3 ± 0.03, which is similar to the value obtained from CD studies (Figure 7B, Table 3). In a control experiment, the fluorescence of the free nucleoside analog **1** did not change at different pH (Figure 8), which indicates that the difference in fluorescence profile displayed by the probe in RC and iM forms is due to the conformational change in the ON. Based on the emission maximum of the free nucleoside analog in different solvents (λ_{em} in water is 427 nm, Chapter 2 - Table 1), it appears that a water-like micropolarity surrounds the nucleoside analog in the iM structure. Further, stacking interaction between the nucleoside analog and adjacent bases can quench the fluorescence intensity upon iM formation. In the NMR structure of the same sequence, the C5-methyl group of T_{10} loop residue located in the groove is solvent exposed and the thymine ring partially stacks on the base pair formed by the cytosines (Figure 2).[31] C5 TFBF-modified analog placed in this position will also experience a similar solvent polarity and stacking interaction with the adjacent C-C$^+$ pair, which is manifested as a weak emission band at 425 nm.

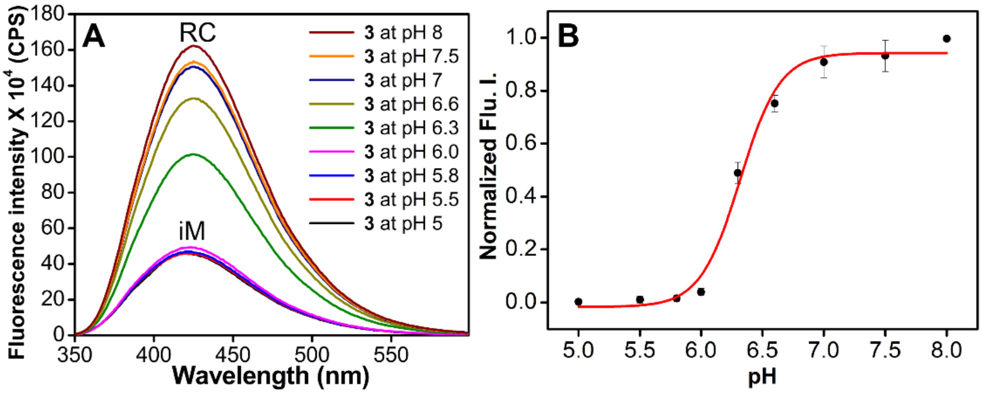


Figure 7. (**A**) Fluorescence spectra of telomeric ON **3** (1 μM) as a function of pH. Samples were excited at 330 nm with excitation and emission slit widths of 3 nm and 4 nm, respectively. (**B**) Curve fit for the plot of normalized fluorescence intensity at 425 nm *versus* pH.

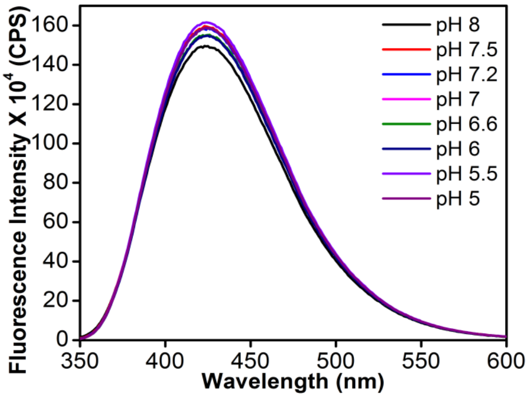


Figure 8. Fluorescence spectra of free TFBF-dU (10 μM) at different pH (5.0 to 8.0). Samples were excited at 320 nm with excitation and emission slit width of 2 nm and 3 nm, respectively.

iMs typically show multiple signals for the imino protons of C-C$^+$ pairs between 15 to 16 ppm, whereas RCs do not show any signal in this region.[31,41,48] Both modified (**3**) and native unmodified (**4**) ONs under acidic conditions produced similar ^{1}H NMR spectra reminiscent of an iM structure (Figure 9 bottom panel and Figure 10). Further, ^{1}H NMR analysis of ON **3** as a function of decreasing pH (7.5 to 5) revealed the transformation of a RC to an iM structure (Figure 9A–9C, bottom panel). The telomeric repeat is known to adopt 5'E (major) and 3'E (minor) iM conformations (population ratio is 4:1), which however is not apparent from the ^{1}H NMR of unlabeled DNA and CD spectrum.[48] Rewardingly, the ^{19}F label of our dual-app probe exhibits a unique signature for each structure adopted by the ON. Using these signatures, we were able to study the conformational changes as a function of pH and temperature, and ascertain the relative population of different structures formed by the C-rich telomeric repeat ON (Figure 9A–9C top panel). ON **3** at 18°C and pH 7.5 exhibits a ^{19}F signal at -61.52 ppm corresponding to a RC structure, which is also consistent with the ^{1}H NMR spectrum and CD profile (Figure 9A). At pH 5.0, where it is known to form two iM structures, we observed three signals–one corresponding to the RC (-61.50 ppm, minor peak) and two upfield-shifted peaks corresponding to iM topologies (Figure 9A, top panel). The major peak (-61.72 ppm) represents 5'E iM form and the minor peak (broad, -61.82 ppm) corresponds to 3'E form. An upfield shift as a result of π-stacking interaction experienced by the probe placed at T_{10} position nicely correlates with the iM structure and quenched emission observed in the fluorescence experiment.[49,50] At near pH_T (6.0 and 6.2) the amount of RC form increases with concomitant decrease in iM forms with a significant amount of 5'E structure still present. As the temperature is increased from 18°C to 25°C and to 35°C, the population of RC at intermediate pH (6.0 and 6.2) increases significantly with practically no iM structures at pH 6.2 and 35°C (Figure 9B and 9C). It is important to mention here that a concentration-dependent ^{19}F NMR analysis of ON **3** at pH 5.0 and at different temperatures gave the same set of peaks indicating the formation of intramolecular iM structures and not intermolecular or higher order structures (Figure 11).

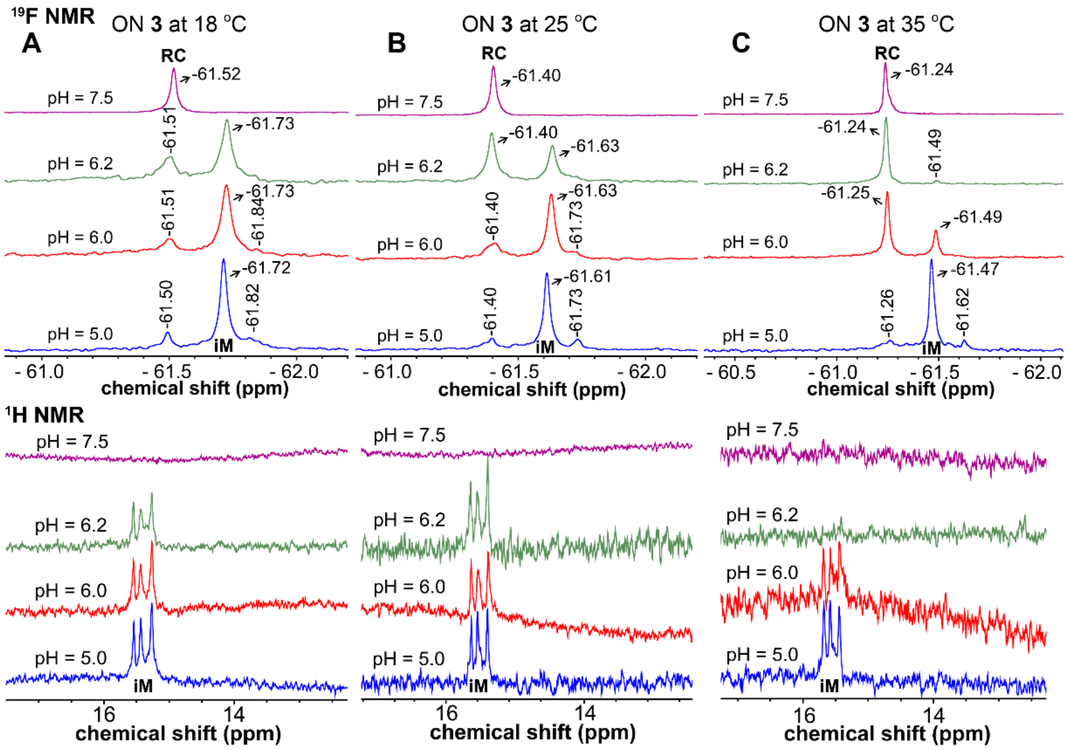


Figure 9. (**A** to **C**) Partial ^{19}F NMR (top panel) and ^{1}H NMR (bottom panel) spectra of modified telomeric ON **3** (10 µM) at different pH and temperatures. The major iM structure is denoted as iM in the ^{19}F NMR spectra.

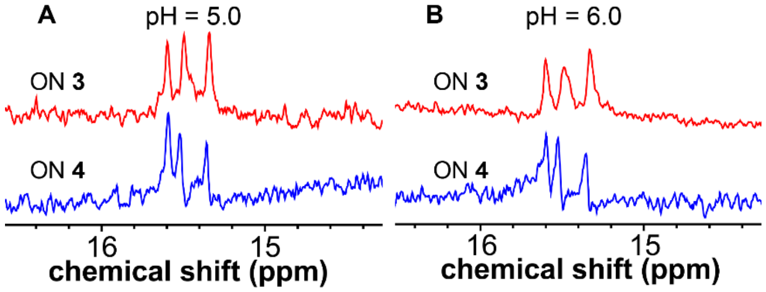


Figure 10. Partial ^{1}H NMR spectra of TFBF-modified ON **3** and control unmodified telomeric ON **4** at pH 5.0 and 6.0 (25°C). Concentration of the ONs is 10 µM.

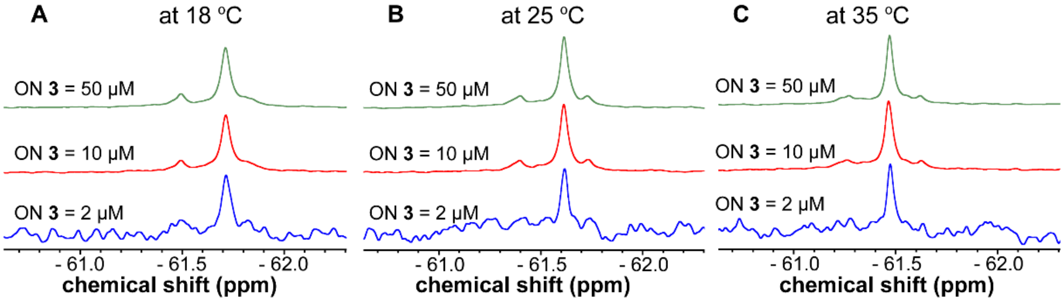


Figure 11. Concentration-dependent ^{19}F NMR spectra of ON **3** at pH 5.0 and at different temperatures.

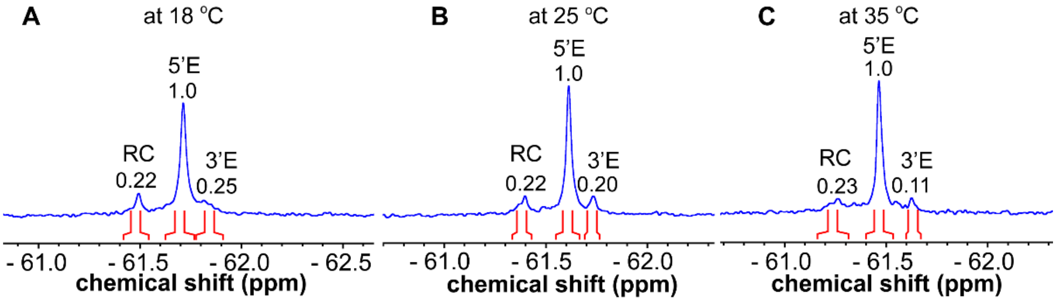


Figure 12. Partial ^{19}F NMR spectra of telomeric ON **3** (10 µM) at pH 5.0 and at 18°C, 25°C and 35°C. RC is random coil. 5'E (major) and 3'E (minor) are iM conformations.

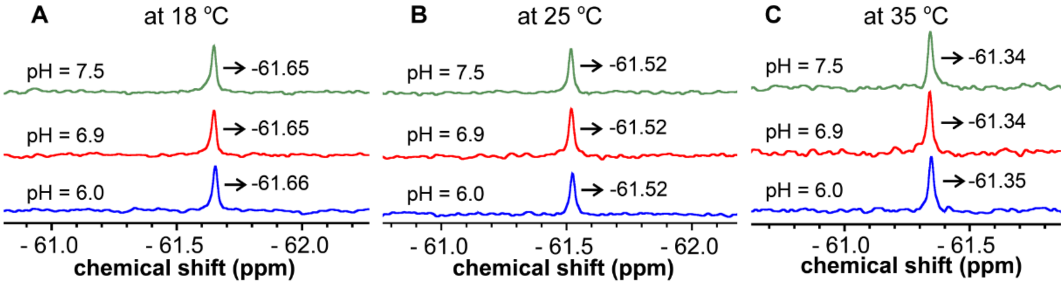


Figure 13. Partial ^{19}F NMR spectra of TFBF-modified nucleoside analog **1** (5 µM) at different pH and temperatures. Changes in pH at a given temperature did not affect the chemical shift of the free nucleoside probe.

Integrating the area of ^{19}F peaks provided direct information on the population equilibrium (Figure 12). At pH 5.0, the relative population of 5'E conformation was found to increase from 68% to 75% with increase in temperature. However, the population of 3'E conformation decreased from 17% to 8% in the above temperature range. This trend in the population equilibrium as a function of temperature is in agreement with the trend observed using a ^{15}N- and ^{13}C-labeled ON sequence.[48] There was no detectable change in RC structure at this pH over different temperatures. These results suggest that telomeric 5'E iM structure is more stable, and hence, more amounts of this conformation is formed at a higher temperature at the expense of the less stable 3'E iM structure. It is worth noting that a slight shift in the individual peaks at different temperatures is due to temperature effect, which was confirmed by recording the NMR of the free nucleoside at different temperatures and pH values (Figure 13). Minor changes in ^{19}F chemical shift due to temperature has been observed earlier also.[51] However, changes in pH at a particular temperature did not affect the chemical shift of the free nucleoside analog indicating that distinct signatures exhibited by the ^{19}F label is due to local conformational differences in the structures adopted by the ON.

3.2.3 Nucleoside probe reports silver ion-induced iM structure

Recent studies show that C-rich DNA ONs can fold into an iM structure in the presence of Ag(I) cations at physiological pH *via* C-Ag$^+$-C base pair formation, which can be unfolded using a chelating agent, namely cysteine.[17] The RC form of ON **3** at pH 7.4 shows high fluorescence intensity, which upon addition of Ag(I) ions displays a significant reduction in intensity as a result of formation of an iM structure (Figure 14A). Addition of cysteine sequestered the Ag(I) ions resulting in the unfolding of ON **3** and concomitant recovery of the fluorescence intensity. However, iM structures formed by intercalated C-C$^+$ pairs at pH 5.0 are not affected by Ag(I) ions (red traces). CD performed at pH 7.4 and 5.0 in the absence and presence of the metal ion using unmodified and TFBF-modified ONs **4** and **3**, respectively, indicate the formation of Ag(I)-mediated iM structure only at pH 7.4, similar to the results obtained from fluorescence study (Figure 14B and 14C). The Ag(I)-induced iM structure of ON **3** produced a distinct ^{19}F signal at -61.34 ppm as compared to the RC form (-61.45 ppm, Figure 15A). Addition of excess of cysteine unfolded the iM structure to give the RC structure. In a control experiment, Ag(I) ions did not affect the iM structures formed at acid pH, which is evident from no changes in chemical shifts of the iMs (Figure 15B). Further, addition of the metal ion and cysteine to the free nucleoside **1** did not affect ^{19}F signal indicating the specific detection of Ag(I)-mediated iM formation by the nucleoside probe (Figure 16).

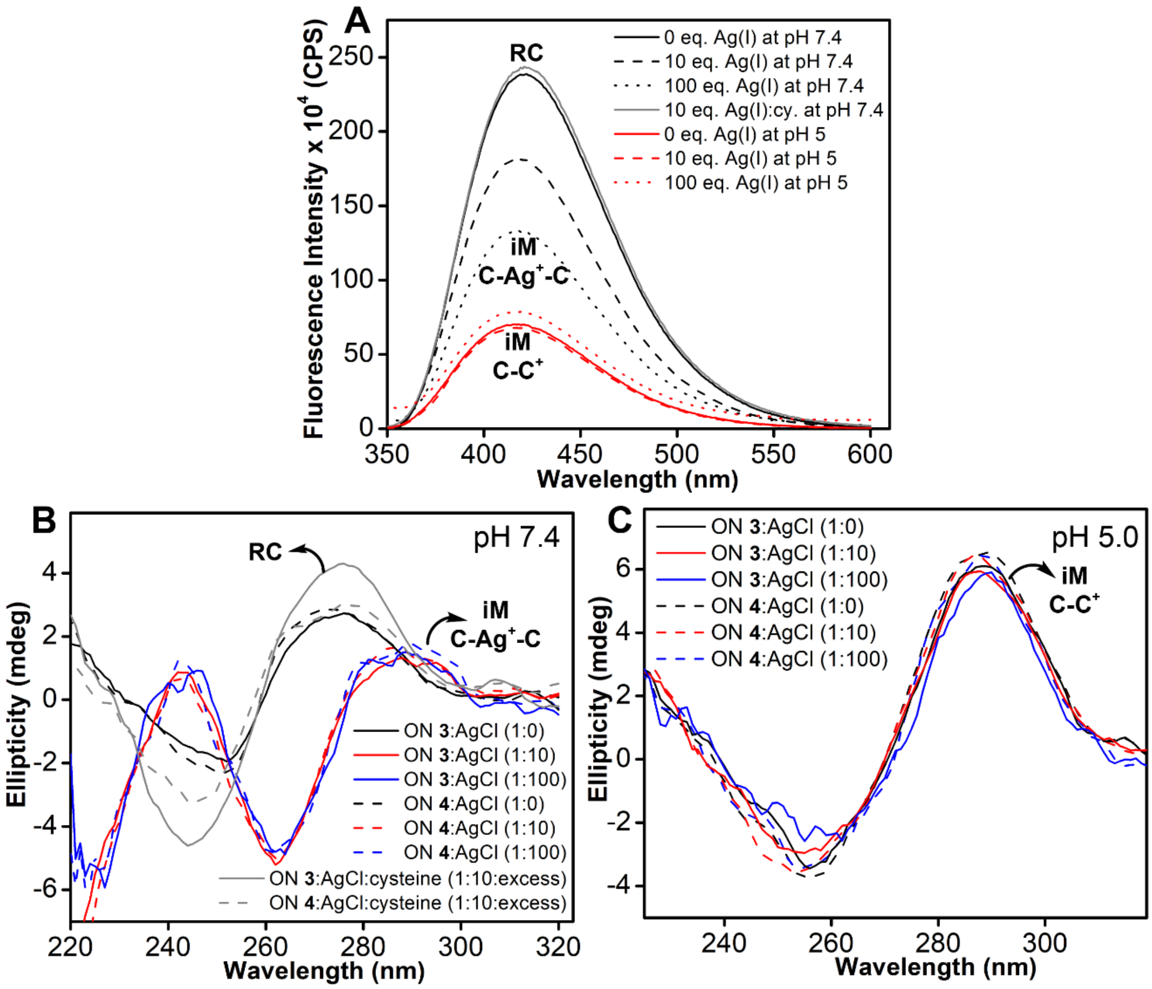


Figure 14. (**A**) Fluorescence detection of the formation of an Ag(I) ion-mediated iM structure of ON **3** (0.5 µM). Samples were excited at 330 nm with excitation and emission slit widths of 5 nm and 5 nm, respectively. (**B** and **C**) Control CD experiments at pH 7.4 and 5.0 gave similar results as fluorescence. Concentration of ON **3** was 5 µM.

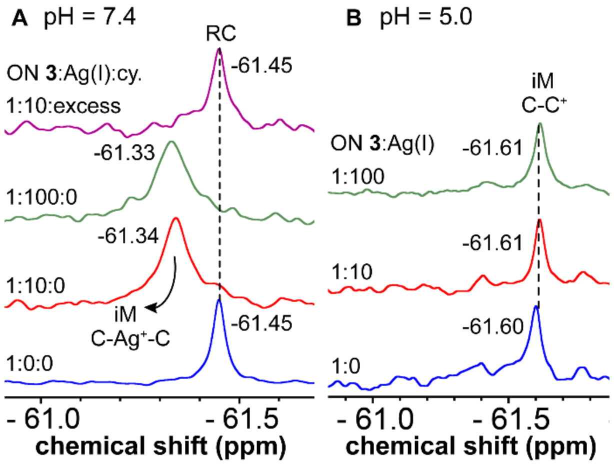


Figure 15. Detection of Ag(I) ion-mediated iM structure formation by ON **3** (5 µM) using ^{19}F NMR (**A**) at pH 7.4 and (**B**) at pH 5.0, respectively.

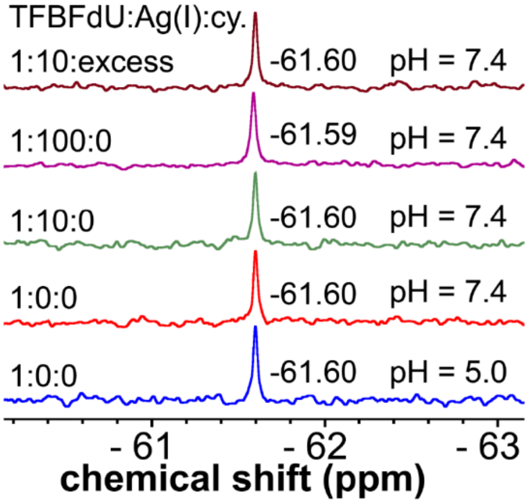


Figure 16 Addition of Ag(I) ions to the free nucleoside analog (5 µM) did not affect the ^{19}F chemical shift.

3.3 Discussion

TFBF-modified 2'-deoxyuridine analog **1** serves as a unique addition to the biophysical tool box, wherein the fluorescence component helps in ensemble analysis (monitoring the iM formation and determination of pH_T) and the ^{19}F label provides distinct signatures for different iMs and enables a systematic analysis of the iM population equilibrium under different conditions. The incorporation of nucleoside analog into C-rich ON sequence is minimally perturbing and does not affect the formation and stability of iM structures. The probe incorporated into the telomeric repeat ON **3** reports the formation of iM structures with reduction in fluorescence intensity, thereby enabling the determination of transition pH for the conversion of iM to RC structure (Figure 7A and 7B). Notably, the ^{19}F label gives unique signatures for all possible structures of ON **3** (RC, 5'E iM and 3'E iM, Figure 9A), which is not easy to delineate by conventional techniques like CD, thermal melting and 1D NMR (requires expensive isotope-labeled ONs for 2D NMR). ^{19}F signatures were further used to estimate the relative population of different forms adopted by telomeric C-rich sequence as a function of pH and temperature in intracellular ionic conditions. Adding to the versatility of our probe, we could also study the silver ion-mediated folding of a C-rich ON by both fluorescence and ^{19}F NMR techniques.

3.4 Conclusions

In summary, we developed a biophysical assay using highly sensitive dual-app nucleoside analog, which functions as a very useful probe in detecting various iM structures by fluorescence and ^{19}F NMR spectroscopy techniques. The probe is minimally perturbing and able to fluorescently report the transition from RC structure to iM form as a function of pH enables the determination of the pH_T value for the human telomeric repeat ON. Interestingly, the modified nucleoside exhibits unique ^{19}F signatures for RC and different iM structures, which enables their quantifications under different conditions. Based on our results, it is clear that the analog can operate under different conditions (e.g., pH, temperature, metal ions). Collectively, our dual-app probe is very important addition to the available biophysical tools to quantitatively probe the dynamics of coexisting nucleic acid structures under different conditions.

3.5 Experimental Section

3.5.1 Materials: Modified nucleoside analog **1** and it's corresponding phosphoramidite **2** for solid-phase DNA synthesis were synthesized using a procedure described in chapter 2. All other reagents/salt (Bio-Ultra grade) used for buffer preparation were purchased from Sigma-Aldrich. *N*-Benzoyl-protected dA, *N*-acetyl-protected dC, *N*,*N*-dimethylformamide-protected dG, and dT phosphoramidite substrates required for solid-phase synthesis were purchased from ChemGenes. Solid supports for DNA synthesis were obtained from Glen Research. All other reagents required for solid phase oligonucleotide (ON) synthesis were purchased from ChemGenes or Sigma-Aldrich. Control DNA ON **4** was purchased from Integrated DNA Technology, purified by denaturing polyacrylamide gel electrophoresis (PAGE) and desalted using Sep-Pak Classic C18 cartridges (Waters Corporation). Autoclaved Millipore water was used for the preparation of all buffer solutions and in all biophysical studies.

3.5.2 Instruments: NMR spectra of small molecules were recorded on a 400 MHz Jeol ECS-400 and Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. Mass analysis was performed using ESI-MS Waters Synapt G2-Si Mass Spectrometry instrument. Modified DNA ON was synthesized on an Applied Biosystems DNA/RNA synthesizer (ABI-394). HPLC analysis was done using Agilent Technologies 1260 Infinity HPLC. Absorption spectra was recorded on a UV-2600 Shimadzu spectrophotometer. Fluorescence of the ON was recorded in a Fluoromax-4 spectrophotometer (Horiba Scientific). 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 ON was recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT) and processed using Bruker TopSpin Software.

3.5.3 Solid-phase DNA ON synthesis: TFBF-modified DNA ON **3** was synthesized in 1 µmole scale (1000 Å CPG solid support) using an ABI-394 DNA/RNA synthesizer. Phosphoramidite **2** was incorporated in the sequence at an appropriate position. At the end of the synthesis, DMT protection was removed and the ON was treated with 30% aqueous ammonium hydroxide solution for 20 h at 55°C and further for 30 min at 70°C. Sample was centrifuged, the supernatant was evaporated and the ON was purified by denaturing polyacrylamide gel electrophoresis (PAGE) (20% gel). The band corresponding to the modified ON product was identified by UV shadowing, which was then cut and the ON was extracted using aqueous ammonium acetate (0.5 M, 3 mL) in a poly-prep column (Bio-Rad) for 12 h. The ON samples were desalted using Sep-Pak C-18 cartridges. The purity of modified ON was confirmed by RP-HPLC and the integrity was confirmed by ESI-MS analysis.

3.5.4. Mass analysis of modified ON. ESI-MS analysis was performed in negative mode by injecting DNA ON (~300 pmol) dissolved in 50% acetonitrile in an aqueous solution of 10 mM triethylamine and 100 mM hexafluoro-2-propanol.

3.5.5 Circular dichroism (CD) analysis. Samples of DNA ONs (5 µM) at different pH (MES 5.5–6.6 or HEPES 6.8–7.5) in intracellular ionic conditions (25 mM HEPES/MES buffer, 110 mM KCl, 10.5 mM NaCl, 1 mM $MgCl_2$, 0.13 mM $CaCl_2$, 0.1 mM EDTA) were prepared and annealed by heating at 90°C for 3 min and later cooled slowly to RT. Samples were incubated at RT for 1 h. CD spectra were recorded from 310 nm to 220 nm on a JASCO J-815 CD spectrometer at 25°C using 1 nm bandwidth. All experiments were performed in triplicate with an average of five scans for each sample. The spectrum of buffer was subtracted from all ON sample spectra.

3.5.6 UV-thermal melting analysis: Samples of DNA ONs (1 µM) were prepared in buffers of different pH (10 mM potassium acetate pH 5.0–5.8 and 10 mM potassium phosphate pH 5.8–6.9) containing100 mM KCl. Samples were annealed by heating at 90°C for 3 min and cooled slowly to RT. Samples were incubated at RT for 1 h and used for thermal melting analysis. Measurements were performed on a Cary 300 Bio UV-Vis spectrophotometer. Temperature was increased from 15°C to 90°C at 1 °C min^{-1} and the absorbance was measured every 1°C at 260 nm. Forward (Fw) and reverse (Rv) cycles were used to determine the T_m values.

3.5.7 Fluorescence of modified ON: Samples of TFBF-modified ON **3** (1 μM) at different pH were prepared in either MES buffer (pH 5.0 to 6.6) or HEPES (pH 6.8 to 8.0) buffer in intracellular ionic conditions (25 mM HEPES/MES buffer, 110 mM KCl, 10.5 mM NaCl, 1 mM $MgCl_2$, 0.13 mM $CaCl_2$, 0.1 mM EDTA). ON samples were annealed in respective buffers by heating the samples at 90°C for 3 min and allowing them cool to RT. Samples were further incubated at RT for 1 h. Fluorescence analysis was performed by exciting the samples at 330 nm with excitation and emission slit widths as mentioned in the figure caption. All fluorescence experiments were performed in triplicate in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on a Fluoromax-4 spectrophotometer (Horiba Jobin Yvon).

Transition pH (pH_T) determination. The pH at which 50% of the ON remains in the folded state (iM) is considered as pH_T. Transition pH of modified ON **3** was determined by plotting normalized fluorescence intensity *versus* pH at its emission maximum (425 nm). pH_T values from CD data were determined by plotting normalized ellipticity at 286 nm *versus* pH. Both fluorescence and CD data were curve fitted using a sigmoidal function of the Boltzmann type (OriginPro. 8.5.0). The χ^2 (goodness of fit) values were very close to unity for all the plots.

3.5.8 ^{19}F and ^{1}H NMR analysis of telomeric ON at different pH: Samples of TFBF-modified ON **3** (10 μM) were prepared at different pH (MES 5.5–6.6 or HEPES 6.8–7.5) in intracellular ionic conditions (see above) containing 20% D_2O. Samples were annealed by heating at 90°C for 3 min and cooled slowly to RT. Further, samples were incubated at RT for 1 h and transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F and ^{1}H NMR spectra were recorded at a frequency of 564.9 MHz and 600 MHz, respectively, on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT). All ^{19}F NMR spectra were referenced relative to an external standard, trifluorotoluene (TFT = −63.72 ppm). ^{19}F and ^{1}H NMR spectra of ON were recorded at 18°C, 25°C and 35°C. Spectral parameters for ^{19}F NMR: excitation pulse: 12 μs; spectral width: 29.90 ppm; transmitter frequency offset: -60.00 ppm; acquisition time: 0.1 s; relaxation delay: 1.0 s; number of scans = 1024. Using these parameters, spectra were obtained in 20–25 min. Spectra were processed with an exponential window function using lb = 5 Hz. ^{1}H NMR spectra were recorded with water suppression using excitation sculpting with gradients. Number of scans were in the range of 400–600.

3.5.9 Detection of Silver ion-induced iM formation: Samples of TFBF-modified telomeric ON **3** and control unmodified ON **4** were prepared in 10 mM sodium cacodylate buffer (pH 5.0 and 7.4) containing 5 mM NaCl. Samples were annealed as mentioned above. CD spectra of ONs (5 µM) in the absence and presence of $AgNO_3$ (0, 10 and 100 equiv.) were recorded from 320 nm to 220 nm using 1 nm bandwidth at 25°C. Cysteine (2 mM) was added to unfold the iM. CD measurements were performed in duplicate with an average of three scans for each sample. Similarly, fluorescence analysis of ON **3** (0.5 µM) was performed by exciting the samples at 330 nm with excitation and emission slit widths of 5 nm and 5 nm, respectively. 0.2 mM of cysteine was added to unfold the iM. Fluorescence experiments were performed in duplicate. ^{19}F NMR spectra of ON **3** (5 µM) in the absence and presence of $AgNO_3$ (0, 10 and 100 equiv.) were recorded in the above buffer containing 20% D_2O (500 scans). Cysteine (2 mM) was added in the unfolding experiment. All the spectra parameters are same as previously mentioned and the data was processed with an exponential window function using lb = 15 Hz.

3.6 References

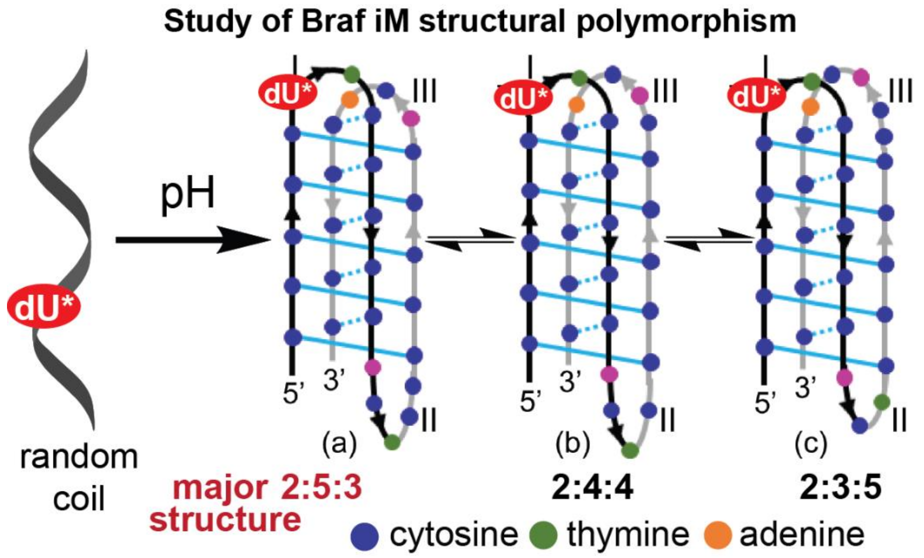
1. Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol*. **2003**, *13*, 275–283.
2. Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat. Rev. Genet.* **2012**, *13*, 770–780.
3. Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 279–284.
4. Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627–8637.
5. Bedrat, A.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res*. **2016**, *44*, 1746–1759.
6. Sengupta, A.; Roy, S. S.; Chowdhury, S. *Acc. Chem. Res.* **2021**, *54*, 46–56.
7. Fleming, A. M.; Nguyen, N. L. B.; Burrows, C. J. *ACS Cent. Sci.* **2019**, *5*, 218−228.
8. Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867–5892.
9. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.
10. Tateishi-Karimata, H.; Sugimoto, N. *Nucleic Acids Res.* **2021**, *49*, 7839–7855.
11. Gehring, K.; Leroy, J.-L.; Guéron, M. *Nature* **1993**, *363*, 561–565.
12. Guéron, M.; Leroy, J.-L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326–331.
13. Day, H. A.; Pavlou, P.; Waller, Z. A. E. *Bioorg. Med. Chem*. **2014**, *22*, 4407–4418.
14. Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res*. **2018**, *46*, 8038–8056.
15. Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863–2874.
16. Rajendran, A.; Nakano, S.-I.; Sugimoto, N. *Chem. Commun.* **2010**, *46*, 1299–1301.
17. Day, H. A.; Huguin, C.; Waller, Z. A. E. *Chem. Commun.* **2013**, *49*, 7696–7698.
18. Bhavsar-Jog, Y. P.; Dornshuld, E. V.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. *Biochemistry* **2014**, *53*, 1586–1594.
19. Assi, H. A.; Harkness, R. W.; V; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2016**, *44*, 4998–5009.
20. Wright, E. P.; Day, H. A.; Ibrahim, A. M.; Kumar, J.; Boswell, L. J. E.; Huguin, C.; Stevenson, C. E. M.; Pors, K.; Waller, Z. A. E. *Sci. Rep.* **2016**, *6*, 39456–39462.

21. Debnath, M.; Ghosh, S.; Chauhan, A.; Paul, R.; Bhattacharyya, K.; Dash, J. *Chem. Sci.* **2017**, *8*, 7448–7456.
22. Sheng, Q.; Neaverson, J. C.; Mahmoud, T.; Stevenson, C. E. M.; Matthews, S. E.; Waller, Z. A. E. *Org. Biomol. Chem.* **2017**, *15*, 5669–5673.
23. Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T.-M.; Tan, J.-H.; Huang, Z.-S.; Li, D. *Chem. Commun.* **2018**, *54*, 2036–2039.
24. 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.
25. 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*, 1710–1723.
26. Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H. *J. Am. Chem. Soc.* **2016**, *138*, 14138–14151.
27. 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.
28. Cheng, M.; Qiu, D.; Tamon, L.; Ištvánková, E.; Víšková, P.; Amrane, S.; Guédin, A.; Chen, J.; Lacroix, L.; Ju, H.; Trantírek, L.; Sahakyan, A. B.; Zhou, J.; Mergny, J.-L. *Angew. Chem., Int. Ed.* **2021**, *60*, 10286−10294.
29. 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.
30. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186.
31. Phan, A. T.; Guéron, M.; Leroy, J.-L. *J. Mol. Biol.* **2000**, *299*, 123–144.
32. B. R. Vummidi, J. Alzeer and N. W. Luedtke, *ChemBioChem* **2013**, *14*, 540–558
33. E. Largy, A. Granzhan, F. Hamon, D. Verga and M.-P. Teulade-Fichou, *Top. Curr. Chem.* **2013**, *330*, 111–177.
34. A. D. Rache and J.-L. Mergny, *Biochimie* **2015**, *115*, 194–202.
35. I. J. Lee, S. P. Patil, K. Fhayli, S. Alsaiari and N. M. Khashab, *Chem. Commun*. **2015**, *51*, 3747–3749.
36. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.
37. Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. *J. Org. Chem*. **2012**, *77*, 1026–1044.
38. Sakamoto, T.; Hasegawa, D.; Fujimoto, K. *Chem. Commun.* **2015**, *51*, 8749–8752.
39. Wee, W. A.; Yum, J. H.; Hirashima, S.; Sugiyama, H.; Park, S. *RSC Chem. Biol.* **2021**, *2*, 876–882.
40. Saady, A.; Wojtyniak, M.; Varon, E.; Böttner, V.; Kinor, N.; Shav-Tal, Y.; Ducho, C.; Fischer, B. *Bioconjugate Chem.* **2020**, *31*, 1188-1198.
41. Lieblein, A. L.; Krämer, M.; Dreuw, A.; Fürtig, B.; Schwalbe, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 4067–4070.
42. Hänsel, R.; Löhr, F.; Trantirek, L.; Dötsch, V. *J. Am. Chem. Soc*. **2013**, *135*, 2816–2824.
43. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
44 Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H. *ChemBioChem* **2015**, *16*, 1647–1656.
45 Mata, G.; Luedtke, N. W. *J. Am. Chem. Soc.* **2015**, *137*, 699–707.
46 Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res.* **2017**, *45*, 2951–2959.
47 Školáková, P.; Renčiuk, D.; Palacký, J.; Krafčík, D.; Dvořáková, Z.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. *Nucleic Acids Res.* **2019**, *47*, 2177–2189.
48. Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. *Angew. Chem., Int. Ed*. **2012**, *51*, 250–253.
49. Sapper, H.; Lohmann, W. *Biophys. Struct. Mechanism* **1978**, *4*, 327–335.

50. Itahara, T. *J. Chem. Soc. Perkin Trans. 2*, **1998**, 1455-1462.
51. Baranowski, M. R.; Warminski, M.; Jemielity, J.; Kowalska, J. *Nucleic Acids Res.* **2020**, *48*, 8209–8224.

Chapter 4:

3FBF-dU probe unravels the complex structural dynamics of Braf i-motif DNAs



Chapter 4 is a reprint of part of the data published in the journal: **Khatik, S. Y.**; Srivatsan, S. G. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.
The thesis author is the main author and researcher for this work.

4.1 Introduction

The iM structures are observed to be more stable and complex in the promoter regions due to the varying length and number of C-stretches.[1–5] Studies using model C-rich ONs suggest that five or more cytosine-containing stretches with a shorter loop tend to form a more stable iM structure compared to shorter C-stretches.[2–5] Further, formation of the secondary structure in the loop region provide additional stability to iM structure. Apart from that external factors namely molecular crowding, negative superhelicity and certain modifications are known to support the formation of iM structure at neutral or near physiological pH.[6–12] Thus, as several promoter region of the oncogene contain long C-rich sequences of varying length of C-stretches, they could formed iM structure at neutral pH or near physiological pH. However, long C-rich sequences could fold into multiple iM conformations using the different combination of C-stretches and loop residues and smaller energy difference between these structures can lead to a complex population equilibrium.[2,5] The CD, UV-thermal melting, NMR, X-ray crystallography and fluorescence techniques have been used to investigate the formation of iM structure and their stability under different conditions.[6,7,13,14] However, when multiple species are present, seldom these techniques provide useful information as they fail to resolve individual structures. In this context, ultrafast time-resolved IR and fluorescence spectroscopy techniques have been used to study the folding behavior of C-rich tracts.[15,16] Notably, a combination of FRET and FCS analysis as a function of pH in bulk and at the single-molecule level supported the coexistence of partially folded structures (iMs) along with the single-stranded structure at neutral pH, which was not evident in CD spectra.[17] A new method was developed based on molecular population dynamics in which the relative population of different species formed by C-rich bcl-2 promoter sequence in the absence and presence of ligands was studied by using mechanical unfolding in laser tweezers.[18] More recently, the fast dynamic nature of iMs was confirmed by a combined data analysis of fluorescence lifetimes, global decay fitting and multivariate analysis.[19] While these techniques use sophisticated experimental setup, a simple and an efficient tool that can provide comprehensive understanding of different iM conformations under a variety of conditions is practically not available. Therefore, development of a robust chemical probe that can resolve multiple conformations of C-rich sequences under different conditions (e.g., pH, temperature and cellular systems) is highly desired. To achieve this, the probe should (i) display distinct spectral signatures for different structures, (ii) resolve population equilibrium and (iii) enable identify physiologically active iM conformations.

In this chapter, we have utilized a TFBF-dU analog to investigate the complex equilibrium of iM structures formed by the long C-rich wild-type and mutated sequences from the promoter region of the B-*raf* gene. As discussed earlier, this is a next generation probe of 5-fluorobenzofuran modified deoxyuridine and developed to enhance the sensitivity of fluorine label. The presence of three equivalent ^{19}F atoms in TFBF modified deoxyuridine analog (**1**) enhanced its ^{19}F NMR detection limit, which could be useful to detect lowly populated structures (Figure 1A). Previously, we have employed TFBF-dU analog to determine the GQ and iM structural equilibrium under different conditions.[20,21] Interestingly, modified analog was able to detect lowly populated 3´E iM structure (8%) at different temperatures. Inspired from these results, we employed the modified analog to explore the complex equilibrium of B-*raf* iM structures under different conditions. Upon incorporation into the B-*raf* C-rich ON, it fluorescently reports the pH dependent formation of iM structures and indicates that the iM structure could even exist at near physiological pH. Notably, fluorine label of the nucleoside displays that this long C-rich sequence exhibits a complex iM structural polymorphism and their equilibrium vary with the surrounding conditions (pH and temperature) and in mammalian cell lysate (Figure 1B). Further, ^{19}F NMR property of the nucleoside analog in mutated sequences were utilized to identify a major iM conformation from the complex iM equilibrium.

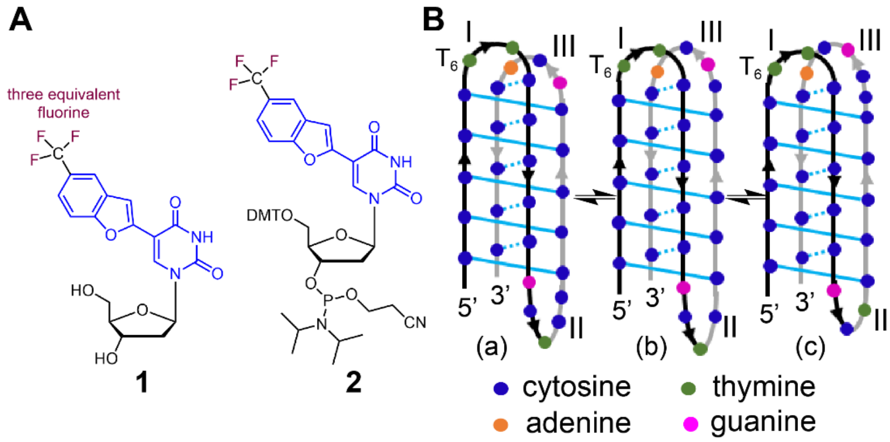


Figure 1. (**A**) Structure of the trifluoromethylbenzofuran (TFBF) modified nucleoside (**1**) and its phosphoramidite **2** and (**B**) Representative iM conformations of native *B-raf* ON **7** with ten C-C$^+$ base pairs. Conformations **a**, **b** and **c** have a loop size of 2:5:3, 2:4:4 and 2:3:5, respectively.

4.2 Results

4.2.1 Incorporation of the TFBF-dU into B-*raf* iM-forming sequences

To determine the efficiency of modified nucleoside (**1**) to investigate the complex iM structural dynamics, we chosen a B-*raf* C-rich sequence for the following reason. The *B-raf* gene encodes for the cytoplasmic serine/threonine kinase, which is important in cell proliferation and apoptosis.[22] Activating mutations in this gene are noticed in melanomas, colorectal cancers and thyroid carcinomas, and hence, considered as an important target for cancer therapy.[23] There are two main G- and C-rich segments in the promoter region of this gene. One called *B-raf*, is located downstream of the transcription start site and other called *B-raf-176* found upstream of the start site at -176 position.[24] *B-raf-176* (ON **7**) has four C-stretches capable of forming a complex population of iM structures. The first two and the fourth stretch contains five cytosines each and the third stretch is composed of seven cytosines. Hence, this sequence may adopt multiple iM conformations by using different combinations of cytosines from the long C-stretches. Using this sequence, we systematically studied the formation and stability of different iM structures as a function of pH and temperature in intracellular ionic conditions.

TFBF-modified *B-raf-176* ON **3** were synthesized by incorporating the phosphoramidite substrate **2** using solid-phase ON synthesis (Table 1). Waller and co-worker reported that C-rich sequences with five or more cytosine-containing C-stretches and a shorter loop could fold into a more stable iM structure.[2] In case of the *B-raf* sequence one could envision such iM structures, wherein the first loop (T_6T_7) should be shorter and more rigid as compared to others loop residues (Figure 1B). The third C-stretch contains seven cytosines, and hence, depending on which cytosines are involved in the folding process, several iM structures with varying loop lengths can form. T_6 position of the first loop is likely to be common for all possible ten intercalated C-C$^+$ base pair structures, but its interaction with nucleobases of loop III will differ with conformation (Figure 1B). Hence, we incorporated the probe at this position to prepare ON **3**. Mutated *B-raf* ONs **4**–**6** were also synthesized to study the role of different cytosines in the iM formation (Table 1). All the ONs were purified using PAGE analysis under denaturing conditions and their purity and integrity were confirmed by RP-HPLC and mass analysis (Figure 2 and 3, Table 2).

Table 1. TFBF-modified and respective control unmodified ON sequences.

ON[a]	5'------------sequence--------- 3'
3	CCC CC**T*** TCC CCC GCT CCC CCC CGC ACC CCC
4	**T**CC CC**T*** TCC CCC GCT CCC CCC CGC ACC CCC
5	CCC CC**T*** TCC CCC GCT **T**CC CCC **T**GC ACC CCC
6	CCC CC**T*** TCC CCC GCT **TT**C CCC CGC ACC CCC
7	CCC CCT TCC CCC GCT CCC CCC CGC ACC CCC
8	**T**CC CCT TCC CCC GCT CCC CCC CGC ACC CCC
9	CCC CCT TCC CCC GCT **T**CC CCC **T**GC ACC CCC
10	CCC CCT TCC CCC GCT **TT**C CCC CGC ACC CCC

ON **4** is a native *B-raf-176* C-rich sequence with the TFBF-dU modification in the first loop (T_6 position). ONs **5**–**7** are mutated *B-raf* ONs. **T*** represents site of modification and **T** represents site of mutation (C to T). ONs **8**–**11** are respective control unmodified ONs.

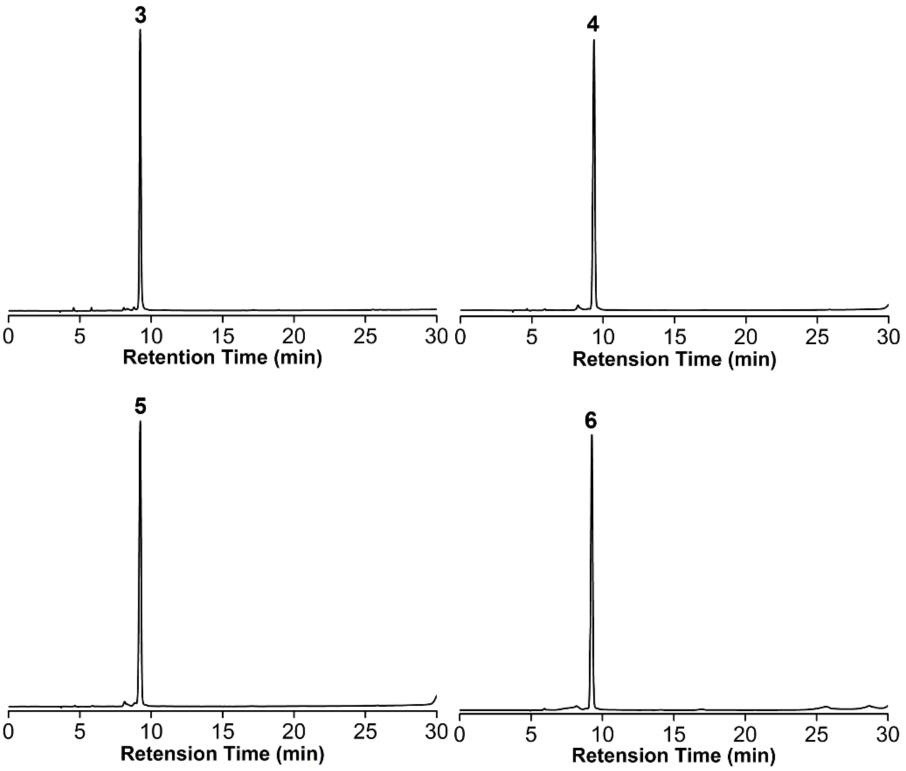
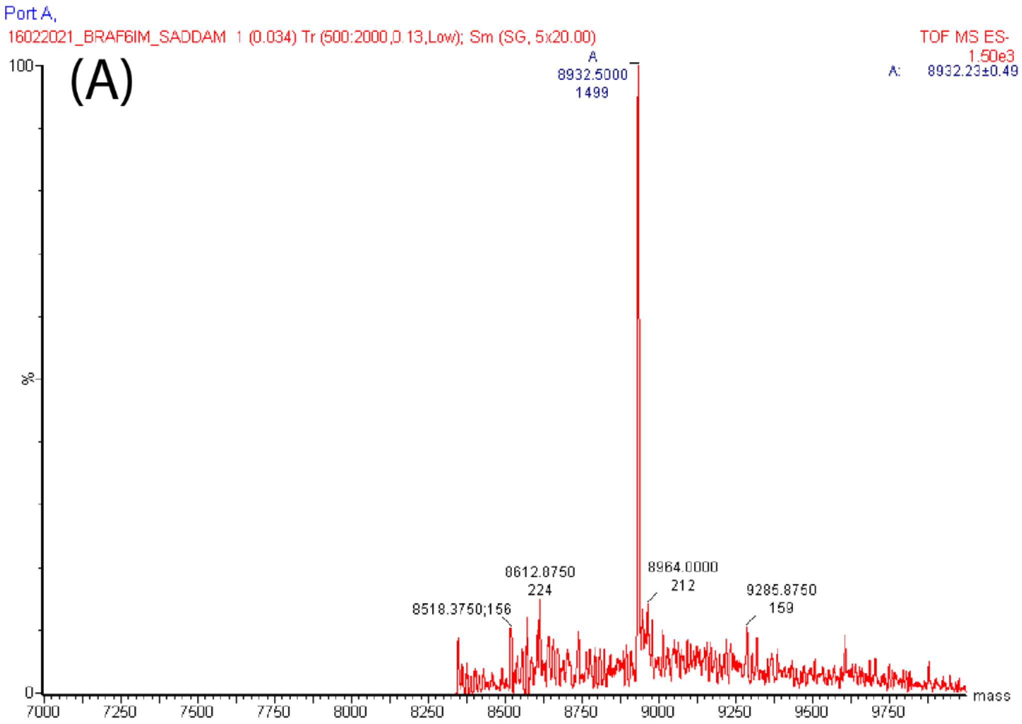
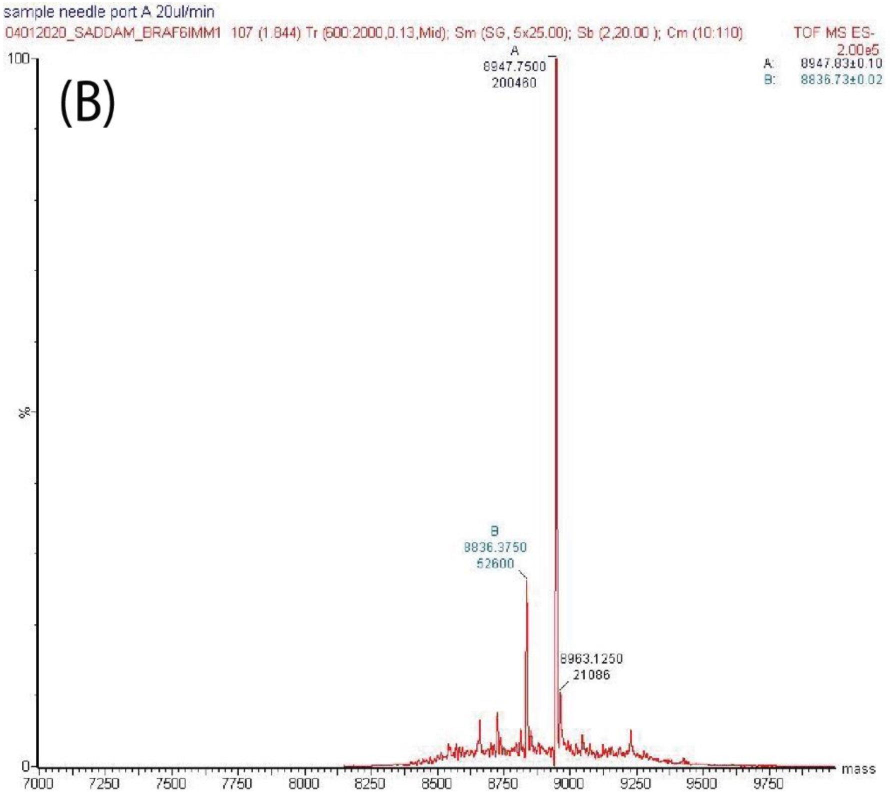
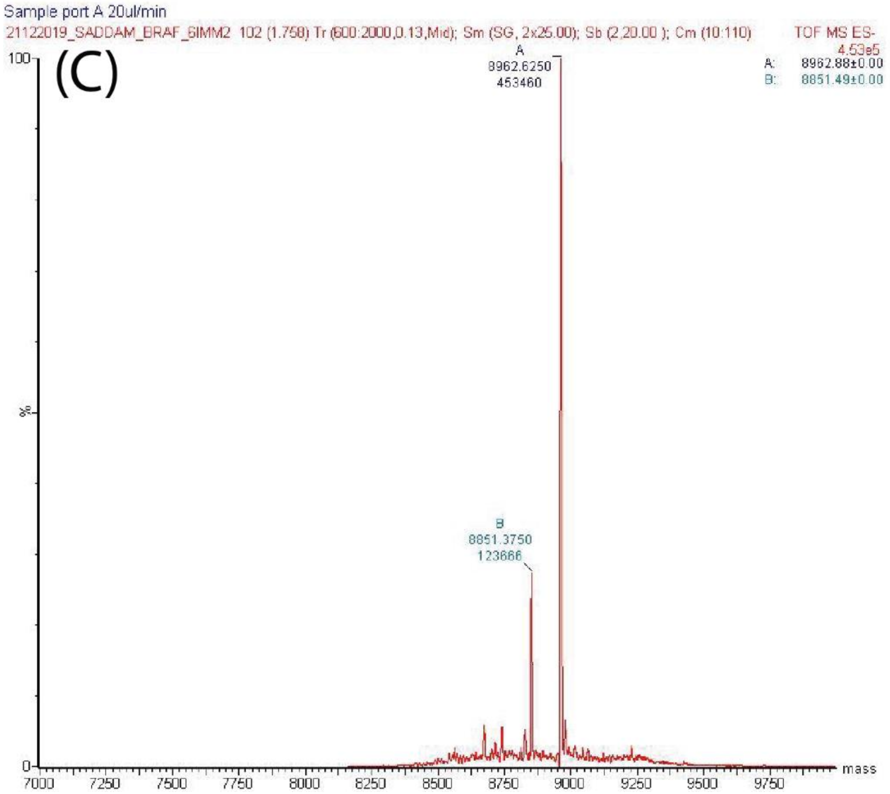


Figure 2. Reversed phase-HPLC chromatograms of PAGE purified ONs **3**–**6** analyzed at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).




sample needle port A 20ul/min
04012020_SADDAM_BRAF6IMM1 107 (1.844) Tr (600:2000,0.13,Mid); Sm (SG, 5x25.00); Sb (2,20.00); Cm (10:110)
TOF MS ES-
2.00e5
A: 8947.83±0.10
B: 8836.73±0.02
(B)
A
8947.7500
200460
B
8836.3750
52600
8963.1250
21086
100
%
0
mass
7000
7250
7500
7750
8000
8250
8500
8750
9000
9250
9500
9750


Sample port A 20ul/min
21122019_SADDAM_BRAF_6IMM2 102 (1.758) Tr (600:2000,0.13,Mid); Sm (SG, 2x25.00); Sb (2,20.00); Cm (10:110)
TOF MS ES-
4.53e5
A: 8962.88±0.00
B: 8851.49±0.00
(C)
A
8962.6250
453460
B
8851.3750
123666
100
%
0
mass
7000
7250
7500
7750
8000
8250
8500
8750
9000
9250
9500
9750

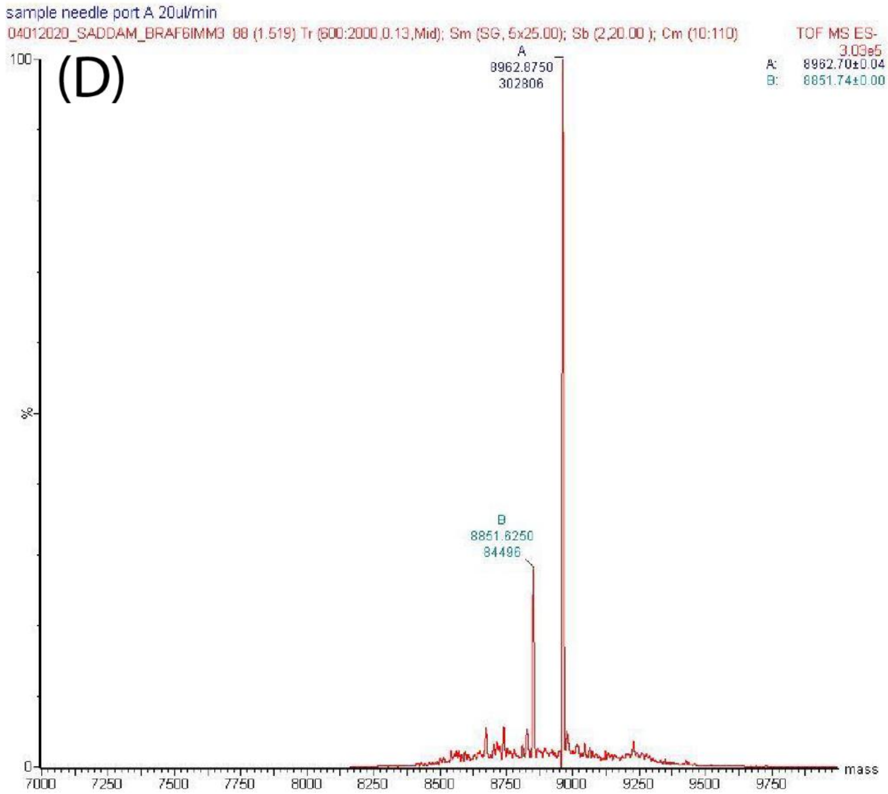


Figure 3. ESI-MS spectra of modified ONs (**A**) **3**, (**B**) **4**, (**C**) **5**, (**D**) **6**.

Table 2. Molar absorptivity and mass of modified DNA ONs.

DNA ON	$\mathcal{E}_{260}$ ^{a} [M^{-1} cm^{-1}]	Calculated mass	Observed mass
3	232 x 10^3	8932.7	8932.5
4	233 x 10^3	8947.7	8947.8
5	233 x 10^3	8962.7	8962.6
6	233 x 10^3	8962.7	8962.9

[a] Molar absorptivity ($\mathcal{E}_{260}$) of modified ONs was determined by using Oligo Analyzer 3.1. The molar absorptivity of modified nucleoside **1** ($\mathcal{E}_{260}$ = 11.4 x 10^3 M^{-1} cm^{-1}) was used in the place of thymidine.

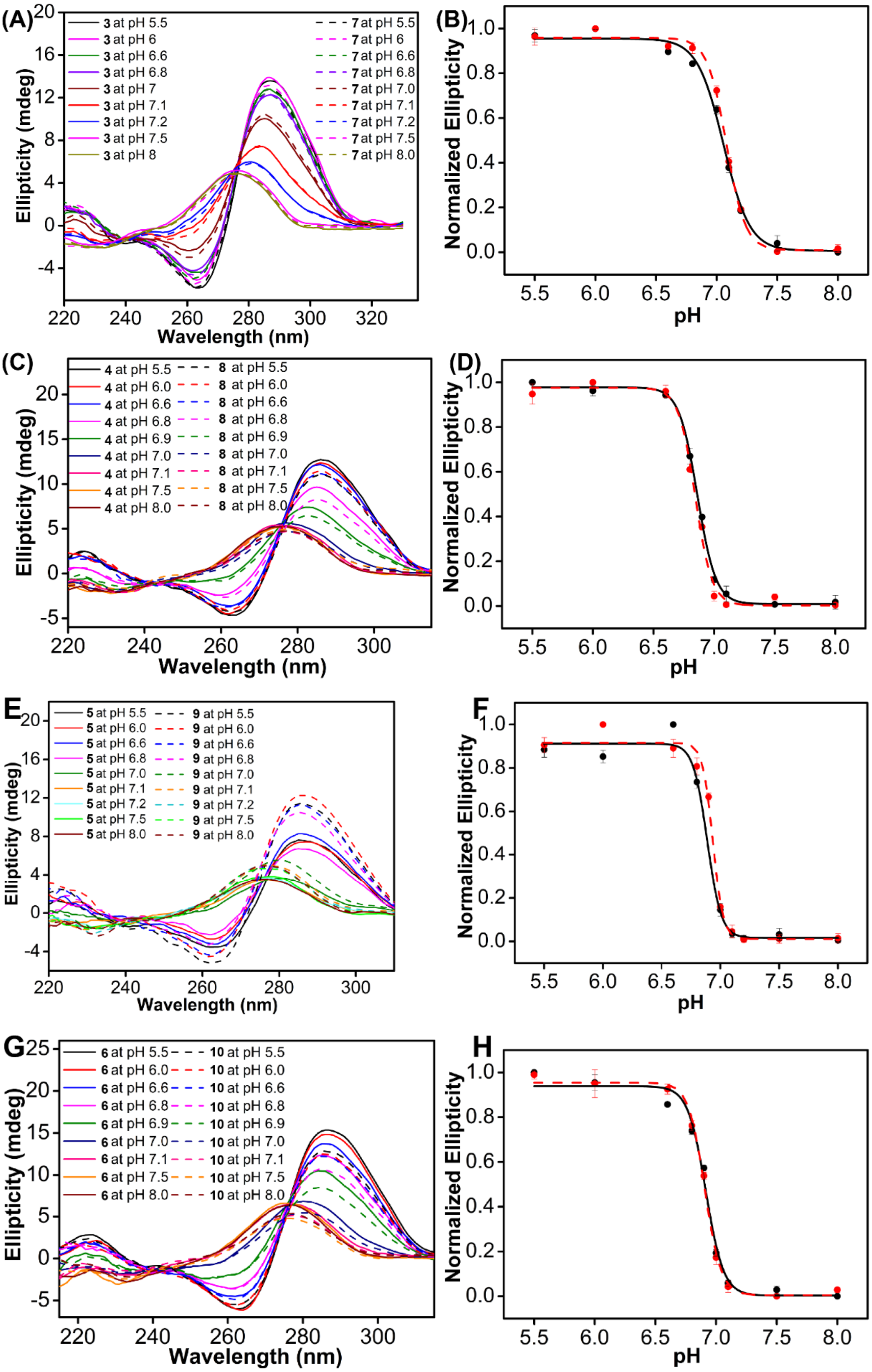


Figure 4. (**A**, **C, E** and **G**) CD spectra (5 µM) of modified DNA ONs **3**–**6** and their control DNA ONs **7**–**10**, respectively, at different pH (pH 5.5–8.0). (**B**, **D, F** and **H**) Curve fitted to the plot of normalized Ellipticity at 286 nm *versus* pH. Solid line is for modified ON and dashed line is for control ON.

Table 3. Transition pH (pH_T) values for modified and control unmodified ONs by CD and fluorescence.

ON	CD	fluorescence
modified **3**	7.1 ± 0.02	7.2 ± 0.02
control **7**	7.1 ± 0.01	-
modified **4**	6.9 ± 0.01	6.9 ± 0.02
control **8**	6.8 ± 0.01	-
modified **5**	6.9 ± 0.02	7.0 ± 0.01
control **9**	6.9 ± 0.01	-
modified **6**	6.9 ± 0.02	6.9 ± 0.02
control **10**	6.9 ± 0.01	-

Samples of the modified wild type ON **3** and mutated ONs **4**−**6** and their corresponding control ONs **7**−**10** were annealed in intracellular ionic conditions at different pH (MES 5.5–6.6 or HEPES 6.8–7.5).[25] CD measurements of these samples display the characteristic CD signature for iM structure i.e. positive peak at 286 nm and a negative peak at 263 nm indicating the formation of iM structure at lower and near neutral pH (Figure 4A, 4C, 4E and 4G). The transition pH (pH_T) values for all modified and control ONs were obtained by plotting normalized ellipticity at 286 nm *versus* pH (Figure 4B, 4D, 4F and 4H, Table 3). pH_T values obtained for a wild-type modified and native ON was in agreement with literature report.[24] Notably, all modified ONs display nearly similar pH_T values compare to its corresponding control ONs. This suggest that the C to T mutation in *B-raf* ON has only minor impact on the formation of iM structure (slightly less pH_T values). Further, UV-thermal melting analysis of all modified and their corresponding control ONs display nearly same melting temperatures at different pH (Figure 5, Table 4). Collectively, this suggested that the modification has negligible effect on the native *B-raf* iM structures. In UV-thermal melting study, hysteresis was observed in melting and annealing cycle, which was more pronounced near the pH_T values of respective ONs. These results indicates that *B-raf* ONs could adopt more than one iM structures in given conditions.[2,3,5]

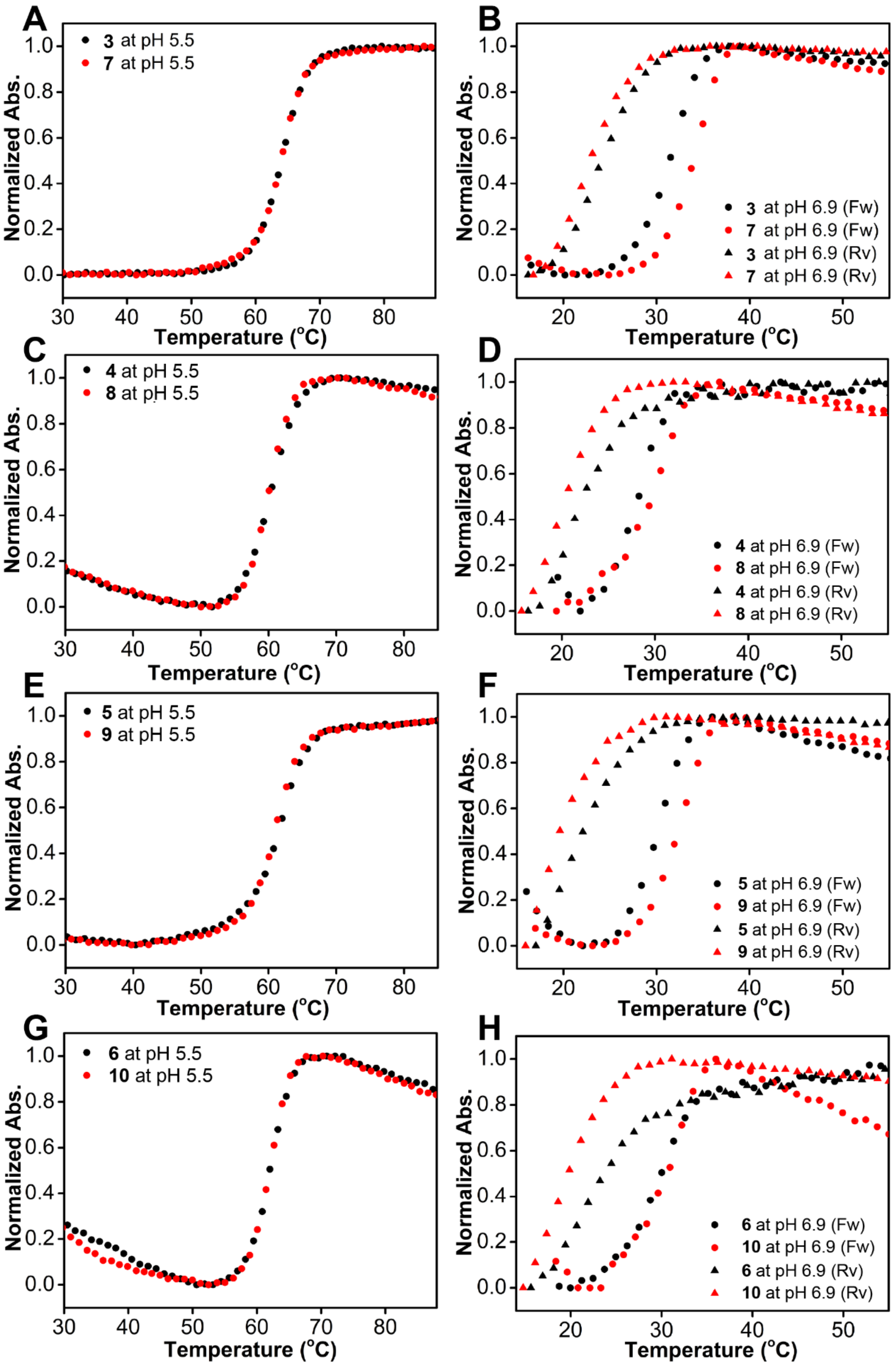


Figure 5. UV-thermal melting profile of TFBF-modified *B-raf* ONs and corresponding control unmodified ONs (1 µM) at pH 5.5 and 6.9. Changes in absorbance were measured at 260 nm.

Table 4. T_m values of TFBF-modified *B-raf* ONs and control unmodified ONs at different pH.

ON	pH 5.5		pH 6.0		pH 6.6		pH 6.9	
	T_m (°C) (Fw)	T_m (°C) (Rv)	T_m (°C) (Fw)	T_m (°C) (Rv)	T_m (°C) (Fw)	T_m (°C) (Rv)	T_m (°C) (Fw)	T_m (°C) (Rv)
3	64.6 ± 0.1	63.1 ± 0.6	53.5 ± 0.2	51.4 ± 0.5	38.2 ± 1.2	35.1 ± 0.5	32.9 ± 0.1	22.5 ± 0.1
7	65.3 ± 0.1	63.2 ± 0.2	53.3 ± 0.1	51.2 ± 0.1	41.6 ± 0.1	32.7 ± 0.2	35.0 ± 0.1	22.3 ± 0.6
4	60.6 ± 0.1	59.4 ± 0.5	47.6 ± 0.1	47.2 ± 0.8	36.4 ± 0.1	30.9 ± 0.2	29.3 ± 0.7	20.1 ± 0.1
8	60.2 ± 0.1	59.3 ± 0.7	51.0 ± 0.2	47.2 ± 0.6	38.5 ± 0.1	28.9 ± 0.2	30.5 ± 1.1	19.3 ± 0.1
5	63.0 ± 0.7	61.7 ± 0.6	52.4 ± 0.8	50.7 ± 0.7	39.4 ± 0.4	31.8 ± 0.5	31.1 ± 0.3	19.5± 0.1
9	62.0 ± 0.1	60.0 ± 0.1	52.0 ± 0.7	47.8 ± 0.2	39.5 ± 0.4	28.3 ± 0.3	33.3 ± 0.2	18.4± 0.2
6	62.8 ± 0.7	62.3 ± 0.1	51.4 ± 0.1	50.2 ± 0.1	39.0 ± 0.7	32.2 ± 0.8	31.2 ± 0.1	20.5 ± 0.1
10	62.8 ± 0.1	61.4 ± 0.1	52.6 ± 0.7	49.0 ± 0.7	40.0 ± 0.1	30.6 ± 0.1	32.1 ± 1.2	19.0 ± 0.7

4.2.2 Nucleoside probe fluorescently reports iM structure of *B-raf*

Samples of modified *B-raf* ON **3** were prepared in intracellular ionic conditions at different pH and were excited at 330 nm. At higher pH (8), random coil form of native ON **3** were found to be highly fluorescent. Upon decreasing the pH, the native ON **3** exhibited a progressive decrease in fluorescence intensity with practically no change in the emission maximum (421 nm) as a result of formation of iM structures (Figure 6A). Since modified nucleoside could experience stacking interactions with adjacent bases upon iM formation and display quenching in the fluorescence intensity. A blue-shifted emission maximum as compared to the free nucleoside analog in water (λ_{em} in water 427 nm, chapter 2 and Table 1), reveals that the incorporated probe is in a slightly less polar environment. Notably, the pH_T value determined by fluorescence (~7.2) and CD (~7.1) was very close (Table 3), which further alludes that this sequence can form iMs at near physiological pH under cellular ionic conditions. Similarly, the mutated ONs **4**–**6** exhibited a very high fluorescence for RC structure at basic pH (8.0) and subsequently quenching in the fluorescence was observed when iM structure formed at lower pH (Figure 6C, 6E and 6G). pH_T values were found to be slightly lower (~6.9) than the native ON **4** (Figure 6D, 6F and 6H, Table 3). As mentioned earlier, the fluorescence profile of the modified nucleoside (**1**) remained unchange at different pH (Chapter 3 and Figure 8), suggesting that the changes in the fluorescence profile observed at different pH for modified ONs were due to the conformational transitions from RC to iM structure.

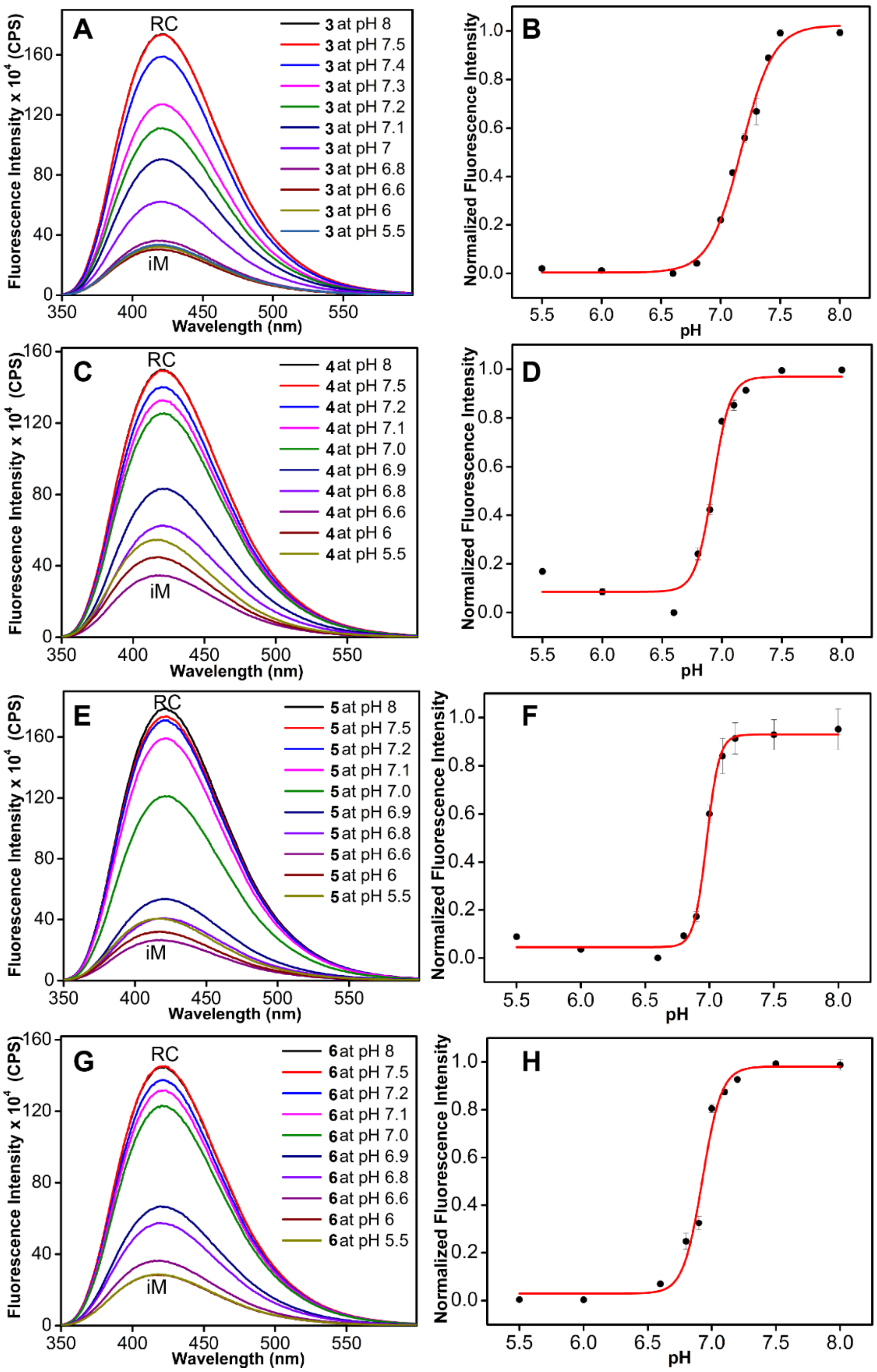


Figure 6. (**A**, **C**, **E** and **G**) Fluorescence spectra of modified native and mutated *B-raf* ONs **3**–**6** (1 µM) as a function of pH. Samples were excited at 330 nm with excitation and emission slit widths of 3 nm and 3 nm, respectively. (**B**, **D**, **F** and **H**) Curve fit for the plot of normalized fluorescence intensity at 421 nm *versus* pH.

4.2.3 Structural polymorphism of *B-raf* sequence depends on pH and temperature

In order to understand the complex structural equilibrium, we recorded NMR spectra of ONs at different pH (below, near and above pH_T) and temperatures (18°C to 35°C). ^{1}H NMR spectra of *B-raf* ON **3** at lower and near transition pH reveal broad peaks in the range of 15–16 ppm, suggesting the formation of more than one iM structure in equilibrium (Figure 7A–7C, bottom panel). A similar spectrum was also observed for the control unmodified ON **7** (Figure 8A). However, with this data it is not possible to identify the major and minor conformers and how the relative population is influenced by changes in the condition. In this context, the fluorine component of our dual-app probe **1** is highly valuable in detecting various conformations adopted by *B-raf* ONs under different conditions, including in mammalian cell lysate. First, we determined the ^{19}F signature of RC and stable iM structures at pH 7.5 and 6.0 (at 18°C), respectively (Figure 7A, top panel). While the RC structure of ON **3** gave a peak at -61.49 ppm, multiple peaks corresponding to iM structures were observed in the range of -61 to -62 ppm. Though signals corresponding to minor iM structures were broad (-61.2 to -61.7 ppm), a major iM structure having a chemical shift of -61.92 ppm was sharp and well resolved. Concentration-dependent ^{19}F NMR at pH 6.0 displayed similar spectra confirming the formation of intramolecular iM structures (Figure 9). As the pH was increased from 6.0 to transition pH (7.1) the major iM peak progressively decreased, which was accompanied by an increase in minor iM structures and RC form.

Further insights were obtained by recording ^{19}F NMR of ON **3** at elevated temperatures (Figure 7B and 7C). Upon increasing the temperature to 25°C and 35°C, iM population at pH 6.0 was not significantly affected. Interestingly, at 25°C and pH 6.6, peaks corresponding to minor iM structures increased and the amount of major iM structure reduced as compared to at 18°C. At a higher pH (6.9 and 7.1), predominantly RC form was present along with small amounts of iM structures. However, at near physiological temperature (35°C), ON **3** did not form iMs beyond pH 6.6. From these results, it is clear that native *B-raf* ON **3** forms multiple iMs and their relative population is sensitive to pH and temperature. Importantly, the major iM structure, which is formed at a lower pH (e.g., 6.0) need not form at physiological conditions. This level of understanding could not be obtained from ^{1}H NMR and CD analysis (*vide supra*).

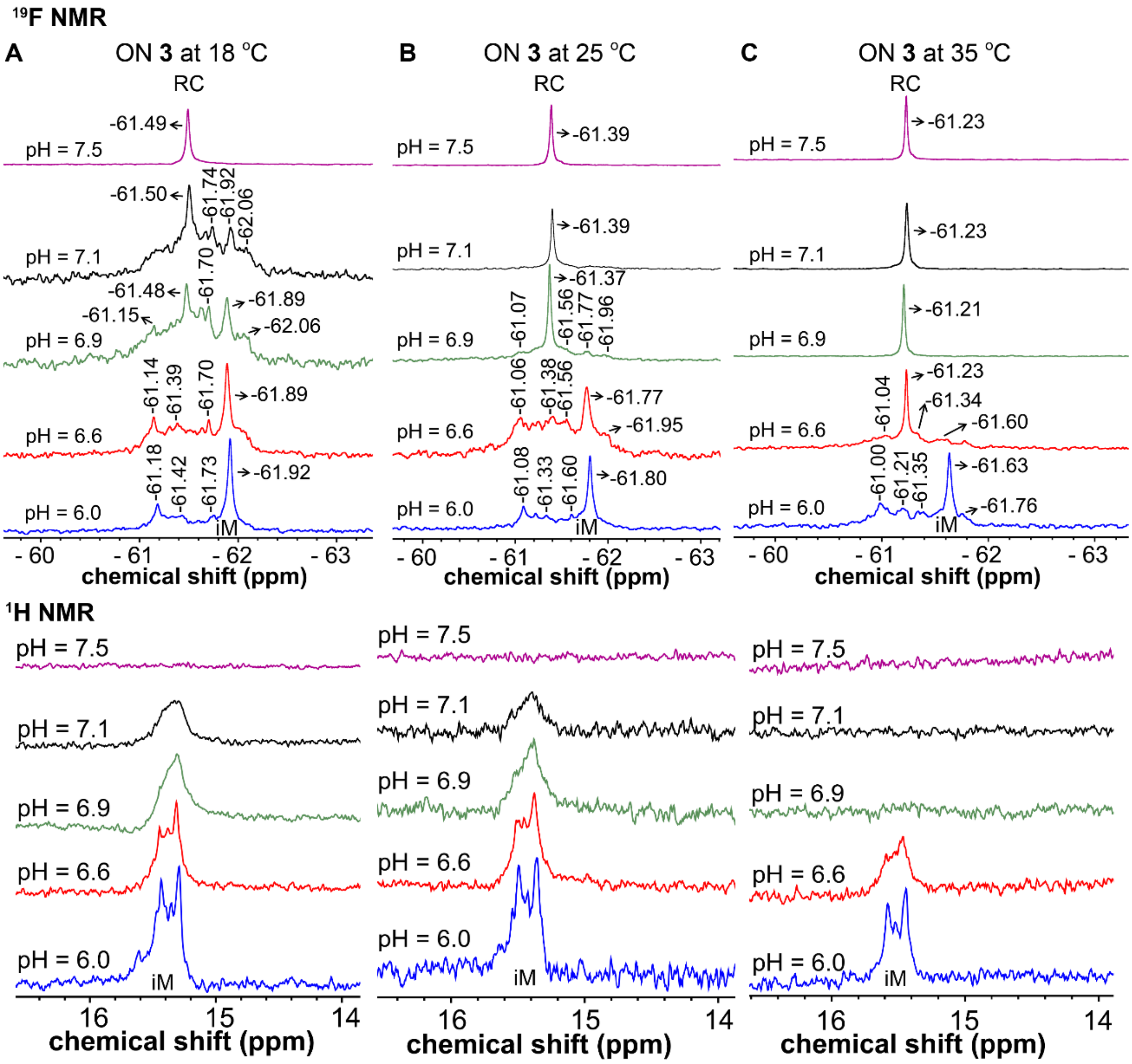


Figure 7. (**A** to **C**) Partial ^{19}F NMR (top panel) and ^{1}H NMR (bottom panel) spectra of native *B-raf* ON **3** (25 µM) at different pH and temperatures. The major iM structure is denoted as iM in the ^{19}F NMR spectra.

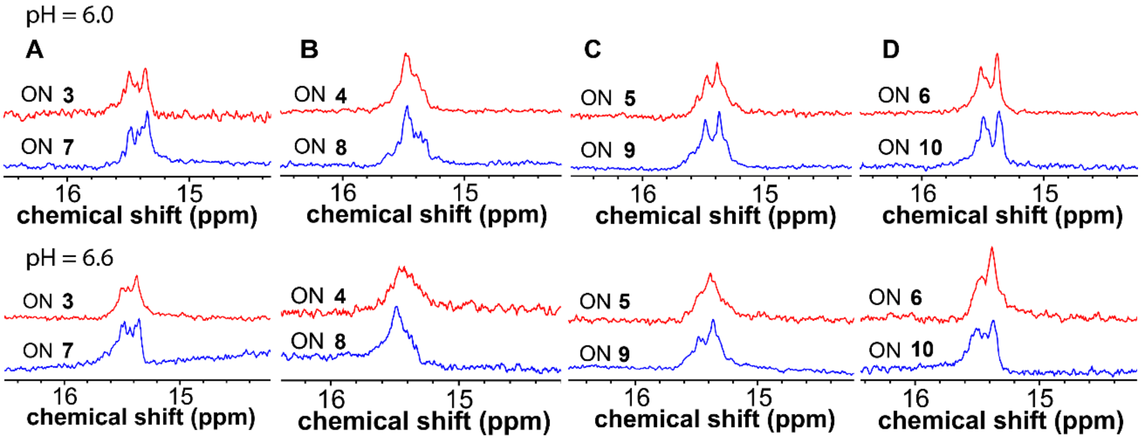


Figure 8. Partial ^{1}H NMR spectra of TFBF-modified (**3**–**6**) and control unmodified *B-raf* (**7**–**10**) ONs at 25 °C and at pH 6.0 and 6.6. Concentration of B-*raf* ONs is 25 µM.

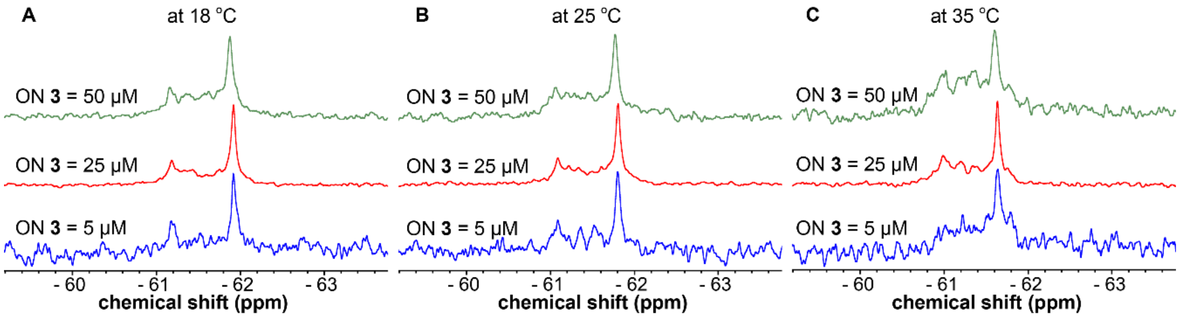


Figure 9. Concentration-dependent ^{19}F NMR spectra of ON **3** at pH 6.0 and at different temperatures.

4.2.4 Effect of mutations in C-stretches on iM structures

^{1}H NMR spectra of mutated *B-raf* ONs **4**–**6** and their control ONs **8**–**10** recorded at lower and near transition pH reveal broad peaks in the range of 15–16 ppm, suggesting the formation of more than one iM structure in equilibrium (Figure 8B–D). Next, ^{19}F NMR of mutated ONs **4**–**6** were recorded at different pH and temperatures to study the involvement of different cytosines in iM stem formation and effect of mutation on iM population. Native ON **3**, in principle could form several iM structures by using different combinations of cytosines. However, iM structures with 10 C-C^{+} pairs formed using each of the 5Cs present in four C-stretches are likely to be more stable (Figure 1B). Hence, ON **4**, containing C_1-T mutation, should not form stable iM structures. In support of this notion, the major peak (-61.92 ppm), which is exhibited by ON **3** was not present in the spectrum of ON **4** at pH 6.0 (Figure 10A). Along with minor iM structures the RC form was also observed, which was found to dominate with increase in pH and temperature (Figure 10). Further, ON **4** exhibited a lower T_m value as compared to the native ON sequence (ΔT_m = ~ 4 °C, Table 4). These results indicate that C_1 residue is very crucial for the formation of the major stable iM structure. Next, we studied the iM formation of ON **5** in which the 3rd C-stretch, containing seven cytosines, was mutated at two different positions (C_{16} and C_{22} to T). ON **5** formed multiple iMs at pH 6.0 as well as at near neutral pH with a similar T_m value as ON **3** (Figure 11, Table 4). However, this set of mutation resulted in the disappearance of the peak at -61.92 ppm, suggesting that these C residues are also important for the major iM structure formation (Figure 7A). Interestingly, when C_{16} and C_{17} positions were replaced with T residues in ON **6**, it produced an intense peak (-61.92 ppm) at pH 6.0 and 18°C, which was identical to the major iM conformation adopted by native ON **3** (Figure 7A). At elevated temperatures also this major iM structure was formed in good amounts (Figure 12). From this data, it appears that C_{16} and C_{17} are not part of the stem region of the major conformer. Further, this mutated sequence formed limited number of iM structures at near neutral pH and 18°C whose peaks were well resolved (Figure 12A). However, above pH 6.6 and at higher temperatures (25°C and 35°C), mainly RC structure was observed (Figure 12B and 12C).

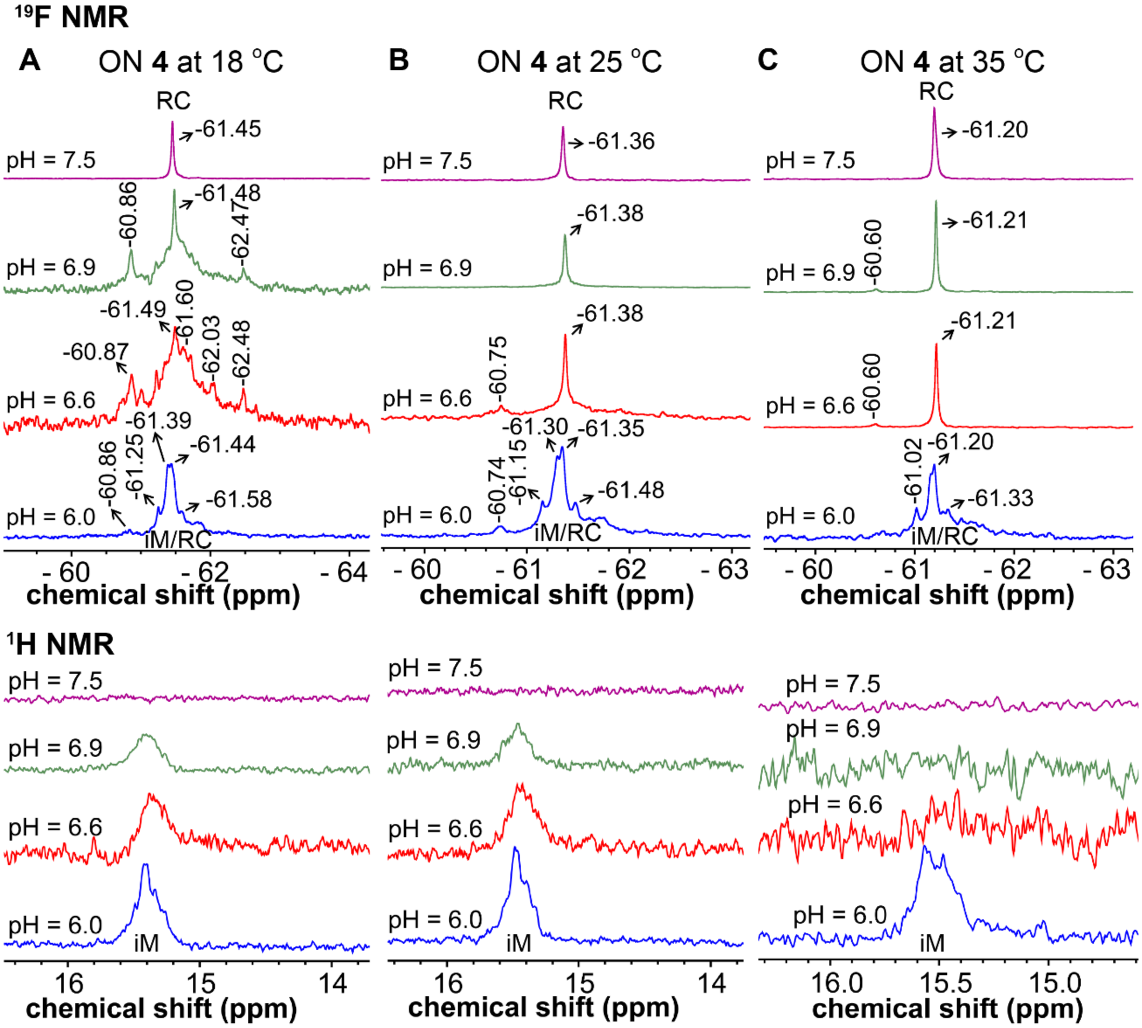


Figure 10. (**A** to **C**) Partial ^{19}F NMR (top panel) and ^{1}H NMR (bottom panel) spectra of mutated *B-raf* ON **4** (25 µM) at different pH and temperatures. Top panel: RC is random coil.

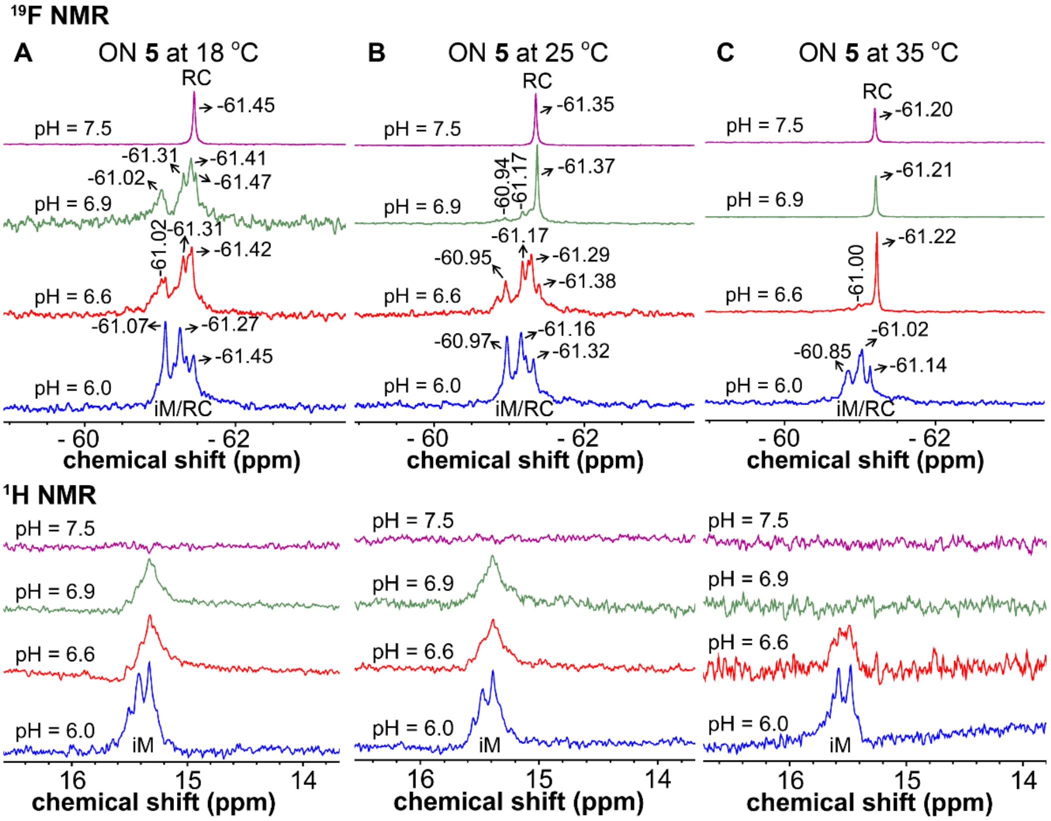


Figure 11. (**A** to **C**) Partial ^{19}F NMR (top panel) and ^{1}H NMR (bottom panel) spectra of mutated *B-raf* ON **5** (25 µM) at different pH and temperatures. Top panel: RC is random coil.

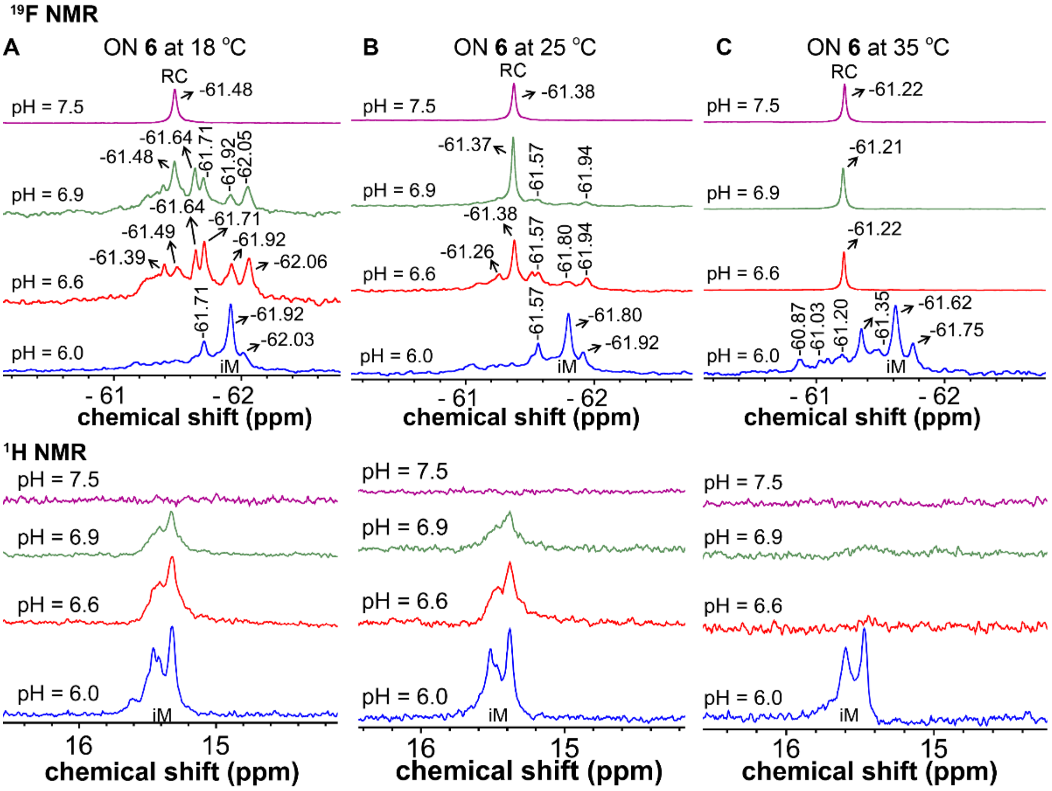


Figure 12. (**A** to **C**) Partial ^{19}F NMR (top panel) and ^{1}H NMR (bottom panel) spectra of mutated *B-raf* ON **6** (25 µM) at different pH and temperatures. Top panel: RC is random coil.

In the *B-raf* iM structure, C_1 can potentially base pair with C_{16}/C_{17}/C_{18} (Figure 13B). However, mutation studies indicate that C_{16} and C_{17} are part of the loop of the major iM form (see the comparison of the ^{19}F NMR spectra of native and mutated ONs at pH 6.0, Figure 13A). Hence, C_1 should pair with C_{18} and this would mean C_{22} pairs with C_5 to form the major structure with maximum C-C$^+$ base pairs. Taken together, it is proposed that the major iM structure formed by the native ON **3** is likely to have a conformation made of a 10 C-C$^+$ pair stem with a 2:5:3 loop size as shown in Figure 13B.

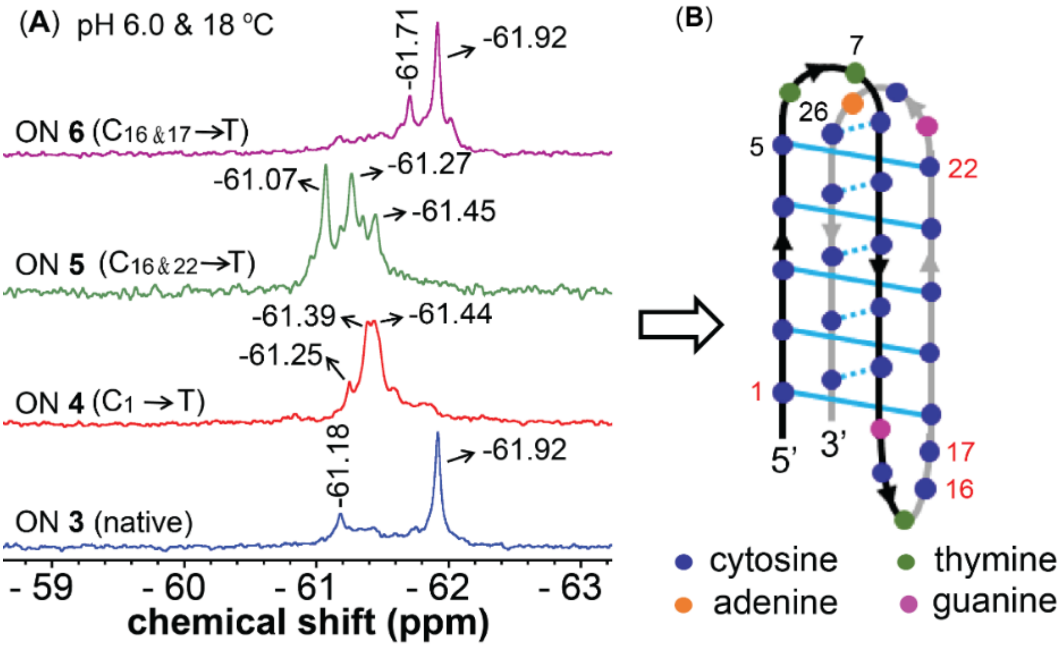


Figure 13. (**A**) Partial ^{19}F NMR spectra of native *B-raf* ON **3** (25 µM) and mutated ONs **4**–**6** (25 µM) at pH 6.0 and 18°C. (**B**) Proposed structure of the major iM conformation of ON **3** on the basis of mutation studies. Mutated points (as in ONs **4**–**6**) are colored in red.

4.2.5 Structure of *B-raf* ON in mammalian cell lysate

Cell lysates are commonly used as ex-vivo models mimicking cellular molecular crowding and ionic conditions to study the structure of nucleic acids.[26,27] ^{19}F NMR spectrum of native *B-raf* ON **3** recorded in HeLa cell lysate buffered at pH 6.9 (near pH_T) indicated the formation of multiple iM structures along with the major conformer at -61.91 ppm (Figure 14A). At pH 7.5 the ON in cell lysate remained as a RC structure. When the temperature was increased from 18°C to 25°C, the ON largely formed a RC structure at both the pH values (Figure 14B). In a control experiment without added ON, cell lysate was found to be free from ^{19}F signal. After NMR experiment, the modified ON was isolated by HPLC and its integrity was confirmed by mass measurement, which revealed that the ON was stable in the lysate during the course of NMR analysis (Figure 15). Good stability of the ON in cell lysate and a very similar ^{19}F NMR results in aqueous buffer and cell lysate suggests that ^{19}F signals from the probe can be used as signatures to detect ON conformations in cellular environment (compare with Figure 7A).

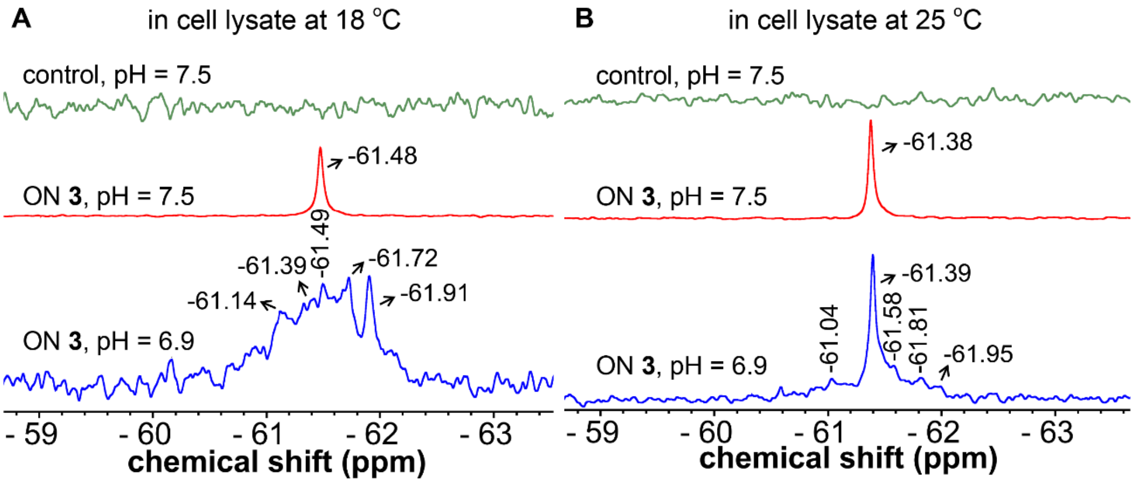


Figure 14. (**A** and **B**) Partial ^{19}F NMR spectra of *B-raf* ON **3** (25 µM) in HeLa cell lysate at different pH and temperatures (blue and red traces, respectively). Cell lysate alone with no added ON is shown in green (control).

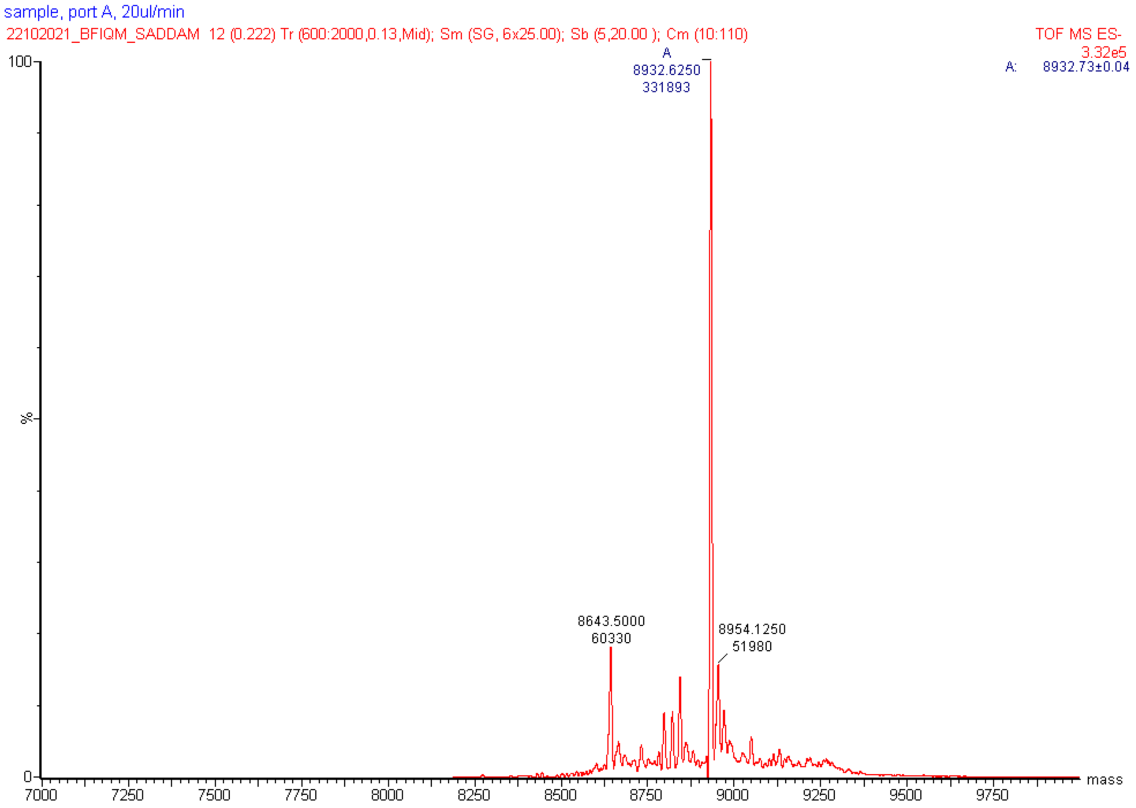


Figure 15. ESI-MS spectrum of modified ON **3** isolated from cell lysate after NMR analysis. Calculated mass: 8932.7 and found: 8932.6.

4.3 Discussion

We used the responsiveness of the nucleoside analog in probing the conformations of native and mutated C-rich *B-raf* ONs, which are capable of forming multiple iM structures. The nucleoside analog fluorescently reports the formation of iM structures with pH_T values close to neutral pH (Figure 6A and 6B). A closer look at the fluorescence profile indicates differences in the quenching pattern exhibited by ONs **3**–**6** (compare in Figure 6). ON **3**, which could form all possible iMs by using various combinations of cytosines, displays a progressive quenching in fluorescence intensity as the pH is lowered. Mutated ON **4** (C_1 to T) has four cytosines in the first C-stretch, and hence, would not be able to form stable iMs with maximum C-C^+ base pairs (5 pairs in each of the parallel stranded duplex). This ON shows a major reduction in intensity between pH 7.0 and 6.9 (Figure 6C). Double mutations created in the third C-stretch containing seven cytosines (ONs **5** and **6**) would form lesser numbers of iM structures containing five C-C^+ base pairs in each of the parallel strands. While ON **5** (C_{16} and C_{22} replaced with Ts) shows two significant dips in intensity, one from pH 7.1 to 7.0 and another from 7.0 to 6.9, ON **6** (C_{16} and C_{17} replaced with Ts) shows dips in intensity between pH 7.0–6.9 and 6.8–6.6 (Figure 6E and 6G). These results suggest that the native and mutated *B-raf* ONs could exist in different structural equilibrium depending on the pH. Again, ^{19}F label provides valuable insights on the conformation equilibrium. As evident from the ^{19}F NMR studies, native *B-raf* ON indeed forms multiple iM structures, wherein the major iM structure formed at a lower pH (6.0) diminishes as the pH is increased giving rise to several minor iM conformations (Figure 7A). In particular, at 18°C, a significant population of various iM structures along with the major conformer exists above neutral pH. However, at near physiological temperature and pH the iMs cease to exist. Experiments in cell lysate also suggest that *B-raf* C-rich motif may not form iM structures under physiological conditions. Nevertheless, our findings allude that ^{19}F signatures obtained *in vitro* using the probe can be potentially used to determine the possibility of C-rich sequences to adopt iM structures in cells.

^{19}F NMR analysis of mutated ONs helped in ascertaining the relevance of certain cytosine residues in the formation of iM structures, and more specifically, proposing the structure of the major iM conformation, which is not possible by CD and ^{1}H NMR experiments. Further, as can be seen from Figure 13, certain mutations in C-stretches need not abolish iM formation, rather can augment the formation of existing or alternate iM structures. While literature precedence suggests that mutations resulting in small changes in T_m would result in the formation of similar structures as the native ON,[28] our results clearly indicate that even small differences in T_m and pH values can have a significant ramification in the structural polymorphism of iMs, which otherwise is difficult to infer by using conventional probes or techniques.

4.4 Conclusions

The modified nucleoside analog **1** is highly useful in monitoring complex iM structural equilibrium under different conditions using fluorescence and ^{19}F NMR techniques. The ability of the probe to fluorescently report the transition from RC structure to iM form as a function of pH enables the determination of the pH_T value for the *B-raf* ONs. Importantly, the probe displays distinct ^{19}F signatures for different iM structures thereby providing an unprecedented means to dissect the iM population dynamics of native and mutated sequences. Further, we believe that this signature would be beneficial to identify physiologically relevant nucleic acid conformations under the cellular condition. Taken together, our new two-channel readout system can provide a simplified solution to study structural complexity and population equilibrium of not only iM-forming sequences but also other non-canonical nucleic acid motifs under various conditions and environments.

4.5 Experimental Section

- Materials and instruments are same as mentioned in section 3.5.1 and 3.5.2, respectively.
- Synthesis of modified DNA ONs and their characterization: Procedure as mentioned in sections 3.5.3, 3.5.4 were used for the synthesis of modified ONs, their purifications and characterization.
- Sample preparations and experimental procedure to perform the CD, UV-thermal melting, fluorescence and NMR analysis were followed as mentioned in section 3.5.5, 3.5.6, 3.5.7 and 3.5.8 respectively.

4.5.1 ^{19}F and ^{1}H NMR analysis of *B-raf* ONs at different pH

The ^{19}F NMR spectra were recorded in 35–40 min with 2000 of scans using the parameters as mentioned in section 3.5.8. Spectra were processed with an exponential window function using lb = 10 Hz. ^{1}H NMR spectra were recorded with water suppression using excitation sculpting with gradients. Number of scans were in the range of 400–600.

4.5.2 ^{19}F NMR analysis of *B-raf* ON 3 in cell lysate

HeLa cells were cultured in DMEM (Gibco) media supplemented with 10% (v/v) fetal bovine serum (Gibco) under 5% CO_2 condition at 37°C. Cells were seeded in a T-175 cm^2 corning flask. After three days, cells with 90% confluence were used for the preparation of cell lysate. 540 µL of 37% formaldehyde and 1 mL of 2.5 M glycine were added to DMEM medium, and cells were incubated at room temperature for 5 min. The media was removed and the cells were washed twice with cold DPBS (Gibco). Cells were collected using a cell scraper and transferred to a 15 mL sterilize centrifuge tube. After centrifugation at 1500 rpm for 5 min at 4°C, the cell pellet was transferred to a 1.5 mL protein low binding centrifuge tube and resuspended with 300 µL of nuclear lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 200 mM NaCl and 1% SDS). The sample was incubated on an ice bath for 5 min after which it was ultra-sonicated on ice and finally, the suspension was centrifuged at 16000 g for 10 min. The supernatant solution was collected and used for NMR experiment. Cell lysate (300 µL containing 25 µM of ON **3**) was subjected to NMR analysis at pH 7.5 and 6.9. pH of the lystate was adjusted from pH 7.5 to 6.9 by adding ~3 µL of 2 N HCl solution. After NMR analysis, sample was heated at 95°C and centrifuged at 16000 g for 10 min. Clear lysate was transferred to an Amicon Ultra 3K tube and filtered. The filtrate was analyzed by RP-HPLC and the peak corresponding to ON **3** was further subjected to mass analysis to confirm the identity of the ON.

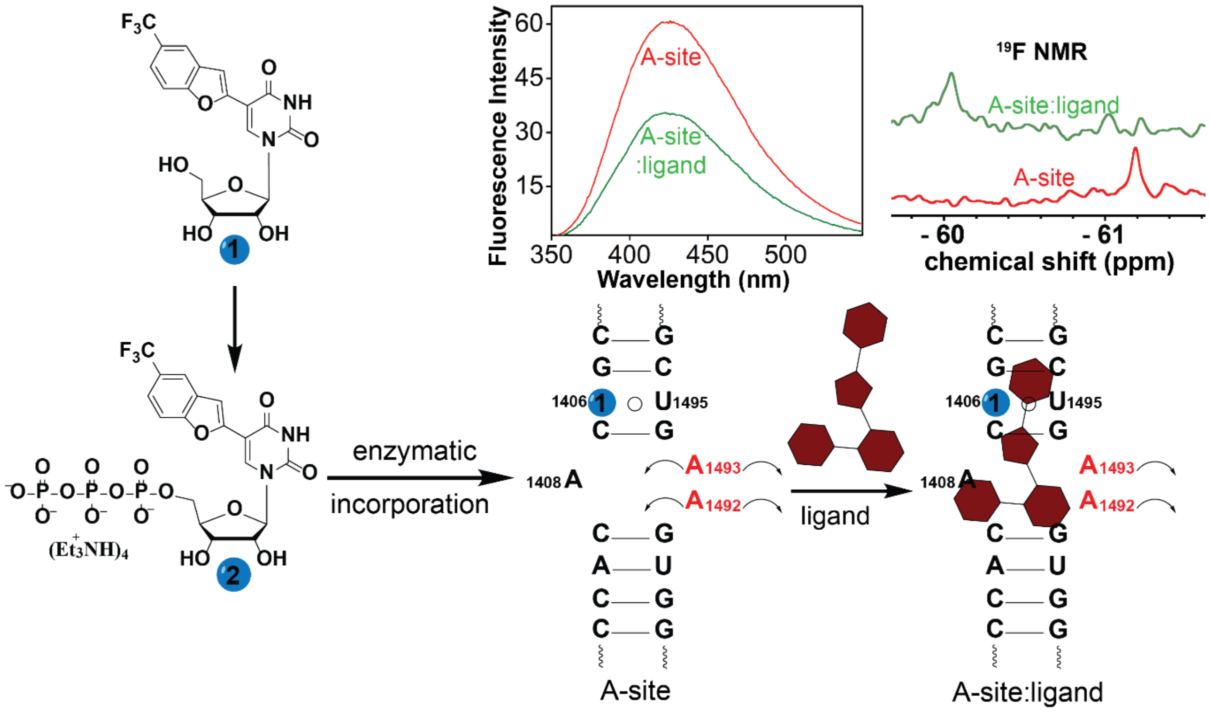
4.6 References

1. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Mol. BioSyst.* **2010**, *6*, 580–586.
2. Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. *Nucleic Acids Res.* **2017**, *45*, 2951–2959.
3. Školáková, P.; Renčiuk, D.; Palacký, J.; Krafčík, D.; Dvořáková, Z.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. *Nucleic Acids Res.* **2019**, *47*, 2177–2189.
4. 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.
5. Cheng, M.; Chen, J.; Ju, H.; Zhou, J.; Mergny, J.-L. *J. Am. Chem. Soc.* **2021**, *143*, 7792–7807.
6. Day, H. A.; Pavlou, P.; Waller, Z. A. E. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.
7. Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2018**, *46*, 8038–8056.
8. Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863–2874.
9. Rajendran, A.; Nakano, S.-I.; Sugimoto, N. *Chem. Commun.* **2010**, *46*, 1299–1301.
10. Day, H. A.; Huguin, C.; Waller, Z. A. E. *Chem. Commun.* **2013**, *49*, 7696–7698.
11. Bhavsar-Jog, Y. P.; Dornshuld, E. V.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. *Biochemistry* **2014**, *53*, 1586–1594.
12. Assi, H. A.; Harkness, R. W.; V; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2016**, *44*, 4998–5009.

13. Phan, A. T.; Guéron, M.; Leroy, J.-L. *J. Mol. Biol.* **2000**, *299*, 123–144.
14. Kang, C.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3874–3878.
15. Keane, P. M.; Wojdyla, M.; Doorley, G. W.; Kelly, J. M.; Parker, A. W.; Clark, I. P.; Greetham, G. M.; Towrie, M.; Magno, L. M.; Quinn, S. J. *Chem. Commun.* **2014**, *50*, 2990–2992.
16. Cohen, B.; Larson, M. H.; Kohler, B. *Chem. Phys.* **2008**, *350*, 165–174.
17. Choi, J.; Kim, S.; Tachikawa, T.; Fujitsuka, M.; Majima, T. *J. Am. Chem. Soc.* **2011**, *133*, 16146–16153.
18. Cui, Y.; Koirala, D.; Kang, H. J.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H. *Nucleic Acids Res.* **2014**, *42*, 5755–5764.
19. Benabou, S.; Ruckebusch, C.; Sliwa, M.; Aviñó, A.; Eritja, R.; Gargallo, R.; De Juan, A. *Nucleic Acids Res.* **2019**, *47*, 6590–6605.
20. Khatik, S. Y.; Sudhakar, S.; Mishra, S.; Kalia, J.; Pradeepkumar, P. I.; Srivatsan, S. G. *Chem. Sci*. **2023**, Just Accepted, doi.org/10.1039/D3SC00519D.
21. Khatik, S. Y.; Srivatsan, S. G. *Bioconjugate Chem.* **2022**, *33*, 1515–1526.
22. Prahallad, A.; Sun, C.; Huang, S.; Di Nicolantonio, F.; Salazar, R.; Zecchin, D.; Beijersbergen, R. L.; Bardelli, A.; Bernards, R. *Nature* **2012**, *483*, 100–104.
23. Dienstmann, R.; Tabernero, J. *Anti-cancer Agents in Medicinal chemistry* **2011**, *11*, 285–295.
24. Greco, M. L.; Folini, M.; Sissi, C. *FEBS Lett*. **2015**, *589*, 2117–2123.
25. Hänsel, R.; Löhr, F.; Trantirek, L.; Dötsch, V. *J. Am. Chem. Soc*. **2013**, *135*, 2816–2824.
26. Bao, H.-L.; Liu, H.-S.; Xu, Y. *Nucleic Acids Res.* **2019**, *47*, 4940–4947.
27. Wee, W. A.; Yum, J. H.; Hirashima, S.; Sugiyama, H.; Park, S. *RSC Chem. Biol.* **2021**, *2*, 876–882.
28. Dai, J.; Hatzakis, E.; Hurley, L. H.; Yang, D. *PLoS One* **2010**, *5*, e11647–e11654.

Chapter 5:

Detection of aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA using 3FBF-U



5.1 Introduction

RNA performs important functions under normal and pathological conditions like catalysis and transfers and regulation of genetic information.[1] It is known to fold into several structures from simple double helical to complex secondary and tertiary structures, and the function of RNA depends on the particular structure formed under cellular conditions.[2] Generally, RNA undergoes a conformational transition to perform different functions, which can arise from post-transcriptional modification, changes in environmental conditions, and also upon interactions with proteins, metabolites and ligands.[2] Hence, in order to understand the mechanism of functional RNA, it is necessary to determine their conformations. Several biophysical techniques (Fluorescence, NMR, EPR and X-ray crystallography) have been developed to understand RNA conformations, dynamics and their interactions with proteins and ligands.[3–6] Importantly, the fluorescence technique aid in the rapid investigation of RNA structures, dynamics and interactions with ligands.[3] On the other hand, NMR and X-ray crystallography techniques provide information on RNA at the atomic level, but they are not straightforward and are time-consuming.[4,6] Recently, notable number of fluorine labeled probes have been utilized to study nucleic acid structures.[7] The absence of fluorine atoms in biomolecules and 100% natural abundance makes the ^{19}F label probe a very sensitive and useful biophysical tool to study nucleic acid structures. Recently, Zhou group has developed 4´-F-uridine and utilized it to study the RNA structures (ssRNA, dsRNA) and enzyme-mediated RNA processing.[8] Kreutz group synthesized ^{19}F-^{13}C labeled RNA and employed it to monitor different RNA secondary structures.[9] Similarly, ^{19}F labeled probes have been used to study RNA dynamics and their interactions with proteins and ligands.[10] Owing to the significance of RNA in therapeutics, we decided to explore the usefulness of 5-trifluoromethyl-benzofuran-modified uridine (TFBF-U) in monitoring ligand-induced conformational changes in RNA structure. For this purpose, we chose one of the well know RNA targets namely, bacterial ribosomal decoding site RNA motif (A-site).

Binding of natural aminoglycoside antibiotics to A-site RNA makes it more prone to errors in protein synthesis of bacteria.[11] In the protein synthesis process, the decoding center of the ribosome (A-site) is responsible for the selection of an accurate tRNA.[12] A-site undergoes a specific conformational change upon interactions of a mRNA codon with an anticodon of the cognate tRNA, in this way it maintains high fidelity of protein synthesis.[13] Importantly, the small bulge subunit of A-site in the 16S rRNA is highly conserved, which contains two flexible adenine residues (A_{1492} and A_{1493}) and a Watson-Crick base pair (C_{1409}-

G_{1491}, Figure 1A).[12,13] Naturally occurring aminoglycoside-antibiotics interact with these residues upon binding with A-site and fix their conformation similar to those formed upon perfect codon-anticodon interactions. Thus, it misleads the selection of the correct cognate tRNA by the bacterial ribosomal decoding site. Over a period of time, bacteria naturally evolve to resist antibiotics by changing their genomic material due to the overuse of antibiotics. Hence, it is necessary to find new antibiotics which could replace available compromised drugs. In this direction, A-site constructs have been labeled with fluorescent nucleosides to determine the binding affinity of naturally occurring aminoglycosides-antibiotics.[14] The structural analysis indicated that the aminoglycosides interact with the A_{1408}, A_{1492} and A_{1493} residues of a bulge part of the A-site through direct H-bonding.[15] Additionally, they interact with the neighboring G-C pairs and a noncanonical $U_{1406}{\circ}U_{1495}$ pair through water-mediated H-bondings. Hence, A_{1492}, A_{1493} and U_{1406} residues have been replaced with fluorescent nucleoside analogs to investigate the aminoglycoside-induced conformational transitions of the A-site motif by recording changes in the fluorescence properties. For this purpose, 2-aminopurine (2AP) is a widely used fluorescent analog incorporated at A_{1492} or A_{1493} to construct fluorescent A-site.[14a,b,e] Apart from that, Hermann group has replaced the A_{1493} residue with 3-methylisoxanthopterin (3MI) and 6-methylisoxanthopterin (6MI).[14c] Tor group has also modified U_{1406} residue with a furan-modified uridine analog and 5-methoxyquinazoline-2,4(1H,3H)-dione as a FRET donor to obtain fluorescent A-site RNA.[14d] Here, to explore the utility of our dual-label nucleoside analog **1**, we environed to incorporate TFBF-UTP at the U_{1406} position of A-site RNA construct.

In the previous chapters, TFBF-dU analog was utilized to investigate the structural polymorphism of coexisting GQs and iMs and also the effect of ligand binding on the GQs conformational equilibrium. Inspired by these results, we decided to expand the utility of the 5-trifluoromethyl-benzofuran-modified probe to investigate RNA conformations. We synthesized the TFBF-modified uridine (TFBF-U) analog **1** and its triphosphate **2** that would help to synthesize the modified RNA ON using an enzymatic method. An initial photophysical study revealed that the modified ribonucleoside **1** displayed distinct fluorescence and ^{19}F NMR signatures upon changing solvent polarity and viscosity like the TFBF-dU analog.[16] Triphosphate of the modified ribonucleoside was fruitfully incorporated into RNA with moderate to good efficiency using T7 RNA polymerase catalyzed *in vitro* transcription reaction. Interestingly, incorporation of the modified ribonucleotide into the model RNA transcript distinguishes different flanking bases by fluorescence and ^{19}F NMR. Further, environment-sensitive properties of the modified ribonucleoside were used to monitor the aminoglycoside antibiotics-induced conformational changes in the bacterial ribosomal decoding site RNA using fluorescence and ^{19}F NMR techniques.

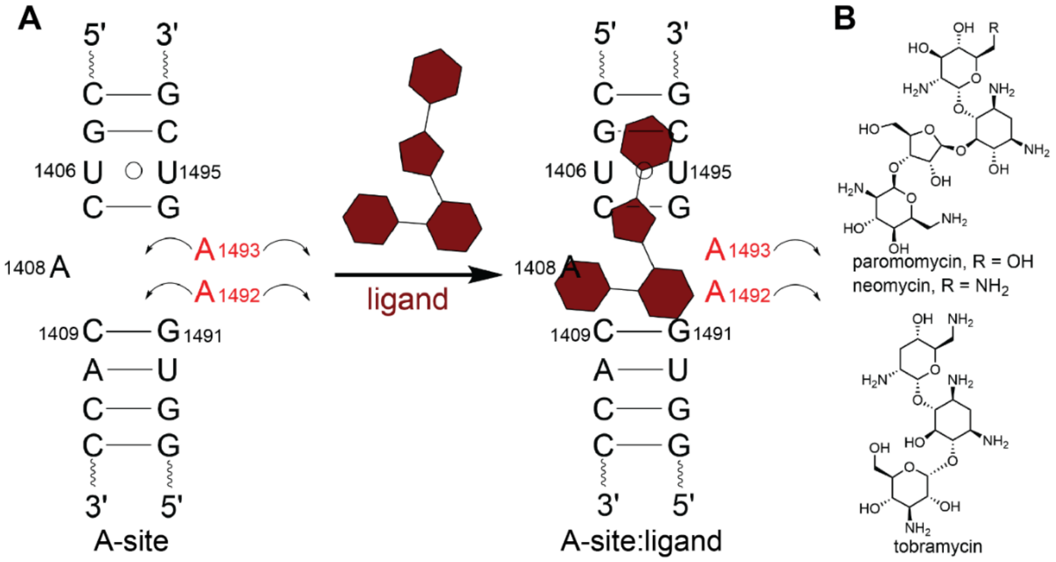
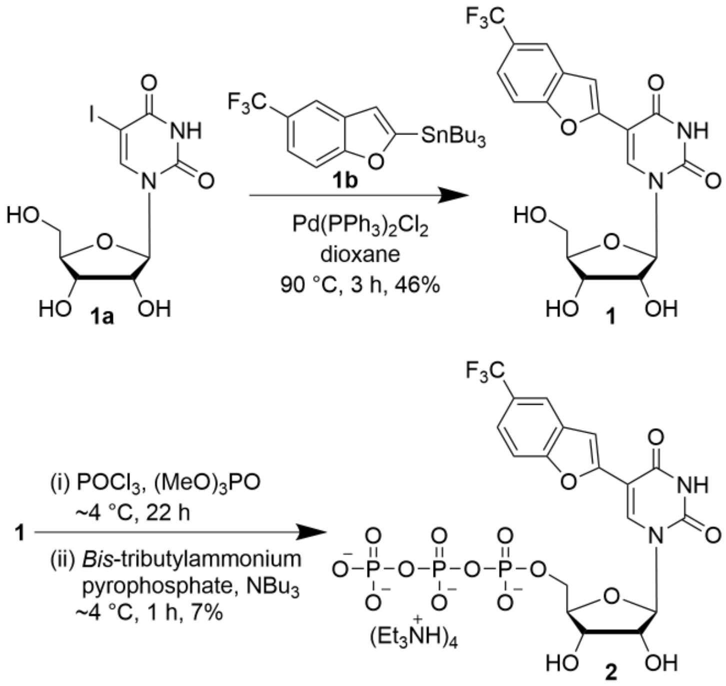


Figure 1. (**A**) Conformations of A-site in aminoglycoside unbound and bound state. (**B**) Structure of aminoglycosides.

5.2 Results and Discussion

5.2.1 Synthesis and photophysical properties of 5-trifluoromethyl-benzofuran-modified uridine 1



Scheme 1. Synthesis of TFBF-U nucleoside **1** and its corresponding triphosphate **2**.

Modified ribonucleoside, 5-trifluoromethyl-benzofuran-modified uridine (TFBF-U) **1**, was synthesized by reacting 5-iodouridine **1a** with stannylated 5-trifluoromethyl-benzofuran **1b** under Stille cross-coupling reaction conditions (Scheme 1). Photophysical and ^{19}F NMR study of the modified nucleoside **1** in solvents of different polarity and viscosity proved the ability of the analog to sense its microenvironment. The ground state configuration of the nucleoside analog **1** was marginally affected upon decreasing the solvent polarity from water to methanol to dioxane (Figure 2A and Table 1). Nevertheless, excited state properties such as emission maxima, quantum yield, and time-resolved fluorescence properties were found to be very sensitive to changes in solvent polarities (Figure 2A, 3A and Table 1). In a polar solvent (water), nucleoside showed an intense emission band (427 nm) with a quantum yield of 0.08 and an average lifetime of 0.97 ns. Upon decreasing the solvent polarity from water to dioxane, it displayed a blue shift in the emission maximum by 27 nm with a hypochromic effect (2 fold) and a shorter lifetime (Figure 2A and Table 1). Moreover, in the TFBF-U analog, 5-trifluoromethyl-benzofuran ring is connected to uridine through a single bond, wherein rotation about this bond would affect the extent of the conjugation[17] and hence, the excited state properties of the modified analog **1**. To evaluate the presence of a molecular rotor element in the modified probe **1**, we performed additional photophysical studies in solvents of similar polarity but different viscosity. Interestingly, significant enhancement in the quantum yield and lifetime of a nucleoside analog **1** was observed with increasing solvent viscosity from water to ethylene glycol to glycerol (Figure 2B and 3B, Table 1). Also, the probe displayed higher anisotropy values in more-viscous solvents. These combined results confirmed the existence of a molecular rotor element.

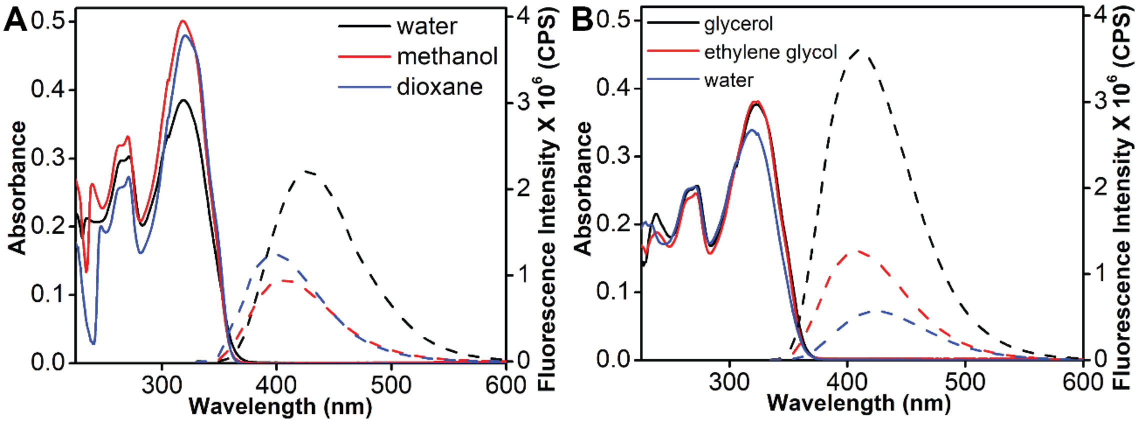


Figure 2. (**A** and **B**) The ground state (25 µM, solid line) and excited state (5 µM, dash line) spectra of modified nucleoside **1** in solvents of different polarity and viscosity, respectively.

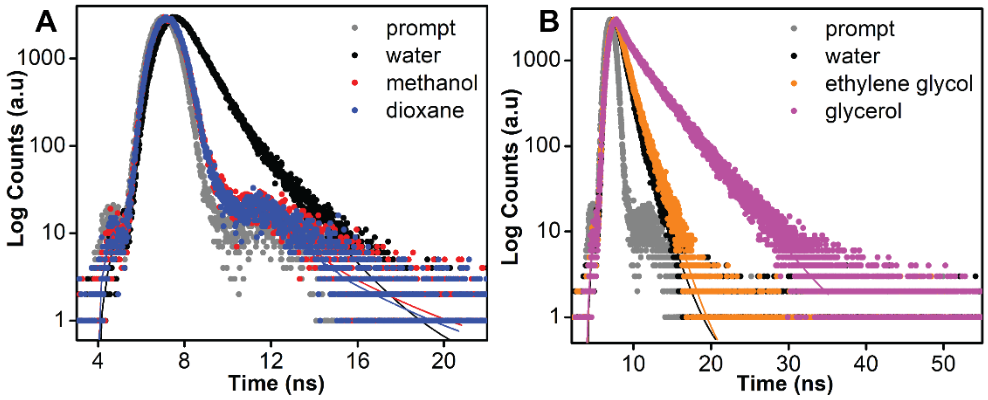


Figure 3. (**A** and **B**) Time-resolved fluorescence (5 µM) spectra of TFBF-U (**1**) in solvents of different polarity (water, methanol, dioxane) and viscosity (water, ethylene glycol, glycerol), respectively. Instrument response (prompt) is indicated in grey dots and curve fits are shown in solid lines.

Table 1. Photophysical properties of TFBF-U (**1**) in different solvents

Solvent	λ_{max}[a] (nm)	λ_{em} (nm)	Φ[b]	τ_{av}[b] (ns)	r[b]
Water	319	427	0.08	0.97	0.10
Methanol	318	406	0.04	0.30	n.d.
Dioxane	320	400	0.05	0.29	n.d.
ethylene glycol	322	407	0.16	1.16	0.31
Glycerol	323	409	0.48	3.21	0.41

[a]Wavelength given corresponds to the lowest energy absorption maximum. [b]Standard deviations for quantum yield (Φ), average lifetime (*τ*av), and anisotropy (*r*) in different solvents are ≤0.005, ≤0.02 ns, and ≤0.001, respectively. n.d. = not determined.

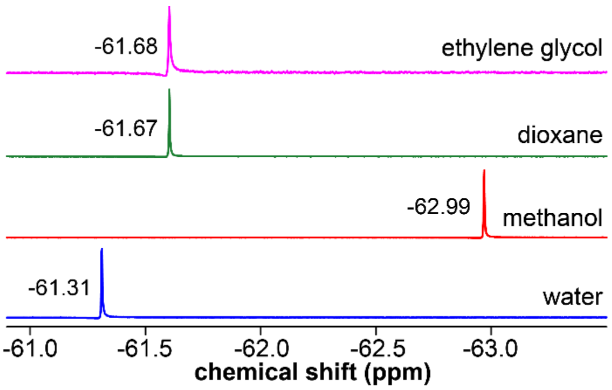


Figure 4. ^{19}F NMR spectra of modified analog (**1**) in different solvents.

Similarly, modified ribonucleoside displayed distinct ^{19}F NMR signals upon changes in the solvent polarity and viscosity (Figure 4). Dipole-dipole interactions between the nucleoside analog and solvent molecules will depend on the polarity of the solvents. Additionally, relative orientation of the heterocyclic ring with uridine ring will vary upon changing the viscosity of the solvents, which can alter the electronic density surrounding the fluorine atom.[18] Hence, the probe could experience different magnitudes of shielding or deshielding effects depending on the polarity or viscosity of the solvents and gave distinct chemical shift signals.

5.2.2 Enzymatic incorporation

Modified RNA can be synthesized by either solid phase synthesis or enzymatic method using RNA polymerases. Here, we have evaluated the incorporation efficiency of triphosphate of TFBF-U (**2**) as a substrate analog of UTP in *in vitro* transcription reaction using T7 RNA polymerase. TFBF-UTP was synthesized as per the steps given in Scheme 1. Firstly, TFBF-U was reacted with dry $POCl_3$, followed by bis-tributylammonium pyrophosphate at ~4 °C to obtain TFBF-UTP. Transcription reactions were performed using different DNA templates **T1**–**T5** (Figure 5). The templates were designed such that the coding region contains one or two dA residues, which will guide the placement of modified nucleotide **2** into the RNA oligonucleotides (ONs) during *in vitro* transcription reaction. DNA templates were annealed with T7 promoter DNA sequence and the transcription reaction was started by adding the enzyme into a solution containing duplex, GTP, CTP, ATP, and UTP/**2**. Reaction mixtures were analyzed by PAGE under denaturing conditions, stained using stains-all, and imaged on a Typhoon scanner (Figure 6).

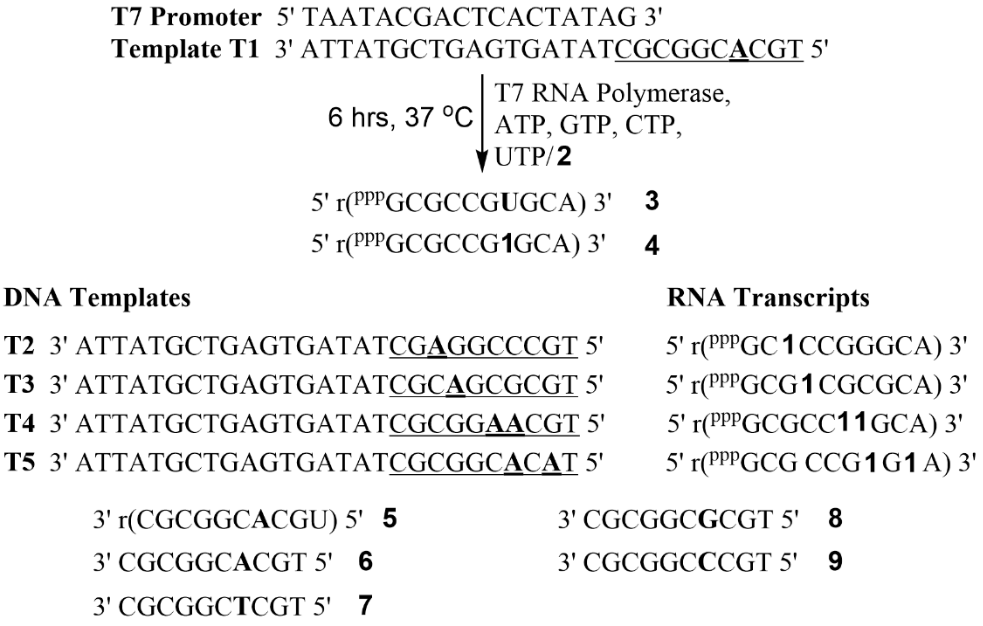


Figure 5. Incorporation of the TFBF-UTP (**2**) into RNA ONs using different DNA templates (**T1**−**T5**) by in vitro transcription reaction. Complementary RNA ON **5** and DNA ONs **6**−**9** were used in this study.

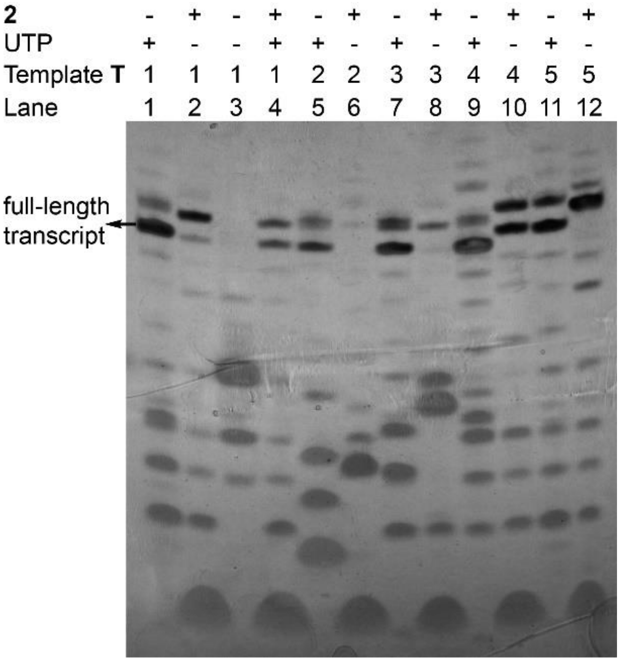


Figure 6. PAGE analysis of products of in vitro transcription reactions performed using DNA templates **T1**–**T5** in the presence of natural UTP and modified triphosphate **2** under denaturing conditions.

Furthermore, full-length transcription products formed in the presence of natural UTP or TFBF-UTP were compared to obtain the relative incorporation efficiency of TFBF-UTP in various templates. In the case of template **T1**, in which the place of the modification was a little far from the transcription start site, the modified UTP produced a full-length transcript **4** with a good yield (~79%) along with the minor non-templated incorporation products (N+1 and N+2) (Figure 6, lane 2). It is well known that the non-templated transcription products could be formed during *in vitro* transcription reactions.[19] In a control experiment, transcription reaction did not yield a full-length transcript in the absence of UTP and **2**. This dismisses any misincorporation possibility by the enzyme (lane 3). Importantly, transcription reaction in the presence of equimolar concentrations of **2** and natural UTP indicated that the enzyme has comparable preferences for both triphosphates (lane 4 – ~52% control RNA and ~48% modified RNA). Next, transcription reactions with templates **T2** and **T3**, in which modification will be near the transcription start site yielded full-length products with a very lower yield. This is because the enzyme poorly tolerates modifications near the transcription start site.[19] Interestingly, templates **T4** and **T5** containing double modifications (adjacent and alternate) produced full-length products with moderate yields. Afterward, a TFBF-U labeled RNA transcript **4** was synthesized by a large-scale transcription reaction with template **T1** and purified by PAGE under denaturing conditions. The purity and integrity of RNA transcript **4** were confirmed from the reverse-phase HPLC and MALDI-TOF analysis (Figure 7A and 7B). Further, incorporation of the modified nucleoside (**1**) in transcript **4** was validated from the enzymatic digestion assay (Figure 7C).

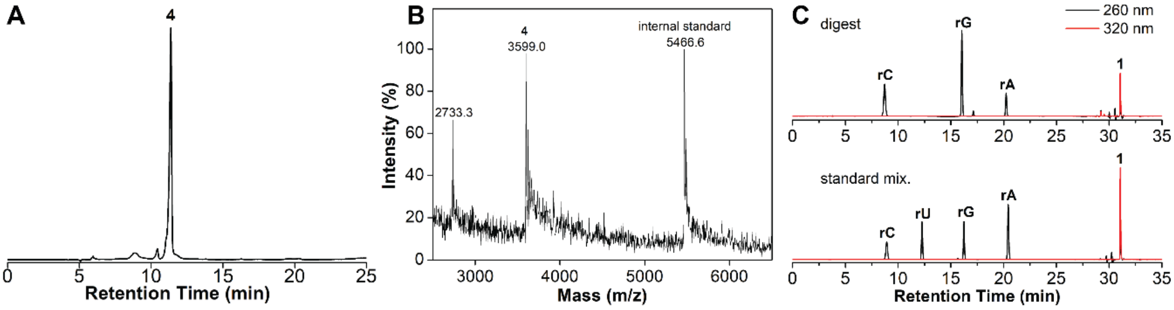


Figure 7. (**A**) RP-HPLC chromatogram of PAGE purified RNA transcript **4**. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron). (**B**) MALDI-TOF analysis of RNA transcript **4**, spectrum calibrated relative to the internal standard of 18-mer DNA ON (m/z of +1 and +1/2 ions are 5466.6 and 2733.3). m/z calculated for transcript **4** = 3598.4, Observed = 3599.0 (**C**) RP-HPLC chromatogram of the reaction mixture of enzymatic digestion of transcript **4** (upper part), and a mixture of natural ribonucleosides and **1** (lower part).

5.2.3 Photophysical properties of TFBF-U depends on changes in neighboring base environment.

Electronic properties of nucleoside probe **1** in ONs could depend on the nature of interactions (e.g., stacking, H-bonding, electron transfer) with neighboring bases. To determine the responsiveness of nucleoside analog **1** upon changes in the neighboring base environment, modified transcript **4** was hybridized with perfect complementary RNA/DNA ONs **5**/**6** and with complementary DNA ONs **7**–**9** containing mismatched bases. UV-thermal melting analysis was performed to understand the effect of modification on the stability of perfect duplexes. Similar T_m were observed for the control unmodified and modified duplexes, suggesting that the modification has negligible impact on the duplex stability (Figure 8, Table 2). Interestingly, modified nucleoside in a single-stranded RNA displayed a very high fluorescence with an emission band centered at 427 nm (Figure 9A and 9B). In the case of duplexes, in which nucleoside **1** was placed opposite to either A/dA in perfect duplexes, or dT/dC in mismatched duplexes, it showed comparably quenched fluorescence with a slight blue shift in emission maxima. This quenching in the fluorescence with blue shift effect could be due to the stacking interactions experienced by the modified nucleoside in duplex form.[20] Notably, when the nucleoside probe was placed opposite dG in a mismatched duplex, it exhibited high fluorescence, indicating that nucleoside selectively detects dG mismatch in RNA-DNA duplex (Figure 9A and 9B). Further, the modified nucleoside displayed a distinct ^{19}F NMR signal for the single-stranded RNA, its perfect duplexes and mismatched duplexes (Figure 9C). Combined results suggest that the fluorescence and ^{19}F NMR properties of the dual-app nucleoside are sensitive to changes in the neighboring base environment. Hence, this dual-channel probe could be useful to investigate RNA conformations and its interactions with ligands by fluorescence and ^{19}F NMR techniques.

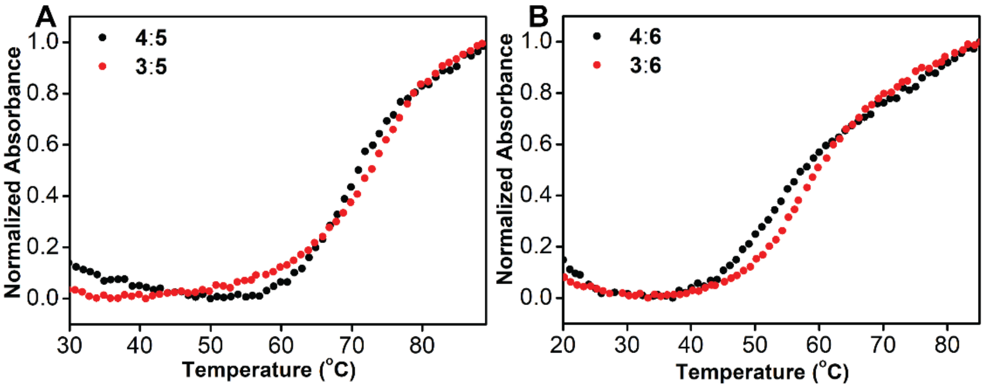


Figure 8. UV-thermal melting study of (**A**) modified (**4**:**5**) and control unmodified (**3**:**5**) RNA:RNA duplexes. (**B**) modified (**4**:**6**) and control unmodified (**3**:**6**) RNA:DNA duplexes.

Table 2. T_m values of modified and control unmodified duplexes.

	4:5	3:5	4:6	3:6
T_m (°C)	69.4 ± 0.4	72.8 ± 0.1	55.5 ± 0.6	57.6 ± 1.4

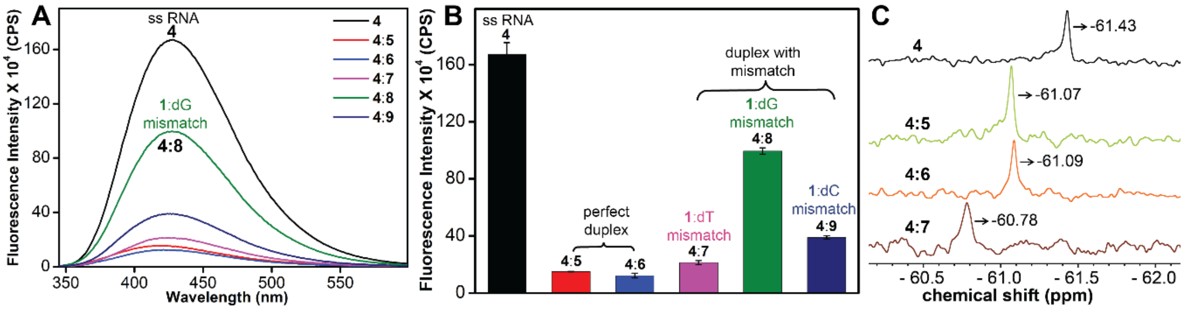


Figure 9. (**A**) Fluorescence spectra, (**B**) fluorescence intensity plotted at 427 nm, and (**C**) ^{19}F NMR spectra of the modified RNA **4** and its duplexes made by hybridizing with complementary ONs **5**–**9**, respectively.

5.2.4 Detection of aminoglycoside-antibiotics and A-site RNA interactions using fluorescence and ^{19}F NMR techniques

Dual-label nucleoside analog **1** was incorporated into the A-site RNA ON construct using enzymatic method. To incorporate single TFBF-UTP in an A-site RNA using an enzymatic reaction, we selected a short RNA ON construct, which is an essential part of the wild-type A-site for the binding of aminoglycosides (Figure 10A).[21] As RNA ON **10** contains a single uridine (U_{1406}), we synthesized the modified ON **10** carrying a single modified analog **1** by large-scale transcription reaction in the presence of TFBF-UTP. The integrity of the RNA transcript was confirmed by mass analysis (Figure 11). Further, the labeled A-site (**10:11**) was constructed by hybridizing modified RNA transcript **10** with its complementary strand **11** (1:1) in HEPES buffer (Figure 10B). UV-thermal melting analysis of the labeled A-site and native unmodified A-site displayed similar T_m, indicating that the modification has negligible impact on the formation of the short RNA duplex (A-site) (Figure 12).

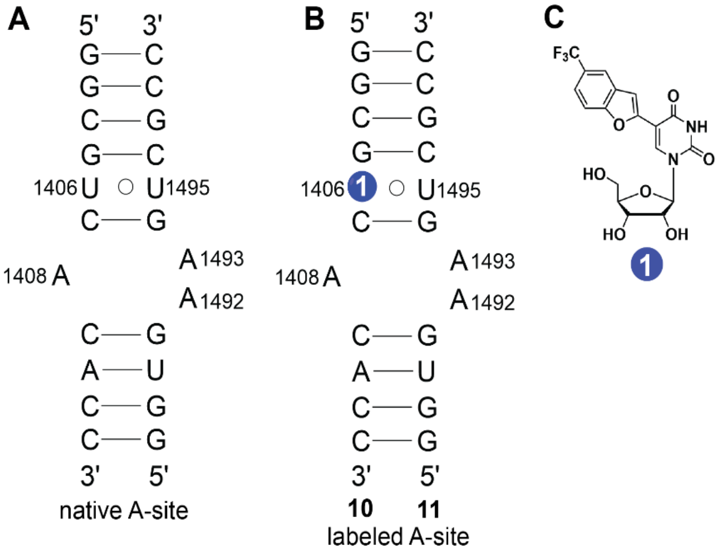


Figure 10. (**A** and **B**) Secondary structure of the native and labeled bacterial A-site constructs that bind to aminoglycosides antibiotics. (**C**) Structure of modified nucleoside.

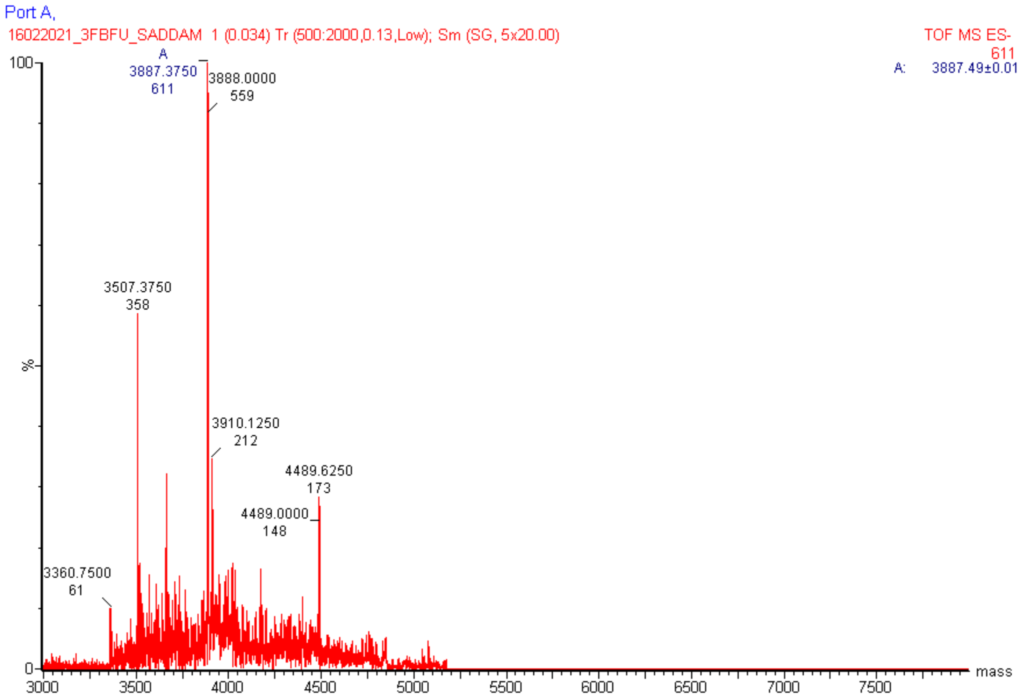


Figure 11. ESI-MS spectra of modified ON **10**. Calculated mass = 3888.21, Observed mass = 3888.00.

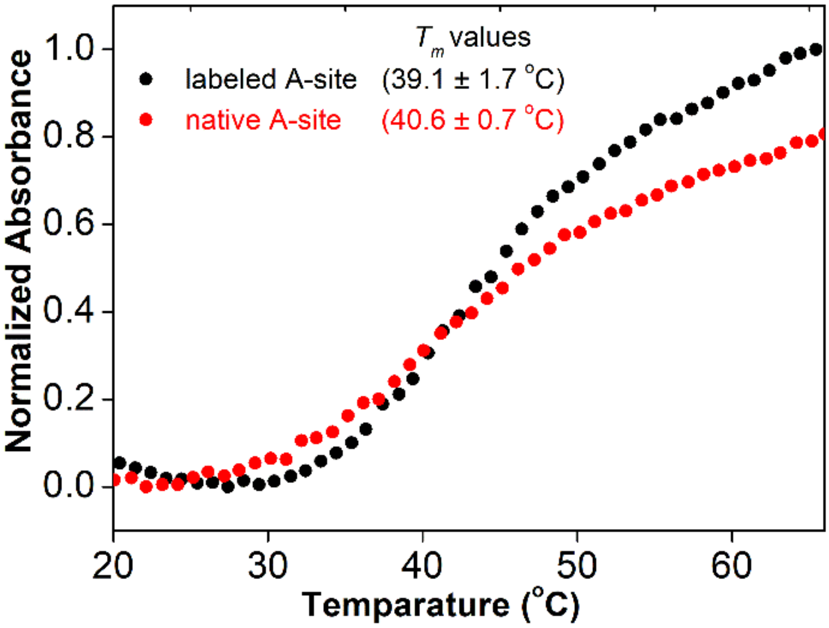


Figure 12. UV-thermal melting study of the labeled A-site and native unmodified A-site. T_m values are mentioned in the figure for the corresponding duplex.

Next, the labeled A-site construct was excited at 330 nm, and change in the fluorescence intensity with an increasing concentration of aminoglycosides was monitored. Titration of paromomycin with A-site displayed dose-dependent quenching in the emission intensity, which is almost a two-fold decrease at the saturation point of paromomycin (Figure 13A). Normalized fluorescence intensity at 425 nm respective to the concentration of paromomycin offered a K_d value of 1.80 ± 0.12 µM (Figure 13D), which is in good agreement with the literature reports ($K_d \sim 10^{-6}$ M).[14] Similarly, binding of the neomycin and tobramycin with modified A-site resulted in quenching in the fluorescence intensity with their corresponding apparent K_d values 0.71 ± 0.06 µM and 1.38 ± 0.10 µM, respectively (Figure 13B, 13C and 13D). The higher binding affinity displayed by the neomycin compared to the paromomycin and tobramycin is in good agreement with the literature reports. Crystal structure data of aminoglycoside and A-site complex revealed that the aminoglycosides form a water-mediated hydrogen bond with O_4 atoms of $U_{1406}{\circ}U_{1495}$ base pair, and also uracil ring of $U_{1406}{\circ}U_{1495}$ base pair undergo conformational transition.[15b] In this conformation, modification at U_{1406} could experience a hydrophobic atmosphere due to hydrogen bonding. Additionally, it could experience stacking interactions with adjacent bases and hence, could produce quench fluorescence. Literature reports suggest that a bulge of the A-site is crucial for the binding of aminoglycosides. In a control experiment, titration of aminoglycosides (paromomycin/neomycin/tobramycin) with a single-stranded RNA ON **10**, which could not form a bulge structure didn't show a significant change in emission intensity (Figure 14A, 14B and 14C). This confirms that difference in the emission profile exhibited by modified nucleoside in the A-site construct was due to the aminoglycoside-induced conformational changes. During this titration, volume of the RNA sample was increased by around 8%. Hence, to understand the effect of dilution, we performed an additional control experiment and found that the fluorescence intensity of the modified A-site was not affected by addition of water aliquots (Figure 14D).

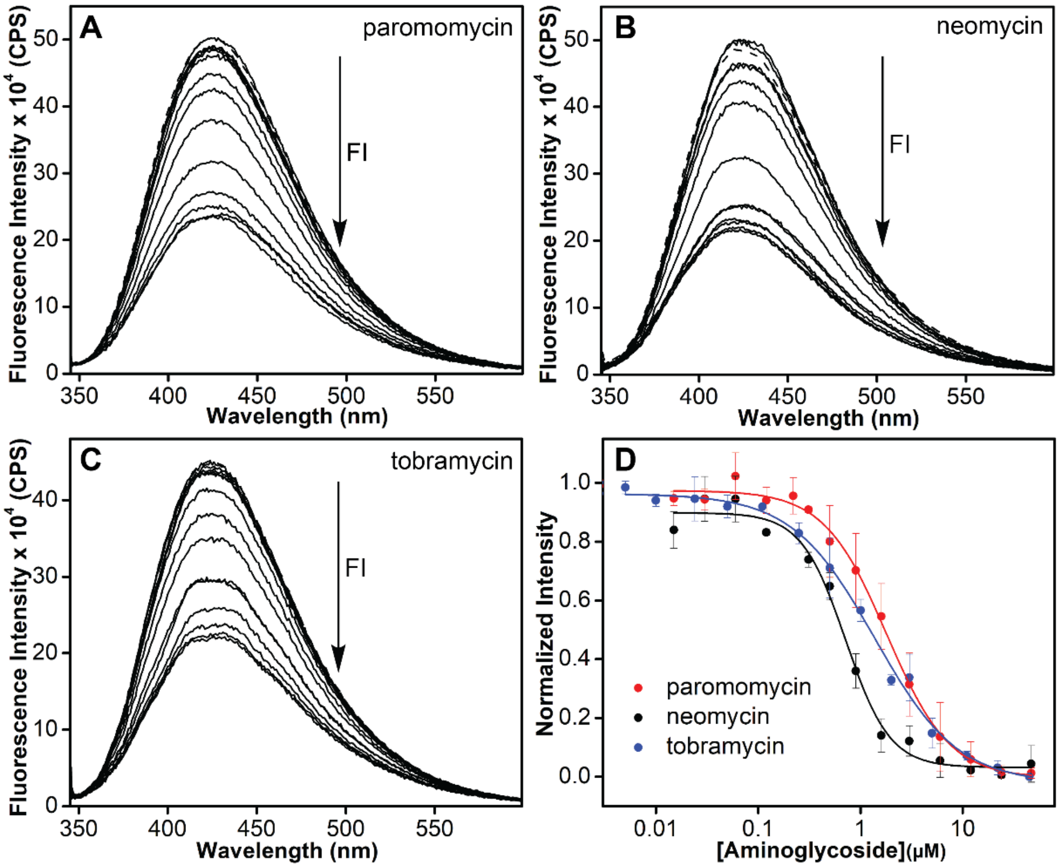


Figure 13. Emission spectra (solid line) of the modified A-site (**10**:**11**) with increasing concentrations of (**A**) Paromomycin, (**B**) Neomycin, and (**C**) Tobramycin. The dashed line indicates the fluorescence emission in the absence of aminoglycosides, FI represents fluorescence intensity. Samples were excited at 330 nm with excitation and emission slit widths of 7 and 8 nm, respectively. (**D**) Curve fits were obtained for the titration of modified A-site with aminoglycosides by plotting normalized fluorescence intensity at 425 nm against the concentrations of aminoglycosides.

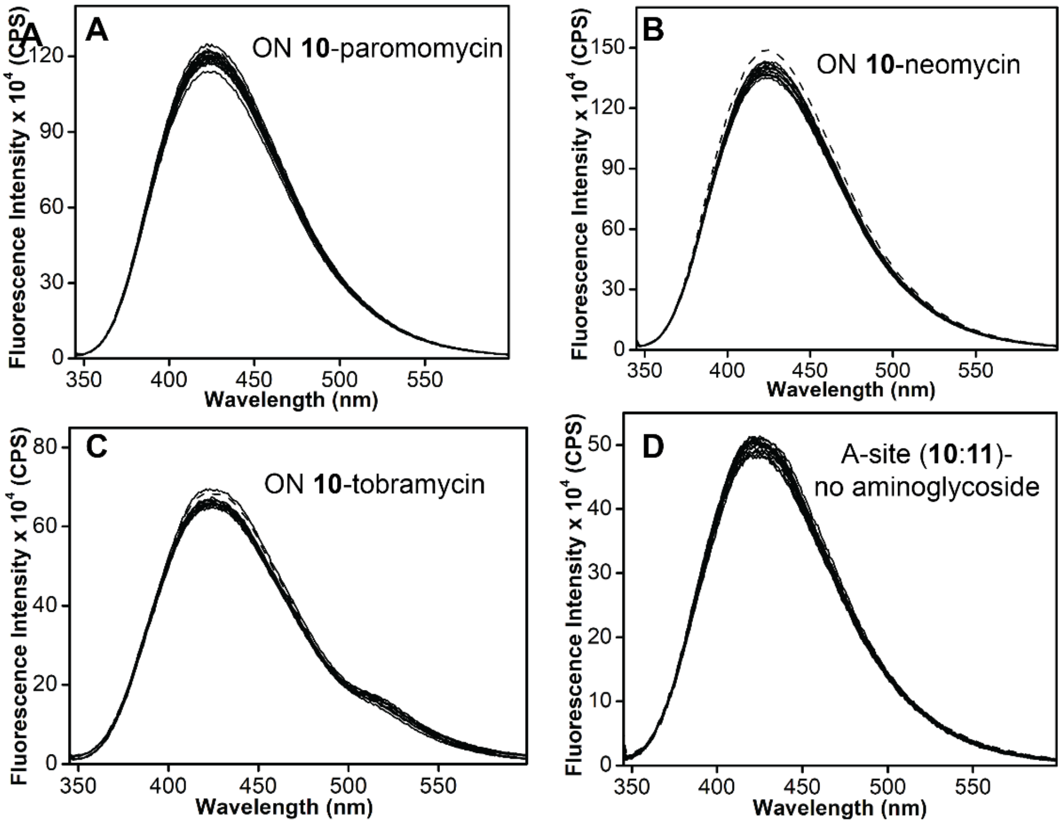


Figure 14. Fluorescence emission spectra of the modified ON **10** with increasing concentrations of (**A**) Paromomycin, (**B**) Neomycin, and (**C**) Tobramycin. (**D**) Fluorescence spectra of the modified A-site constructs (**10**:**11**) after addition of water aliquotes. Samples were excited at 330 nm with excitation and emission slit widths of 7 and 8 nm, respectively.

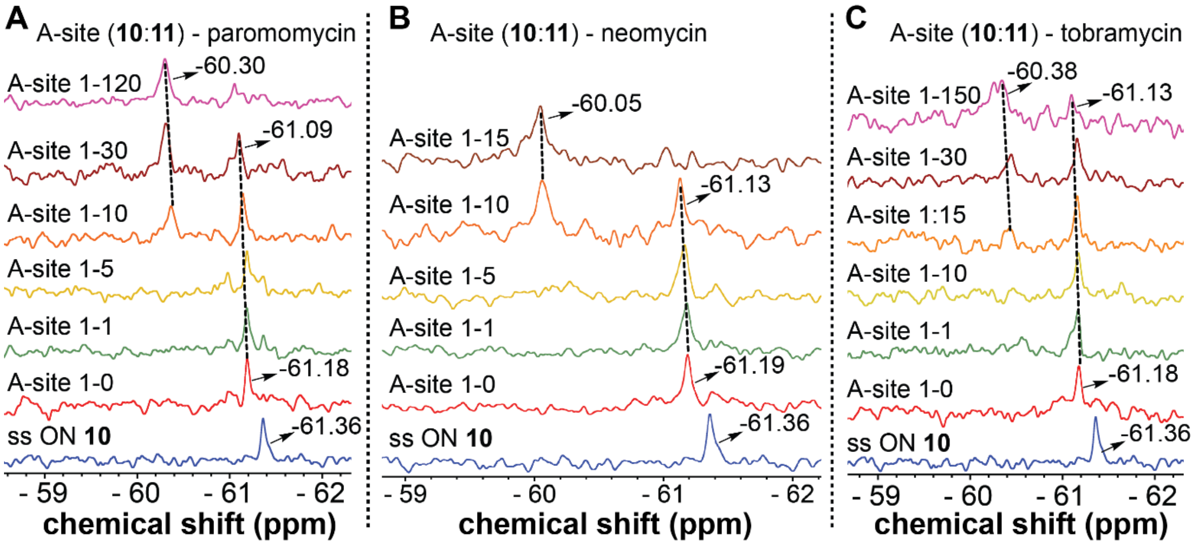


Figure 15. The ^{19}F NMR spectra of ss ON **10**, and A-site constructs (**10**:**11**) with increasing concentration of the aminoglycosides. (ss represent single strand)

We then monitored the interactions of aminoglycosides with labeled A-site construct using ^{19}F NMR technique. The ^{19}F label of our modified nucleoside exhibited unique NMR signatures for individual structures. It displayed a ^{19}F NMR peak at -61.36 ppm for the single-stranded RNA **10**, which shifted downfield (-61.18 ppm) upon formation of the A-site construct (**10:11**) (Figure 15). A similar downfield shift in the ^{19}F NMR was observed for model duplex structures compared to its single strand (compare with Figure 9C). Titration of aminoglycosides with a modified A-site yielded a new downfield shifted peak for the formation of A-site-paromomycin complex (-60.30 ppm), A-site-neomycin complex (-60.05 ppm) and A-site-tobramycin complex (-60.38 ppm), respectively (Figure 15A, 15B and 15C). Collectively, these results illustrated that the dual-app nucleoside analog **1** is an effective probe in detecting drug-induced conformational change in RNA. Previously, the incorporation of 2AP, a fluorescent adenosine analog into the A-site has been utilized to report the binding of paromomycin and tobramycin, but it failed to detect the binding of neomycin which is a strong binder. In this context, dual-label TFBF-U nucleoside analog **1** has an upper edge over other fluorescent nucleosides and could be useful to examine RNA structures and identify new binders by two powerful techniques like fluorescence and ^{19}F NMR.

5.3. Conclusions

We have developed a microenvironment-sensitive ribonucleoside analog, which acts as a dual channel probe and reports changes in solvent polarity and viscosity by fluorescence and ^{19}F NMR. Modified nucleotide act as a good substrate for T7 RNA polymerase and yielded labeled RNA ONs with moderate to good efficiency. Interestingly, when incorporated into RNA ONs, it reports changes in its flanking bases, particularly, it recognizes G-mismatched in the duplex structure by fluorescence and ^{19}F NMR. Further, responsiveness of the modified nucleoside was utilized in monitoring the aminoglycoside antibiotics-induced conformational changes in the A-site RNA motif. Fluorescence and ^{19}F NMR-based methods were successfully developed using one probe system to detect the binding event. Taken together, TFBF-U could be used to develop a biophysical platform to investigate the RNA structure, dynamics and also to set up a small-molecule screening platform to identify A-site binders.

5.4 Experimental Section

5.4.1. Materials

Bis(triphenylphosphine)-palladium(II) dichloride, tributylamine, trimethyl phosphate and all reagents/salt (Bio-Ultra grade) used for buffer preparation were purchased from Sigma Aldrich. 5-Iodouridine (**1a**)[22] and tributyl (5-trifluoromethyl benzofuran-2-yl) stannane (**1b**)[16] was synthesized using previously reported procedures. $POCl_3$ was procured from Acros Organic and was freshly distilled before use. T7 RNA polymerase, ribonuclease inhibitor (RiboLock), NTPs, RNase A and RNase T1 were purchased from Thermo Fisher Scientific. Calf intestinal alkaline phosphatase (CIP) and snake venom phosphodiesterase I were purchased from Invitrogen and Sigma-Aldrich, respectively. RNA ON **5**, **11** and DNA ONs **6**–**9** were procured from Dharmacon RNAi Technologies and Integrated DNA Technology, respectively. All custom DNA and RNA ONs were purified using polyacrylamide gel electrophoresis under denaturing conditions and desalted on Sep-Pak Classic C18 cartridges (Waters Corporation). Autoclaved Millipore water was utilized in the preparation of all buffer solutions and in all biophysical studies.

5.4.2. Instrumentation

NMR analysis of small molecules was performed on a 400 MHz Jeol ECS-400 and Bruker AVANCE III HD ASCEND 400 or 600 MHz spectrometer. NMR spectra were processed using Mnova software from Mestrelab Research. Mass of the small molecules was recorded using an ESI-MS Waters Synapt G2-Si Mass Spectrometry instrument or on an Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer. HPLC analysis was carried out using reverse-phase flash chromatography (C18 RediSepRf column) on Agilent Technologies 1260 Infinity HPLC. Modified nucleotide was purified using Teledyne ISCO, Combi Flash Rf. UV absorption spectra were acquired on a Shimadzu UV-2600 spectrophotometer. Steady-state fluorescence analysis of modified nucleoside and RNA transcripts were performed on a Fluoromax-4 spectrophotometer. Fluorescence decay of a modified nucleoside was recorded on a TCSPC Fluorolog 3 (Horiba Jobin Yvon). A UV-thermal melting experiment of ONs was carried out on Cary 300 Bio UV-Vis spectrophotometer. NMR spectra of ONs were recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with a Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT) and processed using Bruker TopSpin Software.

5.4.3. Synthesis of 5-(5-trifluoromethyl-benzofuran-2-yl)-uridine and its corresponding triphosphate

5.4.3.1. 5-(5-trifluoromethyl-benzofuran-2-yl)-uridine (1)

In dry dioxane, 5-iodouridine (**1a**)[22] (0.50 g, 1.35 mmol, 1.0 equiv.) was added and solution was degassed for 15 min by purging N_2 gas. Further, compound **1b**[16] (1.16 g, 2.43 mmol, 1.8 equiv.) and bis(triphenylphosphine)-palladium (II) dichloride (0.05 g, 0.07 mmol, 0.05 equiv.) were added to the above solution and the reaction mixture was heated at 90 °C for 3 h under a nitrogen atmosphere. The reaction was monitored by TLC. After completion of the reaction, a suspension was filtered through a celite pad, and the celite pad was rinsed with warm dioxane. The filtrate was evaporated to dryness and the crude product was purified using reversed-phase flash chromatography (~60% methanol in water) to afford pure product **1** as a white solid (0.27 g, 46%). TLC (10% MeOH in DCM), R_f = 0.47; ^{1}H NMR (d_6-DMSO, 400 MHz): δ (ppm) 11.83 (bs, 1H), 9.02 (s, 1H), 8.07 (bs, 1H), 7.78 (d, 1H, J = 8.8 Hz), 7.64 (dd, 1H, J = 8.8 Hz, 1.6 Hz), 7.46 (s, 1H), 5.86 (d, 1H, J = 4 Hz), 5.51 (d, 1H, J = 5.2 Hz), 5.45 (t, 1H, J = 4 Hz), 5.13 (d, 1H, J = 5.6 Hz), 4.16–4.08 (m, 2H), 3.97 (m, 1H), 3.83 (dt, 1H, J = 11.6 Hz, 3.2 Hz), 3.69 (d, 1H, J = 11.2 Hz); ^{13}C NMR (*d6*-DMSO, 100 MHz): δ (ppm) 160.2, 154.6, 151.3, 149.5.0, 138.3, 129.3, 124.8 (q, 270.1 Hz), 124.1 (q, 31.2 Hz), 121.1 (q, 3.5 Hz), 118.6 (q, 3.9 Hz), 111.6, 104.0, 103.7, 89.0, 84.5, 74.5, 69.2, 59.9; ^{19}F NMR (*d6*-DMSO , 376.6 MHz): δ (ppm) -59.25; HRMS: m/z calculated for $C_{18}H_{14}F_3N_2O_7$ $[M-H]^-$ = 427.0753, found = 427.0755; λ_{max} (H_2O) = 265, 271 and 319 nm, ε_{265} = 11.75 x 10^3 $M^{-1}cm^{-1}$, ε_{271} = 12.00 x 10^3 $M^{-1}cm^{-1}$, ε_{319} = 15.32 x 10^3 $M^{-1}cm^{-1}$, ε_{260} = 10.91 x 10^3 $M^{-1}cm^{-1}$.

5.4.3.2. 5-(5-trifluoromethyl-benzofuran-2-yl)-uridine triphosphate (2)

5-(5-trifluoromethyl-benzofuran-2-yl)-uridine (**1**) (70 mg, 0.16 mmol, 1.0 equiv.) was dissolved in dry trimethyl phosphate, a solution was cooled in ice bath for 15 min and freshly distilled $POCl_3$ (38 µL, 0.41 mmol, 2.5 equiv.) was added to that. The reaction mixture was stirred for 24 h at ~ 4 °C. From TLC, it was found that the starting material was not fully consumed even after 24 h. Further, bis(tributylammonium) pyrophosphate solution in DMF (0.5 M, 1.64 ml, 0.82 mmol, 5.0 equiv.) and tributylamine (0.43 ml, 1.80 mmol, 11.0 equiv.) were simultaneously added to the reaction mixture at ~ 4 °C. The reaction was stirred for 30 min at 4 °C and stopped by adding 15 mL of 1 M triethyl ammonium bicarbonate buffer (pH 7.5). The solution was washed with ethyl acetate (2 x 15 ml) and the aqueous layer was concentrated over the rotary evaporator. The crude product was initially purified using a DEAE Sephadex-A25 anion exchange column (mobile system 10 mM–1M TEAB buffer, pH 7.5) followed by reverse phase flash chromatography (C18 RediSepRf, 0–50% acetonitrile in 100 mM triethylammonium acetate buffer, pH 7.4, flow rate 5 ml/min, run time 65 min). The appropriate fraction was evaporated to dryness and lyophilized three times to get the desired triphosphate **2** as a salt of tetraethyl ammonium (13 mg, 7%). ^{1}H NMR (D_2O, 600 MHz): δ (ppm) 8.38 (s, 1H), 7.92 (s, 1H), 7.84 (d, 1H, J = 8.4 Hz), 7.63 (d, 1H, J = 8.4 Hz), 7.31 (s, 1H), 6.00 (d, 1H, J = 5.4 Hz), 4.52 (t, 1H, J = 5.4 Hz), 4.48−4.47 (m, 1H), 4.34−4.29 (m, 3H); ^{13}C NMR (D_2O, 150 MHz): δ (ppm) 180.3, 161.8, 155.1, 150.6, 149.5, 137.5, 128.8, 124.8 (q, 31.4), 124.7 (q, 269.0), 121.8, 118.6, 112.2, 106.2, 104.9, 88.5, 83.8, 83.7, 73.7, 70.0, 65.4; ^{19}F NMR (D_2O, 564.9 MHz): δ (ppm) -61.59; ^{31}P NMR (D_2O, 162 MHz) δ (ppm) -10.57 (br, P_γ), -11.79 (br, P_α), -23.04 (br, P_β); HRMS: m/z Calculated for $C_{18}H_{17}F_3N_2O_{16}P_3$ $[M-H]^-$ = 667.2490, found = 666.9430.

5.4.4. Photophysical properties of 5-(5-trifluoromethyl-benzofuran-2-yl)-uridine 1 in different solvents

5.4.4.1. UV absorption and steady-state fluorescence

UV absorption analysis of the modified ribonucleoside **1** (25 µM) was performed in solvents of different polarity (water, methanol, and dioxane) and viscosity (water, ethylene glycol, and glycerol) in a quartz cuvette (Hellma, path length 1 cm) and in triplicate reading. Around 2.5% of DMSO was contained in each sample to solubilize probe **1**. Steady-state fluorescence of the analog **1** (5 µM) was recorded in different solvents with triplicate. Each sample contains 0.5% DMSO. Samples in a micro-fluorescent cuvette (Hellma, path length 1 cm) were excited at the wavelength corresponding to their lowest energy, and measurement was carried out on a Fluorolog-3 spectrophotometer (Horiba Jobin Yvon). For Figure 2A, 2B excitation and emission slit widths were kept at 3 and 4 nm and 3 and 3 nm, respectively.

5.4.4.2. Quantum yield calculation

The quantum yield of nucleoside **1** was calculated in different solvents relative to 2-aminopurine as a standard using the following equation.[16]

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

Here, s stands for the standard, x is nucleoside **1**, A is absorbance at the excitation wavelength, F is the area under the emission curve, n is the refractive index of the solvent, and Φ_F is quantum yield. The quantum yield of 2AP in water is 0.68.

5.4.4.3. Time-Resolved Fluorescence

The fluorescence decay of modified nucleoside **1** (5 μM) was recorded in various solvents on a TCSPC instrument (Fluorolog-3, Horiba Jobin Yvon). The samples were excited using a 339 nm LED source (IBH, UK, NanoLED-339 L), and the decay of probe at the corresponding emission wavelength was acquired in triplicate. The decay profiles were fitted using IBH DAS6 software with χ^2 (goodness of fit) values nearly one to obtain the individual lifetime values.

5.4.4.4. ^{19}F NMR experiment of Nucleoside 1

^{19}F NMR analysis of the nucleoside analog **1** (150 μM) was performed in solvents of different polarity and viscosity. All samples contained 15% d_6-DMSO, and each NMR spectra was calibrated using trifluorotoluene (TFT) (-63.72 ppm) as an external standard.

5.4.5. Transcription reaction

Promoter-template duplexes (5 μM) were constructed by hybridizing DNA templates **T1**–**T5** with T7 promoter DNA sequence (1:1) in an annealing buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) at 90 °C for 3 min. Samples were slowly cooled to room temperature, later placed on crushed ice for 30 min and stored at -40 °C. In vitro transcription reaction (20 μL reaction volume) in 40 mM Tris-HCl buffer (pH 7.8) was carried out at 37 °C using 300 nM of 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 ATP, CTP, GTP and UTP/modified UTP **2**, 3 U/μL (total 60 units) T7 RNA polymerase. To stop the reaction, 20 μL of loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8) was added to that after 4 h. The samples were heated at 75 °C for 3 min, cooled it on an ice bath and 10 μL of each sample were loaded on a 20% denaturing polyacrylamide gel. The electrophoresis experiment was carried out for 4 h at constant power (12 W). RNA was stained on the gel using stains-all and imaged on a Typhoon scanner. The incorporation efficiency of TFBF-UTP **2** into RNA transcripts was estimated by comparing with the incorporation efficiency of natural UTP.

5.4.6. Large-scale transcription reaction

Transcription reactions were performed on a large-scale (reaction volume = 250 μL) using a template **T1** and the experimental conditions as mentioned above to get control transcript **3** and modified transcript **4**. Reaction was performed using 300 nM of promoter-template DNA duplex, 2 mM of ATP, CTP, GTP and UTP or TFBF-UTP, 20 mM of $MgCl_2$, 0.4 U/μL RNase inhibitor (Riboblock), 800 units T7 RNA polymerase and incubated at 37 °C for 12 h. The reaction mixture was centrifuged to remove the precipitated salt of pyrophosphate, and speed-Vac to reduce the volume by nearly 1/3 of the total. Around 40 µL of the loading buffer (7 M urea in 10 mM Tris-HCl, 100 mM EDTA, pH 8) was added to that and heated at 75 °C for 3 min. The solution was cooled on an ice bath and loaded on 20% denaturing PAGE. The gel electrophoresis was continued for 5.3 h at constant power of 25 W. The desired band of a full-length transcript was recognized by UV shadowing. The gel band was transferred in the poly-prep column (Bio-rad) and crushed that. The gel mass was soaked into 0.5 M of ammonium acetate solution for 12 h and desalted using Sep-Pak classic C18 cartridges (Waters) to extract the RNA transcript. Approximately 8 to 12 nmoles of transcript **4** (ε_{260} = 92853 $M^{-1}cm^{-1}$) was obtained from each reaction. Similarly, TFBF-U modified A-site oligonucleotide **10** was synthesized using a DNA template 5′ GGTGTGACGCCTATAGTGAGTCGTATTA 3′. A Large-scale transcription reaction for this template gave nearly 6 to 7 nmoles (ε_{260} = 102753 $M^{-1}cm^{-1}$) of RNA transcript **10**.

5.4.7. Mass analysis of modified transcripts.

5.4.7.1. MALDI TOF

The mass of transcript **4** was obtained using Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer. A solution containing 2 µL of transcript **4** (200 µM), 2 µL of an internal DNA standard (100 µM), 4 µL of an 8:2 solution of 3-hydroxypicolinic acid and ammonium citrate buffer (100 mM, pH 9) was desalted by adding an ionexchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form). Solution (2 µL) was spotted on a MALDI plate, and used for mass analysis. The MALDI spectrum was referenced relative to the mass of an internal DNA standard. Internal DNA standard sequence 5' TAATACGACTCACTATAG 3', m/z of +1 and +2 ions are 5466.6 and 2733.3.

5.4.7.2. ESI-MS

The molecular weight of transcript **10** was determined using ESI-MS analysis in negative mode by injecting ON (~300 pmol) dissolved in 50% acetonitrile in an aqueous solution of 10 mM triethylamine and 100 mM hexafluoro-2-propanol.

5.4.8. Enzymatic digestions of transcript 4

Transcript **4** (3.5 nmole) was added to a solution containing snake venom phosphodiesterase I (0.01 U), calf intestinal alkaline phosphatase (10 µL, 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 and incubated for 12 h at 37 °C. Next, RNase T1 (0.2 U/µL) was added to the above solution and was incubated further for 4 h at 37 °C. The solution of RNA digest was filtered through a 0.2 micron syringe filter and analysed by reverse-phase HPLC using Phenomenex-Luna C18 column (250 × 4.6 mm, 5 micron) at 260 and 320 nm. Mobile phase A: 50 mM TEAA buffer (pH 7.5), mobile phase B: acetonitrile. Flow rate: 1 ml/min. Gradient: 0−10% B in 20 min, 10−100% B in 10 min.

5.4.9. UV-thermal denaturation experiments

Modified and control RNA ON **3** and **4** were annealed with their complementary RNA/DNA ON **5**/**6** (in 1:1 ration) in 20 mM cacodylate buffer (pH 6.9, 250 mM NaCl, 0.5 mM EDTA) by heating at 90 °C for 3 min. Samples (1 µM) were slowly cooled to room temperature and stored at 4 °C for 1 h. Similarly, native and labeled A-site constructs were hybridized in 20 mM cacodylate buffer (pH 6.9, 100 mM NaCl, 0.5 mM EDTA). UV-thermal melting analysis was performed at 260 nm.

5.4.10. Fluorescence study of transcript 4 and its duplexes

Modified transcript **4** was annealed with its perfect complementary RNA/DNA sequence **5**/**6** and also with mismatch DNA sequences **7**−**9** (in a 1:1 ratio) in 20 mM cacodylate buffer (pH 6.9, 250 mM NaCl, 0.5 mM EDTA) as described above. Samples (2 µM) were excited at 330 nm with excitation and emission slit width 5 nm and 6 nm, respectively. Fluorescence measurements were performed in triplicate.

5.4.11. ^{19}F NMR study of transcript 4 and its duplexes

Transcript **4** and its perfect and mismatch duplexes were annealed in 20 mM cacodylate buffer (pH 6.9, 250 mM NaCl, 0.5 mM EDTA) containing 20% D_2O by heating at 90 °C for 3 min. Samples (5 µM) were allowed to cool slowly to room temperature and incubated at 4 °C for 1 h and transferred to a Shigemi tube (5 mm advance NMR microtube). ^{19}F NMR spectra were recorded at a frequency of 564.9 on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with a Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT). All ^{19}F NMR spectra were calibrated corresponding to an external standard, trifluorotoluene (TFT = −63.72 ppm). ^{19}F NMR of samples were recorded at 25 °C using the following spectral parameters. Excitation pulse: 12 µs; spectral width: 29.90 ppm; transmitter frequency offset: −60.00 ppm; acquisition time: 0.1 s; relaxation delay: 1.0 s; the number of scans: 1500 and an exponential window function using lb = 15.

5.4.12. Fluorescence and ^{19}f NMR-based binding assay

5.4.12.1. Fluorescence study

Single-stranded (ss) RNA transcript **10** and its duplex formed by hybridizing ON **10** with its complementary RNA ON **11** (in 1:1 ratio) in 20 mM HEPES buffer pH 7.5 containing 100 mM NaCl and 0.5 mM EDTA by heating at 75 °C for 5 min. Samples were cooled slowly to room temperature and incubated at 4 °C for 1 h. Samples (1 µM) were excited at 330 nm with excitation and emission slit width 7 nm and 8 nm. The concentration of aminoglycoside in samples (200 µL) of A-site construct (**10**:**11**) and single-stranded ON **10** were increased by adding aliquots (1 µL) of varying concentrations of aminoglycoside and emission spectra were recorded after each addition. In a control experiment, a series of water aliquots (1 µL) were added into a sample of A-site (**10**:**11**) and fluorescence spectra were acquired with excitation and emission slit width 7 nm and 8 nm. The experiment was performed in triplicate. Fluorescence intensity of the blank sample without A-site construct and aminoglycoside was subtracted from all titrations. Normalized fluorescence intensity (F_N) against the concentrations of aminoglycosides (AG) was plotted and a graph was fitted by a Hill equation using Origin Pro 8.5 to obtain the apparent K_d value for the corresponding aminoglycoside. Hill coefficient (n) was found between 1.1–2.0.

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

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

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

5.4.12.2. ^{19}F NMR study. ss RNA ON **10** and A-site construct (**10**:**11**) were annealed in 20 mM HEPES buffer pH 7.5 containing 100 mM NaCl, 0.5 mM EDTA and 20% D_2O by heating at 75 °C for 5 min. Sample (5 µM) were cooled slowly to room temperature and incubated at 4 °C for 1 h. ^{19}F NMR spectra (ns = 1500) of ssRNA ON **10** and A-site constructs (**10**:**11**) was recorded with increasing concentrations of aminoglycosides using the same spectral parameter as mentioned in section 5.4.11. ^{19}F NMR spectra were processed using an exponential window function using lb = 25.

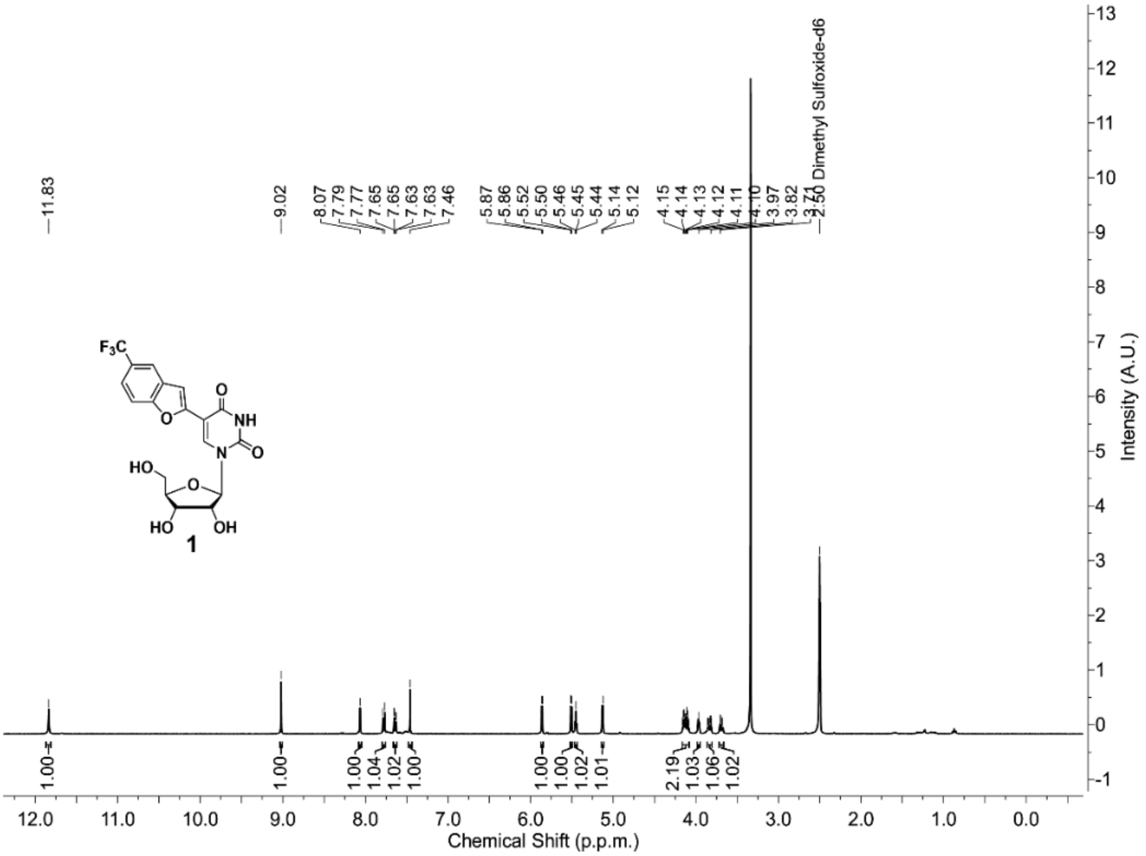
5.5. References

1. Esteller, M. *Nat. Rev. Genet.* **2011**, *12*, 861–874.
2. (a) Staple, D. W.; Butcher, S. E. *PLoS Biology* **2005**, 3 e213−e216. (b) Li, H.; Zhu, D.; Zhang, C.; Han, H.; Crandall, K. A. *Biomed Res Int.* **2014**, *2014*, 690340−690349. (c) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. *Nat. Rev. Mol. Cell Biol*. **2004**, *5*, 1013–1023. (d) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. *Nat. Rev. Mol. Cell Biol*. **2019**, *20*, 474−489.
3. (a) Okamoto, A.; Saito, Y.; Saito, I.; *Photochem. Rev.* **2005**, *6*, 108–122. (b) Holzhauser, C.; Wagenknecht, H.-A. *J. Org. Chem.* **2013**, *78*, 7373–7379. (c) Wetmore, S. D.; Maderville, R. A. *Chem. Sci*. **2016**, *7*, 3482–3493. (d) Manna, S.; Srivatsan, S. G. *RSC Adv.* **2018**, *8*, 25673–25694.
4. (a) Al-Hashimi, H. M.; Walter, N. G. *Curr. Opin. Struct. Biol*. **2008**, *18*, 321–329. (b) Schnieders, R.; Knezic, B.; Zetzsche, H.; Sudakov, A.; Matzel, T.; Richter, C.; Hengesbach, M.; Schwalbe, H.; Fürtig, B. *Curr. Protoc. Nucl. Acid Chem*. **2020**, *82*, e116. (c) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H.; *ChemBioChem* **2003**, *4*, 936−962.
5. (a) Schiemann, O.; Weber, A.; Edwards, T. E.; Prisner, T. F.; Sigurdsson, S. T. *J. Am. Chem. Soc*. **2003**, *125*, 3434–3435. (b) Sicoli, G.; Wachowius, F.; Bennati, M.; Höbartner, C. *Angew. Chem. Int. Ed*. **2010**, *49*, 6443–6447.
6. (a) Mooers, B. H. M. *Methods* **2009**, *47*, 168–176. (b) Peselis, A.; Gao, A.; Serganov, A. *Methods Mol. Biol*. **2016**, *1320*, 21–36.
7. (a) Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. *Chem. Soc. Rev*. **2013**, *42*, 7971–7982. (b) Hammann, C.; Norman, D. G.; Lilley, D. M. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5503–5508. (c) Olejniczak, M.; Gdaniec, Z.; Fischer, A.; Grabarkiewicz, T.; Bielecki, L.; Adamiak, R. W.; *Nucleic Acids Res*. **2002**, *30*, 4241–4249. (d) Olsen, G. L.; Edwards, T. E.; Deka, P.; Varani, G.; Sigurdsson, S. T.; Drobny, G. P. *Nucleic Acids Res*. **2005**, *33*, 3447–3454. (e) Granqvist, L.; Virta, P. *J. Org. Chem.* **2015**, *80*, 7961–7970.
8. Li, Q.; Chen, J.; Trajkovski, M.; Zhou, Y.; Fan, C.; Lu, K.; Tang, P.; Su, X.; Plavec, J.; Xi, Z.; Zhou, C. *J. Am. Chem. Soc.* **2020**, *142*, 4739–4748.
9. Nußbaumer, F.; Plangger, R.; Roeck, M.; Kreutz, C. *Angew. Chem. Int. Ed*. **2020**, *59*, 17062–17069.
10. (a) Baranowski, M. R.; Warminski, M.; Jemielity, J.; Kowalska, J. *Nucleic Acids Res*. **2020**, *48*, 8209–8224. (b) Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. *Angew. Chem. Int. Ed.* **2006**, *45*, 3450–3453. (c) 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. (d) Kiviniemi, A.; Virta, P. *J. Am. Chem. Soc.* **2010**, *132*, 8560–8562. (e) Sakamoto, T.; Hasegawaa, D.; Fujimoto, K. *Analyst* **2016**, *141*, 1214–1217. (f) Kieger, A.; Wiester, M. J.; Procissi, D.; Parrish, T. B.; Mirkin, C. A.; Thaxton, C. S. *Small* **2011**, *14*, 1977–1981. (g) Olszewska, A.; Pohl, R.; Hocek, M. *J. Org. Chem*. **2017**, *82*, 11431–11439.
11. Moazed, D.; Noller, H. F. *Nature* **1987**, *327*, 389–394.
12. Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *Trends Biochem. Sci*. **2003**, *28*, 259–266.
13. Ogle, J. M.; Brodersen, D. E.; Clemons, W. M.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. *Science* **2001**, *292*, 897–902.
14. (a) Srivatsan, S. G.; Tor, Y. *J. Am. Chem. Soc*. **2007**, *129*, 2044−2053. (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. (c) Parsons, J.; Hermann, T. *Tetrahedron* **2007**, *63*, 3548−3552. (d) Xie, Y.; Dix, A. V.; Tor, Y. *J. Am. Chem. Soc*. **2009**, *131*, 17605−17614. (e) Kaul, M.; Barbieri, C. M.; Pilch, D. S. *J. Am. Chem. Soc*. **2004**, *126*, 3447−3453. (f) Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G. *Bioconjugate Chem*. **2013**, *24*, 1367−1377.

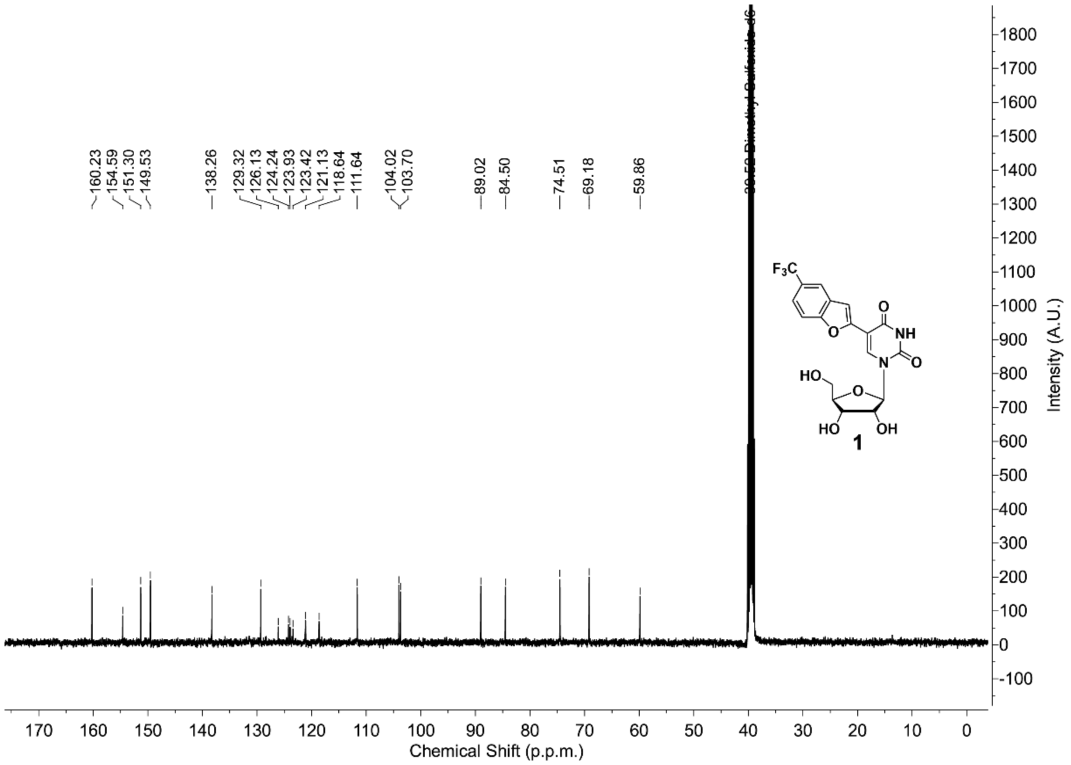
15. (a) Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. *Science* **1996**, *274*, 1367−1371. (b) Vicens, Q.; Westhof, E. *Structure* **2001**, *9*, 647−658. (c) Nuthanakanti, A.; Boerneke, M. A.; Hermann, T.; Srivatsan, S. G. *Angew. Chem. Int. Ed.* **2017**, *56*, 2640–2644.
16. Khatik, S. Y.; Srivatsan, S. G. *Bioconjugate Chem.* **2022**, *33*, 1515−1526.
17. Sinkeldam, R. W.; Wheat, A. J.; Boyaci, H.; Tor, Y. *ChemPhysChem* **2011**, *12*, 567−570.
18. Giam, C. S.; Lyle, J. L. *J. Am. Chem. Soc.* **1973**, *95*, 3235−3239.
19. Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, *180*, 51−62.
20. Seidel, C. A. M.; Schulz, A.; Sauer, M. H. M. *J. Phys. Chem.* **1996**, *100*, 5541−5553.
21. Purohit, P.; Stern, S. *Nature* **1994**, *370*, 659−662.
22. Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem.* **1994**, *5*, 508–512.

5.6 Appendix-II: NMR and mass data of synthesized compounds

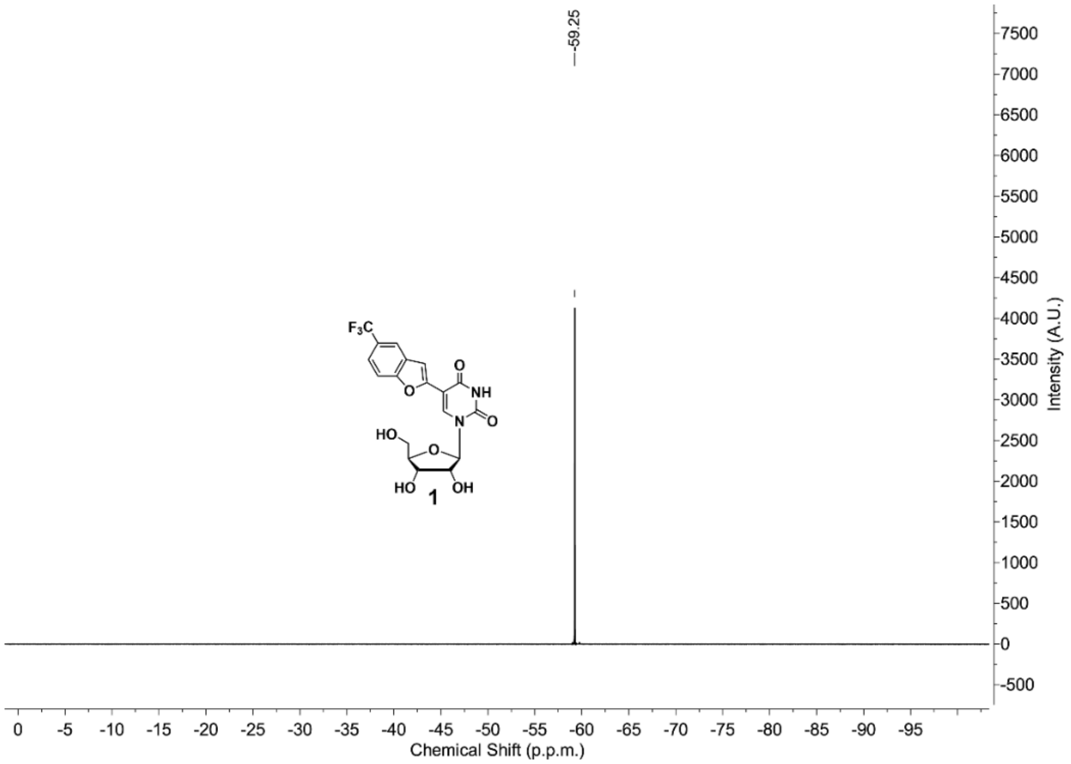
^{1}H NMR of **1** (400 MHz, d_6-DMSO)



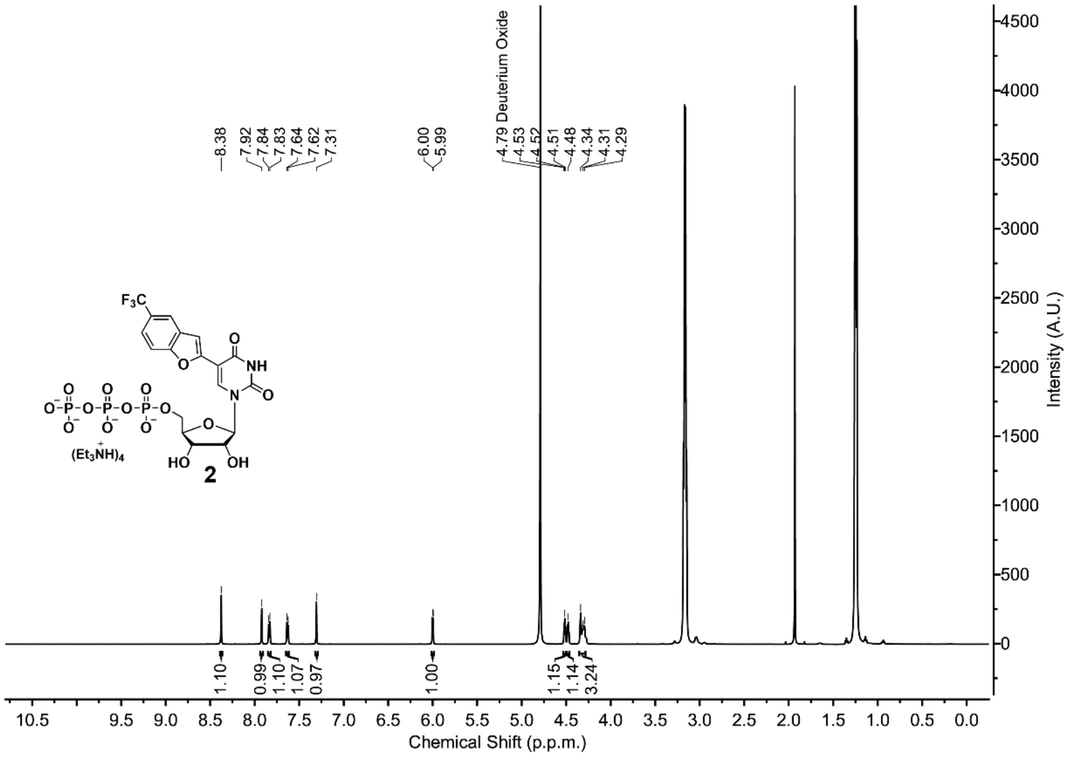
^{13}C NMR of **1** (100 MHz, *d$_6$*-DMSO)



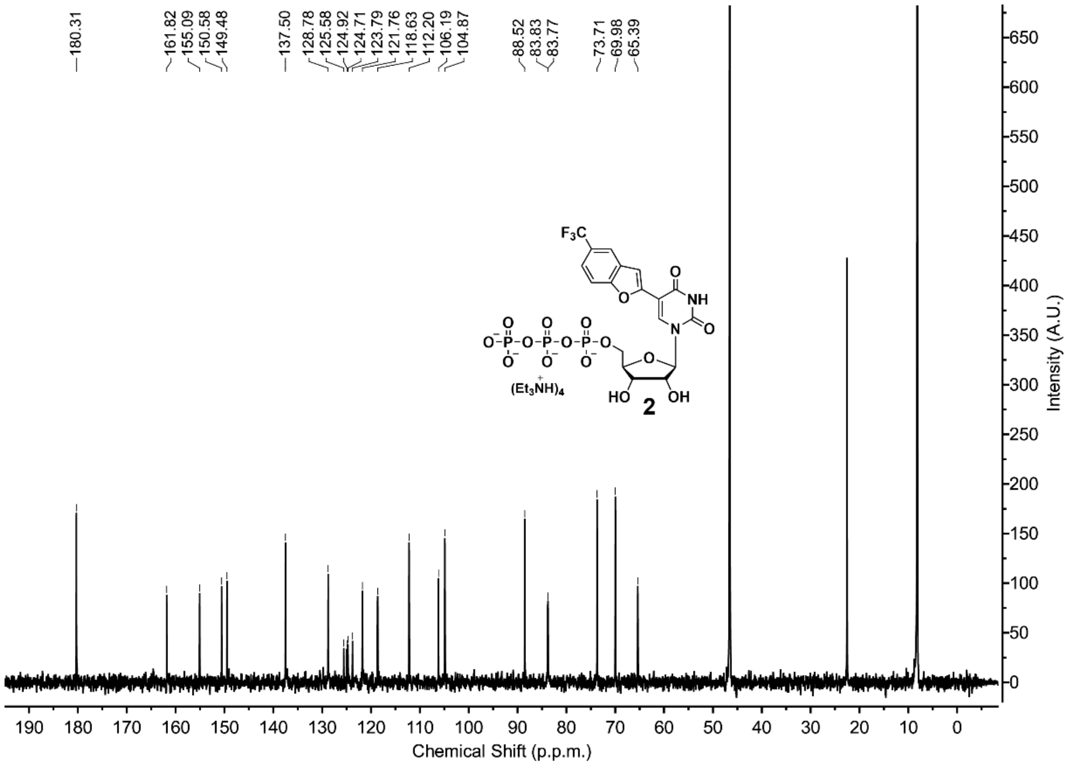
^{19}F NMR of **1** (376.6 MHz, *d$_6$*-DMSO)



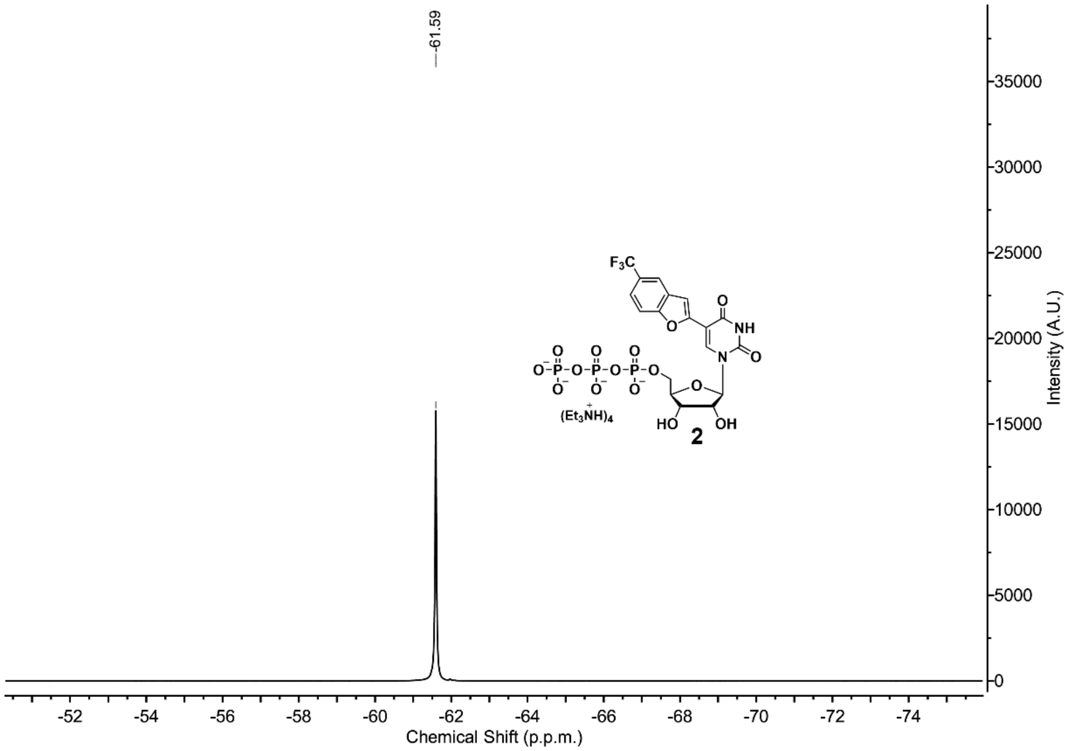
^{1}H NMR of **2** (600 MHz, D_2O) (Trace amount of triethylammonium acetate buffer is present)



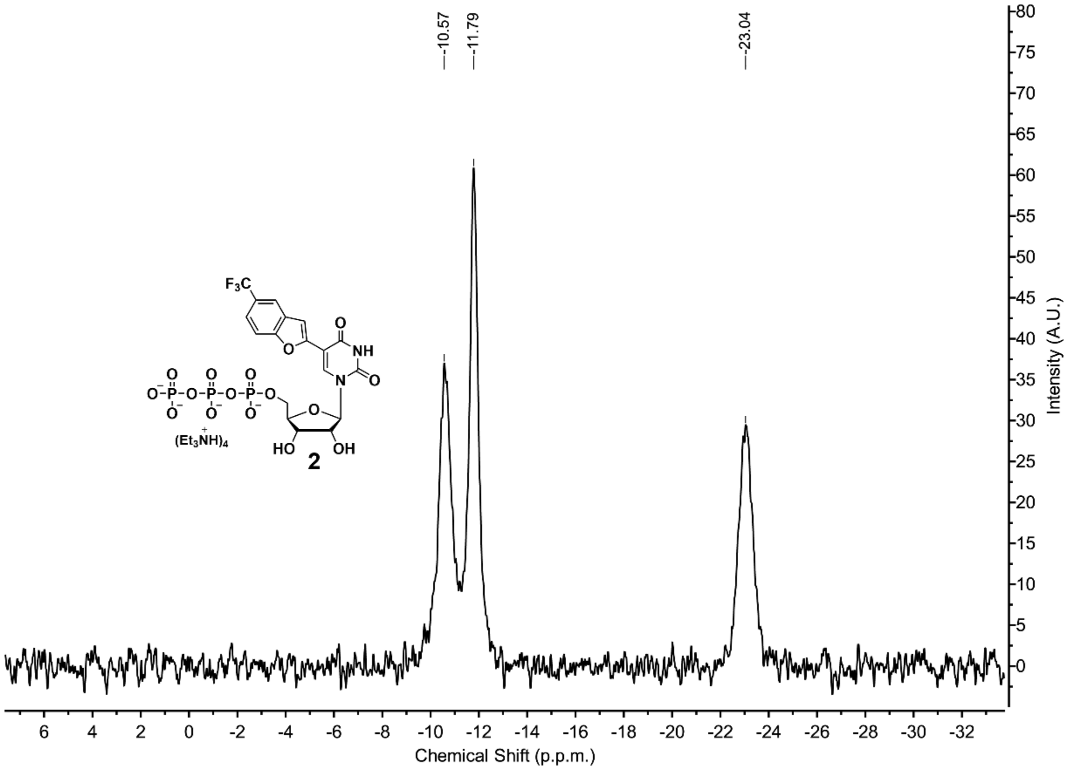
^{13}C NMR of **2** (150 MHz, D_2O) (Trace amount of triethylammonium acetate buffer is present)



^{19}F NMR of **2** (564.9 MHz, D_2O)



^{31}P NMR of **2** (162 MHz, D_2O)



Summary and Future Scope

The core intent of this thesis was to develop a biophysical assay to probe the dynamic equilibrium of the co-existing nucleic acid structures under different conditions. Here, using a simple design approach, we have developed highly sensitive dual-app nucleoside analogs containing fluorescence and ^{19}F NMR tags. Modified analogs efficiently report changes in their surrounding environment in terms of fluorescence and ^{19}F NMR and act as good substrates for the synthesis of DNA/RNA using the solid phase synthesis method and also by enzymatic method. Importantly, modified probes are non-envasive when incorporated into the DNA/RNA and retain their probe-like properties. These properties were useful to investigate the G-quadruplex structural dynamics formed in the promoter region of the EGFR gene in the presence and absence of ligands. Interestingly, the ^{19}F NMR tag provided an opportunity to identify physiologically relevant GQ topology under cellular conditions. Further, modified analog was also applicable to understand the complex iM structural dynamics in the H-Telo and *B-raf* promoter region under different conditions and even able to detect lowly populated iM structures. Similarly, the enzymatic incorporation of a modified ribonucleoside into RNA was able to distinguish changes in its flanking bases and to observe aminoglycoside-antibiotics-induced structural changes in the ribosomal decoding site RNA motif. Collectively, these probes are applicable to study the varieties of nucleic acid structures and even detect minor structures.

While the promoter regions of several oncogenes are prone to adopt multiple GQ/iM structures, it is highly desirable to know the biologically active conformations under cellular conditions. This information will strengthen our understanding of nucleic acid structure-function relationship and further help to develop structure-specific ligands. In this scenario, our dual-app analogs would be highly beneficial to probe the GQ/iM conformational equilibrium formed by the different oncogenes under cellular conditions. Also, the fluorescence and ^{19}F NMR properties of modified nucleosides would be useful to develop a small-molecule screening assay to identify structure-specific ligands.