Probing structure and druggable space of viral G-quadruplexes using dual-functional nucleoside analogs

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

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

द्वारा / By

सरूपा रॉय / Sarupa Roy

पंजीकरण सं. / Registration No.: **20193662**

शोध प्रबंध पर्यवेक्षक / Thesis Supervisor:

प्रो. एस. जी. श्रीवत्सन / Prof. S. G. Srivatsan

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
2024

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 structure and druggable space of viral G-quadruplexes using dual-functional nucleoside analogs*", submitted by Ms. Sarupa Roy, 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

Prof. S. G. Srivatsan

Date: 23/12/2024

Tel.: +91-20-2590 8086 Fax: +91-20-2590 8186 E-mail: srivatsan@iiserpune.ac.in

Declaration by Student

Name of Student: **Ms. Sarupa Roy**

Reg. No.: **20193662**

Thesis Supervisor(s): **Prof. S. G. Srivatsan**

Department: **Chemistry**

Date of joining program: **01/08/2019**

Date of Pre-Synopsis Seminar: **16/08/2024**

Title of Thesis: **Probing structure and druggable space of viral G-quadruplexes using dual-functional nucleoside analogs**

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I declare that I have acknowledged collaborative work and discussions wherever such work has been included. 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.

The work reported in this thesis is the original work done by me under the guidance of Prof. S. G. Srivatsan.

Date: 23/12/2024

Signature of the student

For the seekers, the dreamers, and the questioners–

This one's for you….

Acknowledgements

I extend my deepest gratitude to my supervisor, Prof. S. G. Srivatsan, for his invaluable guidance, unwavering support, and insightful mentorship throughout my doctoral journey. His passion for scientific discovery, meticulous attention to detail, and commitment to fostering academic excellence have been a constant source of inspiration. I am immensely grateful for the freedom he granted me to explore new ideas and the trust he placed in my abilities, which allowed me to grow as an independent researcher. His encouragement during challenging times and his constructive feedback have not only honed my scientific acumen but also strengthened my perseverance and critical thinking skills.

Prof. Srivatsan's ability to see the silver lining in every situation has been especially motivating, instilling in me a sense of resilience and optimism even during setbacks. His generosity and sense of humor, in both significant and everyday moments, have made this journey even more memorable. I am particularly thankful for his patience in refining scientific figures and demonstrating how to get things just right—skills that have made all the difference.

Thank you, Sir, for being a mentor, guide, and role model in every sense of the word.

I would like to express my sincere gratitude to the members of my Research Advisory Committee (RAC), Prof. H. N. Gopi and Dr. Moneesha Fernandes for their invaluable guidance and insightful feedback throughout my doctoral research journey. Their support and suggestions have been instrumental in shaping my work and enhancing its quality.

I would like to express my gratitude to Prof. Pradeepkumar P. I., Priyasha, and Shruti from IIT Bombay for their invaluable contribution to carry out the MD simulation part of this work. I would extend my thanks to Prof. Jeet Kalia and Satyajeet for helping in the surgery of frog and extraction of eggs for the NMR experiment. Their expertise and efforts were immensely helpful and greatly appreciated. A special thanks to Dr. Bharathwaj and Vinay from IISER Bhopal for their efforts in exploring the NMR aspect of RNA. Your contributions are sincerely appreciated.

I would like to express my heartfelt gratitude to my labmates, Dr. Jerrin, Dr. Manisha, Dr. Manisha Jagtap, Dr. Ragini, Dr. Saddam, Dr. Pulak, Dr. Apeksha, Swagata, Arindam, Amit, Harsh, Vivek, Punnya, Kartik for their camaraderie, collaboration, and unwavering support throughout my Ph.D. journey. Their insightful discussions, constructive feedback, and constant encouragement have been invaluable to my research. Beyond the science, I will always treasure the countless hours spent brainstorming not just experiments but also party venues! From planning celebrations to navigating deadlines, your humor and positivity have been a constant source of joy and motivation. The lively conversations, shared laughter, and memorable moments have made my time in the lab an experience I will always cherish. Working alongside such talented, enthusiastic, and fun-loving colleagues has been a privilege. A one-liner for the present lab members which I will cherish forever -

Swagata- Our all-time playful banter and scientific discussions were some of the highlights of this journey. These moments of lighthearted teasing, combined with deep scientific exchanges, made every challenge a little easier and every success even sweeter.

Arindam: All-time partner in mischief and childish antics, always taking each other's stuff and making fun of each other. These moments of playful fun, alongside our scientific discussions, made this journey memorable and full of laughter

Apeksha: Our Goa girl! Thanks for suggesting the best libations!

Harsh: Serious discussions about lab safety and other important stuff—always a mix of learning and laughter, but we made sure we stayed on top of the essentials!

Vivek and Punnya: I loved teaching you guys! Thanks for making this journey so memorable and for always being there to lend a helping hand when needed. You guys really made it all easier!

Thank you all for making the lab a place of learning, growth, and friendship and for being an integral part of this journey.

A special thanks goes to Pulak, who has been an unwavering source of support throughout my journey. Whether it was for deep lab discussions, navigating the ups and downs of this research journey, or simply offering a moment of lightheartedness, Pulak was always there. Your constant presence, thoughtful insights, and ability to bring joy and laughter into even the busiest of days made a world of difference. Thank you for being there every step of the way, not just as a colleague, but as a true friend and surrendering me the very thing you received in the company roadshow. Your contributions have made this journey all the more memorable.

And, of course, a big shoutout to Ragini Di, who was always there, cheering me on during my presentations and offering encouragement just when I needed it the most. From being a steady presence in moments of doubt to sharing laughter and advice in between, your unwavering support has meant the world to me. Whether it was patiently listening to my practice runs or simply believing in me when I wasn't sure myself, you've been a constant source of strength throughout this journey. Thank you for everything, Di—you truly made this experience so much better!

I would also like to extend my deepest gratitude to my mother, whose unwavering love, support, and patience have been a constant source of strength throughout this journey. Not only has she been my rock, but she also somehow managed to partially earn a Ph.D. alongside me—especially after learning just how difficult RNA is to handle compared to DNA! I'm pretty sure she now knows more about RNA's temperamental nature than most people in this field. Thank you, Mom, for your endless encouragement, your belief in me, and for being my biggest cheerleader through every experiment, both successful and (often) disastrous. Your support made all the difference, and I couldn't have done this without you!

I would like to express my deepest gratitude to my father, whose silent support has been a constant throughout this entire journey. Though he may not always have voiced his encouragement, I have always felt his unwavering presence and belief in me. His calm and steady demeanor, along with his quiet confidence in my abilities, has been a source of strength during moments of doubt. Sometimes, the most powerful support is given in silence, and for that, I am forever grateful. Thank you, Dad, for being there in every way that truly matters.

A big thank you to my brother *and Bouthan for always having my back. Your constant support, love, and encouragement have meant the world to me and kept me going through this journey.*

*I would also like to extend my heartfelt gratitude to my IISER family, particularly Anindita **a**nd Writakshi. Thank you both for being more than just friends during this journey—you were my constant support system. Anindita, I will forever remember how you scolded me for eating while cooking and not having proper meals! Your concern for my well-being, even when I was too absorbed in my work, was a much-needed reminder to take care of myself. Writakshi, your patience, understanding, and always being there with a listening ear or a word of advice made all the difference. I'm lucky to have had you both by my side, keeping me grounded, well-fed, and supported through the highs and lows of this Ph.D. experience.*

I extend my heartfelt gratitude to everyone who has knowingly or unknowingly contributed to this journey as well as to my friends and family.

I would like to express my gratitude to the universe for guiding me through this journey, for aligning the right opportunities, people, and moments, and for giving me the strength to persevere through challenges.

"**The stars began to burn through the sheets of clouds,**

and there was a new voice which you slowly

recognized as your own,

that kept you company as you strode deeper and deeper

into the world, determined to do the only thing you could do—

determined to save the only life you could save."

- *Mary Oliver*

Table of Contents

Abbreviations

λ	Wavelength
μM	Micromolar
ε	molar extinction coefficient
μL	microliter
μg	microgram
A	adenine
2-AP	2-Aminopurine
Abs	Absorbance
ACN	Acetonitrile
Ag	Silver
C	cytosine
CD	circular dichroism
COSY	Correlated Spectroscopy
Cy	Cysteine
3D	Three Dimensional
DEER	Double Resonance
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	N, N-dimethyl sulfoxide
DMT	dimethoxytrityl
DNA	deoxyribonucleic acid
DOX	doxorubicine
ds	double stranded
DTT	dithiothreitol
dT	2′-deoxy thymidine
EDTA	Ethylenediaminetetraacetic Acid
ESI-MS	Electrospray Mass Spectroscopy
em	Emission
EPR	electron paramagnetic resonance
FAM	Fluorescein Amidites
FI	Fluorescence Intensity
G	Guanosine
GQ	G-quadruplex
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
H-Telo	Human Telomeric
in vitro	Outside living organism
in vivo	Inside living organism
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation-Time of flight
max	maximum
MeOH	methanol
mg	milligram
MHz	megahertz
mM	millimolar
ng	nanogram
nm	nanometer
nmol	nanomoles
nM	nanomolar
NMR	nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NTP	Nucleotide triphosphate
PDB	Protein Data Bank
PELDOR	Pulsed Electron Double Resonance
ON	oligonucleotides
Pd	Palladium
PAGE	polyacrylamide gel electrophoresis
ppm	Parts per million
PDS	pyridostatin
RC	Random coil
RNA	Ribonucleic acid
R_f	Retention factor
ss	Single strand
T	thymine
TBAF	Tetrabutylammonium Fluoride
TBDMS	Tert-Butyldimethylsilyl
TCA	Trichloroacetic acid
tC	1,3-diaza-2-oxophenothiazine-2′-cytidine
tC_{nitro}	7-nitro-1,3-diaza-2-oxophenothiazine-2′-oxophenoxazine-2′-deoxycytidine
TDS	Thermal differential spectra
TEA.3HF	Triethylamine hydrofluoride
TEAA	Triethylammonium Acetate
TERRA	Telomeric Repeat-Containing RNA
TFT	Trifluorotoluene
THF	tetrahydrofuran

TLC	Thin layer chromatography
T_m	Thermal melting
TMEDA	*N*, *N*, *N'*, *N'* Tetramethylethylenediamine
TOCSY	Total Correlation Spectroscopy
TO	Thiazole orange
ThT	Thioflavin T
Tris	Tris(hydroxymethyl)amino methane
U	Uridine/Uracil
UV	Ultraviolet
WC	Watson-Crick

Synopsis

Probing structure and druggable space of viral G-quadruplexes using dual-functional nucleoside analogs

Background and Aim

Nucleic acids exhibit structural diversity beyond their well-known double-helical form, adopting alternative structures including bulges, hairpins, branched junctions, triplexes and G-quadruplexes (GQs).[1] These non-canonical structures are closely tied to their functional roles in biological systems. Investigating the interplay between nucleic acid structures and their biological functions is crucial in structural biology and holds immense promise for designing nucleic acid-based diagnostic and therapeutic strategies. To this end, numerous studies have utilized techniques like fluorescence spectroscopy, CD, NMR, EPR, and X-ray crystallography to explore nucleic acid structures, dynamics, and their interaction with proteins.[2] The majority of these methods typically require external labeling of nucleic acids, as they lack inherent detection tags or labels. Besides structural characterization, these methodologies aid in identifying small molecules that specifically bind to nucleic acid structures, paving the way for novel therapeutic approaches. Nevertheless, most of these studies are conducted under *in vitro* conditions, which may not accurately replicate the native intracellular environment.[3] Factors such as ionic strength, molecular crowding, confinement, and pH significantly influence nucleic acid folding, meaning their cellular conformations often differ from those observed *in vitro*.[4] Emerging tools, including antibodies, fluorescent probes, and in-cell applications of NMR and EPR have enhanced the study of nucleic acid conformations within cellular contexts.[5] However, distinguishing closely related conformations, such as different GQ topologies, in dynamic equilibrium within complex cellular environments remains a persistent challenge. To overcome these limitations, we proposed developing minimally invasive nucleoside probes capable of detecting environmental changes and producing unique signatures for diverse nucleic acid topologies, both *in vitro* and in cellular systems.

Scope of the Thesis

This thesis focuses on utilizing dual-labeled nucleoside analogs equipped with fluorescence and ^{19}F NMR tags to study nucleic acid conformations, dynamics, and interactions with small molecules. Specifically, modified nucleosides were synthesized by conjugating a fluorobenzofuran group to the C5-position of 2′-deoxyuridine (FBFdU) or uridine (FBFU).

These probes demonstrated high sensitivity to environmental factors such as solvent polarity and viscosity while maintaining the structural integrity of DNA and RNA when incorporated into oligonucleotides. This versatile platform enables the investigation of nucleic acid motifs in both cell-free and cellular environments, with a particular emphasis on secondary structures of conserved G-rich viral sequences, including G-quadruplexes. By leveraging the sensitivity and specificity of these dual-labeled probes, the thesis delves into characterizing the formation, stability, and interactions of viral G-quadruplex structures, which are crucial regulators of viral genome functions such as replication, transcription, and translation. This approach provides deeper insights into the structural dynamics of these critical elements in the context of viral pathogenesis.

In this context, the dual-functional nucleoside, in combination with 5-fluoro-2′-deoxyuridine, generated distinct fluorescence and ^{19}F NMR signatures for the individual GQ structures formed by the G-rich DNA sequence of the HIV-1 long terminal repeat (LTR). Importantly, ^{19}F NMR signals facilitated the detection of a physiologically relevant HIV-1 GQ structure under cellular conditions. Furthermore, this probe system was employed to monitor the formation and polymorphism of a unique hybrid hairpin/GQ structure, as well as to map small-molecule ligand interactions with the hybrid hairpin-GQ system. Similarly, the modified ribonucleoside was utilized to investigate GQ polymorphism within the Nsp3 gene of SARS-CoV-2. Collectively, these dual-labeled nucleoside analogs provide an innovative and efficient approach for exploring complex nucleic acid structural equilibria and offer significant potential for identifying topology-specific binders aimed at targeting viral sequences.

The organization of the thesis is outlined as follows.

Chapter 1. Biophysical approaches for investigating structure, dynamics, and function of nucleic acids

This introductory chapter discusses the formation of non-canonical nucleic acid structures, with a particular focus on GQ structures, and their biological functions. The structure-function relationship of GQs has been explored using a variety of biophysical techniques, including CD, NMR, fluorescence, EPR and X-ray crystallography.[2] The principles behind these techniques, recent advancements, and challenges in studying GQ structures and their interactions with proteins and ligands are briefly outlined. Additionally, the chapter covers the use of antibodies, fluorescence light-up probes, in-cell NMR, and EPR techniques to investigate GQ structures under cellular conditions. Furthermore, the development and application of micro-environment-sensitive fluorescent nucleoside analogs for probing GQ structures both *in vitro*

and *in vivo* are reviewed in detail. The limitations of these methods in studying nucleic acid structures are discussed, leading to the motivation of the present thesis on the development and utilization of dual-app nucleoside analogs, which are elaborated in the final section of the chapter.

Chapter 2. Structural elucidation of HIV-1 G-quadruplexes in a cellular environment using responsive ^{19}F-labeled nucleoside probes

Understanding the structure of highly conserved regulatory segments in integrated viral DNA genomes, which adopt unique and complex topologies, is crucial for devising novel therapeutic strategies against chronic infections. A notable example is the long terminal repeat (LTR) G-rich region of HIV-1, which exemplifies such conservation and structural complexity.[6] In this study, we developed and employed a highly sensitive probe system based on nucleoside analogs, 5-fluoro-2′-deoxyuridine (FdU) and 5-fluorobenzofuran-2′-deoxyuridine (FBFdU), to explore the structural polymorphism of GQs formed by the G-rich sequences in the HIV-1 long terminal repeat (LTR). Using fluorescence and ^{19}F NMR techniques, these analogs served as effective sensors for detecting distinct topologies of GQs and their interactions with other structural motifs. FdU and FBFdU demonstrated distinct spectral signatures that enabled identification and characterization of different GQ topologies within the LTR G-rich sequences. Notably, systematic ^{19}F NMR analysis in *Xenopus laevis* oocytes provided unprecedented insights into the structural conformation of the LTR G-rich regions within a cellular environment. The data revealed that this region adopts a unique hybrid GQ-hairpin architecture, which represents a highly specific and accessible hotspot for selective targeting by small molecules.

In addition to experimental data, structural models generated using molecular dynamics (MD) simulations offered detailed insights into how the probe system senses various GQ topologies. These simulations highlighted the probe's sensitivity in distinguishing between structural variations, further validating its utility in mapping complex nucleic acid architectures. This study offers insights into the structural landscape of HIV-1 LTR GQs, which could serve as a foundation for designing targeted and structure-specific therapeutic agents.

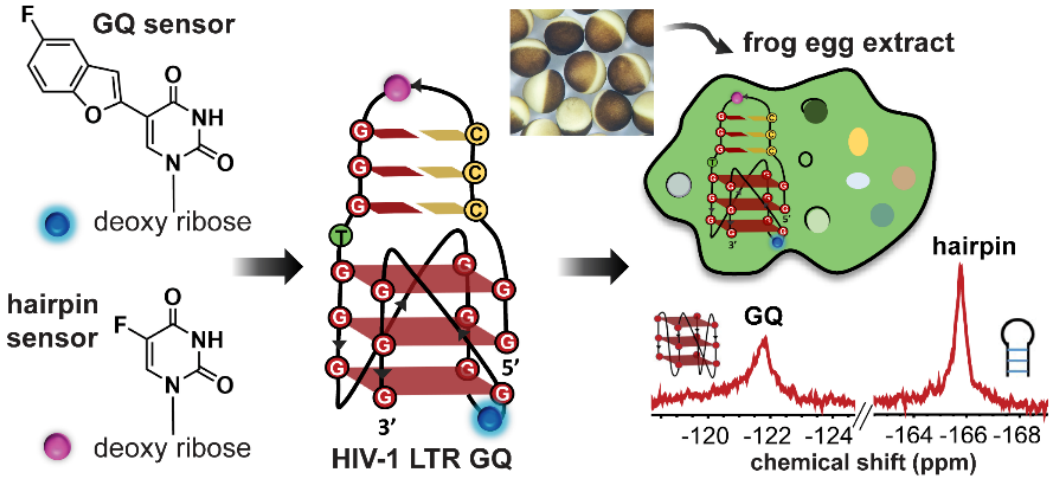


Figure 1: Microenvironment-sensitive probe system detecting HIV-1 LTR G-rich region in cell-free and cellular environment.

Chapter 3. Probing druggable space of HIV-1 LTR G-quadruplexes using responsive ^{19}F-labeled nucleoside probes

Given the highly conserved nature and unique structural features of these GQs, we sought to investigate how small molecules interact with the juxtaposed hairpin-GQ regions. To achieve this, we employed an environment-sensitive probe system comprising FdU and FBFdU, designed to map the druggable space of individual LTR segments using the spectral signatures of fluorescence and ^{19}F NMR. The fluorescence component of the dual-labeled probe was pivotal in estimating the binding affinities of various ligands with distinct spectral properties. Interestingly, ^{19}F NMR analysis provided a deeper understanding by revealing distinct signals corresponding to the specific regions of the hybrid GQ-hairpin structure targeted by the ligands. This finding underscores the ability of small molecules to selectively bind structure-specific regions within the hybrid motif.

To extend this work, we examined a longer LTR region encompassing LTR-(III+IV), which was stabilized using TMPyP4, a GQ binder, and doxorubicin, a duplex binder. Notably, ^{19}F NMR revealed that TMPyP4 specifically interacted with the G-tetrad regions, while doxorubicin preferentially occupied the hairpin region of the hybrid structure. These observations highlighted the differential binding preferences of ligands, shedding light on their structure-specific targeting capabilities.

Building on these results, we conducted a *Taq* DNA polymerase stop assay to assess the potential of simultaneous targeting of the GQ-hairpin motif. The assay suggested that dual targeting of these juxtaposed motifs could be a rational strategy to selectively and effectively inhibit viral pathogenesis. In this context, the combination of FBFdU and FdU offers a versatile

platform not only for studying the hybrid GQ motifs but also for devising screening assays to identify hybrid ligands capable of targeting GQ-hairpin motifs. This approach paves the way for designing novel therapeutic interventions against viruses leveraging the unique structural features of their nucleic acid elements.

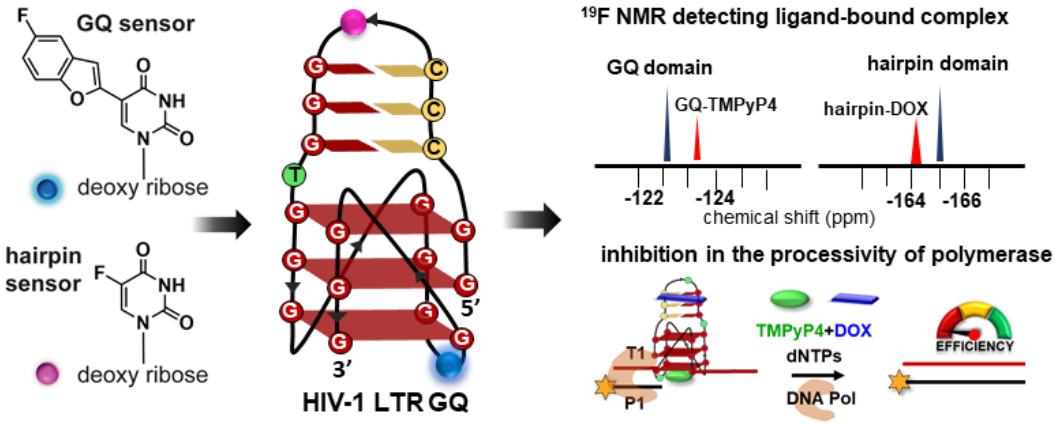


Figure 2. Microenvironment sensitive probe system mapping the druggable space of the unique hairpin juxtaposed GQ structure.

Chapter 4. Probing viral G-quadruplexes of SARS-CoV-2 with ^{19}F dual-app nucleoside analog

The fluorobenzofuran-modified dual-labeled RNA analog (FBFU) was employed to investigate RNA viral sequences, broadening its utility in studying viral GQs. Given the frequent emergence of SARS-CoV-2 variants,[7] which pose challenges to the efficacy of existing vaccines and antiviral therapies, targeting viral GQs represents an innovative and promising therapeutic strategy.[8] In this work, we focused on the structural and functional characterization of the highly conserved G-rich region encoding the non-structural protein Nsp3 of SARS-CoV-2, a sequence implicated in critical viral processes. The dual-labeled probe was strategically incorporated into the Nsp3 G-rich region by substituting a natural uridine residue. This incorporation enabled the detection of structural polymorphisms within the GQ, unveiling the coexistence of multiple GQ conformations. Detailed temperature-dependent studies revealed further insights into the polymorphic nature of these GQs, highlighting their dynamic structural landscape. Furthermore, the probe was instrumental in assessing ligand interactions with Nsp3 GQs. Binding studies demonstrated a high affinity of the well-known GQ-binding ligands TMPyP4 and BRACO19, emphasizing the therapeutic potential of targeting these structures. These findings underscore the versatility and sensitivity of the probe in elucidating the complex structural distribution and ligand recognition properties of viral

GQs. This understanding of GQ structural diversity and ligand interactions provides a valuable foundation for the rational design of antiviral drugs. By tailoring molecules to specifically target the unique and conserved structural element, this approach opens a new avenue in advancing therapeutic strategies against RNA viruses like SARS-CoV-2.

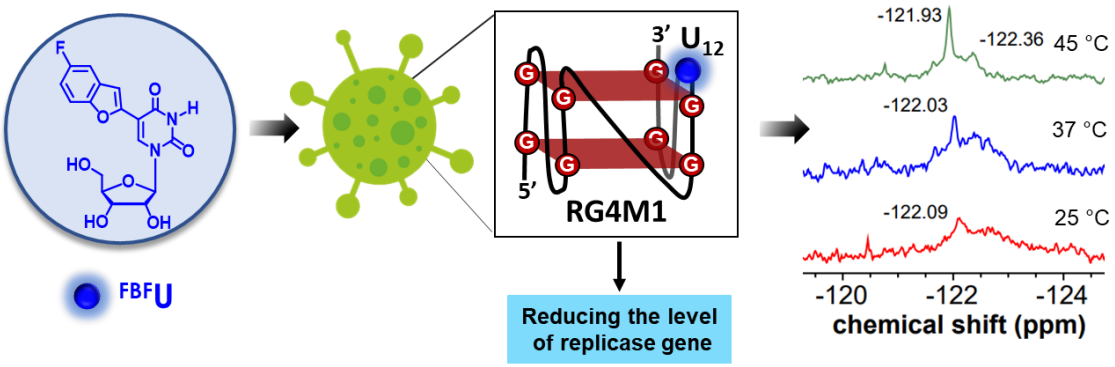


Figure 3. Microenvironment-sensitive RNA probe system detecting Nsp3 region in GQs of SARS-CoV-2.

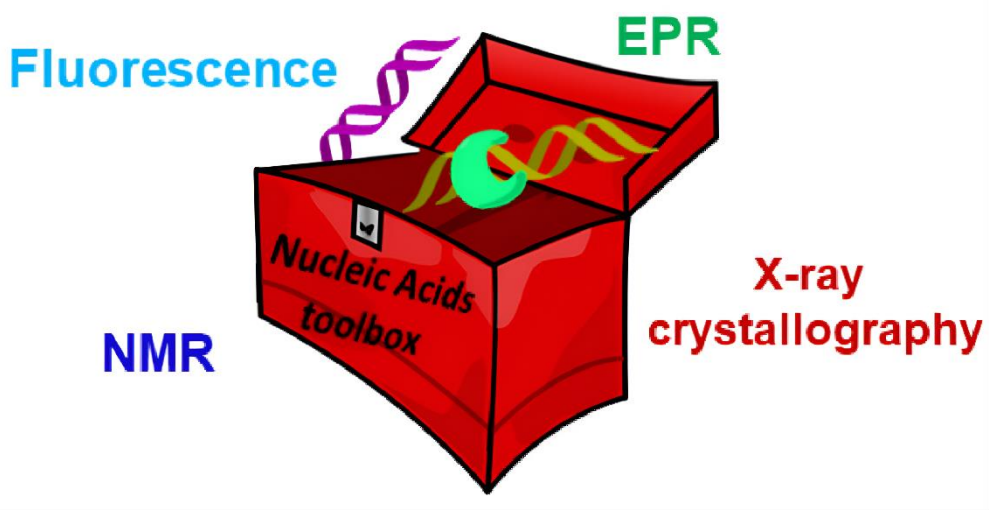
References

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) Leamy, K. A.; Assmann, S. M.; Mathews, D. H.; Bevilacqua, P. C. *Q Rev Biophys.* **2016**, *49*, e10. (b) Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. *FEBS Letters* **2002**, *526*, 77–81. (c) Nakano, S.; Miyoshi, D.; Sugimoto, N. *Chem. Rev.* **2014**, *114*, 2733. (d) Assi, H. A.; Garavís, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2018**, *46*, 8038–8056.
5. (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.
6. (a) Butovskaya, E.; Heddi, B.; Bakalar, B.; Richter, S. N.; Phan, A. T. *J. Am. Chem. Soc.*, **2018**, *140*, 13654–13662. (b) 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.
7. Harvey, W. T.; Carabelli, A. M.; Jackson, B. *et al*. *Nat Rev Microbiol*, **2021**, *19*, 409–424
8. Pathak R. *Viruses* **2023**, *15*, 2216.

Chapter 1

Biophysical approaches for investigating structure, dynamics and function of nucleic acids



1.1 Introduction

1.1.1 Nucleic Acids: A hall of fame for structural and functional versatility

Nucleic acids are indispensable molecules in life, performing dual roles as carriers of genetic information and regulators of vital cellular processes.[1] Beyond their traditional role as blueprints for protein synthesis, their versatility is rooted in their ability to adopt diverse secondary and tertiary structures.[2] These include hairpins, triplexes, cruciforms, G-quadruplexes, i-motifs, bulges, loops, Holliday junctions, pseudoknots, and etc.[2,3] Such dynamic conformations are critical for regulating fundamental cellular activities, including replication, transcription, repair, and recombination.[4]

This structural diversity is particularly evident in RNA, which exhibits extraordinary complexity.[5] Non-coding RNAs like ribozymes and tRNAs fold into intricate three-dimensional shapes, enabling catalytic activities and precise molecular recognition.[6] Such specialized structures allow nucleic acids to interact with ions, water molecules, and proteins, such as histones and DNA polymerases, thereby acting as versatile scaffolds, regulators, and effectors in a variety of cellular mechanisms.[7] In addition to their structural versatility, nucleic acid functions are finely modulated by environmental factors, including pH, temperature, and chemical modifications.[8] Modifications such as DNA methylation and diverse RNA base alterations introduce additional regulatory layers, allowing nucleic acids to adapt dynamically to cellular and environmental cues. These modifications, together with interactions between nucleic acids and proteins, underpin essential biological processes, including DNA repair, recombination, transcription regulation, and viral infection.[9]

Nucleic acid structures have been studied using a variety of biophysical techniques, including circular dichroism (CD), UV-Vis spectroscopy, fluorescence, X-ray crystallography, electron paramagnetic resonance (EPR) and nuclear magnetic resonance (NMR).[10] However, the lack of intrinsic labels in native nucleic acids suitable for techniques like fluorescence, NMR, EPR, and X-ray methods poses a significant challenge for structural and interaction studies. To overcome this challenge, responsive nucleoside labels have been developed, enabling investigations of nucleic acid architectures, their dynamics, and their interactions with proteins and ligands.[10b-e] These methods offer valuable insights into the structural properties and functions of nucleic acids under *in vitro* conditions. However, studying these features within living cells remains challenging due to the complexity of the cellular environment.[11] Since nucleic acid structures can be highly sensitive to environmental changes, it remains uncertain whether structures observed *in vitro* behave the same way within cells. Recently, the

development of antibodies and fluorescent probes has significantly advanced our understanding of where and how certain nucleic acid structures form within cells.[12–14] In addition, in-cell NMR and EPR techniques have proven useful for studying nucleic acid structure and interactions *in vivo*, though they still face certain limitations in cellular applications.[14d]

This chapter provides a comprehensive overview of non-canonical tetraplex nucleic acid structures, with a specific focus on G-quadruplexes (GQs). GQs have garnered significant attention due to their widespread presence in genome regulatory regions, including telomeres, oncogene promoters, and untranslated regions of mRNA.[15] Their ability to regulate key cellular processes such as transcription, translation, replication, and genomic stability underlines their biological importance.[4] To explore these functions, the chapter delves into the biophysical techniques used to study GQ structures, emphasizing their strengths and limitations. Traditional methods such as CD spectroscopy and UV thermal melting are discussed for their utility in determining GQ formation and stability, while advanced approaches like NMR and X-ray crystallography are highlighted for detailed structural and dynamic insights.

Particular emphasis is given to the use of base-modified nucleosides and site-specific labels, which have revolutionized the study of GQs. These modified probes, such as fluorescent or ^{19}F NMR-labeled analogs, facilitate monitoring of GQ stability, folding polymorphism, and interactions with biomolecules like proteins or ligands. This section also reviews the application of these probes in uncovering subtle conformational changes and their functional implications in biological contexts. In addition to analytical techniques, the chapter explores advancements in tools for visualizing GQs within cellular environments, highlighting structure-specific antibodies, small-molecule ligands, and chemical probes as powerful methods for detecting GQs *in situ*. These tools have enabled researchers to map GQ localization, identify their functional roles in live cells, and evaluate their involvement in disease pathways. Finally, the chapter addresses critical challenges in the field, including the transient nature of GQs, their structural heterogeneity, and the difficulty in distinguishing different biologically relevant GQs. Together, these challenges and developments lay the groundwork for the present research aimed at understanding GQs and exploring their potential as therapeutic targets.

1.2 G-quadruplexes

G-tetrads are formed when guanine bases are arranged into planar quartets, with each guanine connected to two others through Hoogsteen base pairing (Figure 1A). Each pair of guanines forms hydrogen bonds involving four donor and acceptor atoms: N1, N7, N2, and O6, resulting in a G-quartet with a total of eight hydrogen bonds (four N2–H---N7 and four N1–H---O6). The four carbonyl oxygen atoms (O6) create a negatively charged core at the center of the G-quartet (Figure 1A).[16,17] When two or more G-quartets stack on top of each other, they form a stable right-handed helical structure called a G-quadruplex (GQ), with individual G-quartets separated by approximately 3.3 Å.[16] Monovalent cations or other metal ions can intercalate into the central anionic core of a G-tetrad or bind between adjacent G-quartets, facilitating coordination and stabilization of hydrogen-bonded tetrads and enhancing base-stacking interactions.[17,18] The effectiveness of monovalent cations in stabilizing GQs follows the order: of K^+>$NH4^+$>Na^+>Li^+.[19] This variation is closely tied to the cation's ionic radius and level of dehydration, which influence how well the ions fit within the GQ structure.[19b] K^+ ions are particularly effective stabilizers, typically residing between two adjacent G-quartets, where their optimal size and hydration properties enable strong coordination with the oxygen atoms of guanine residues. In contrast, Na^+, ions being smaller and more dehydrated, tend to localize at individual tetrad planes rather than between them. This cation-specific behavior plays a critical role in determining the folding, topology, and stability of GQs.[19c]

1.2.1 Structural polymorphism of GQs

GQs exhibit remarkable structural diversity due to factors like strand orientation, loop configurations, cations and the glycosidic bond angles of guanine bases. These variations allow GQs to adopt distinct conformations that can serve specific biological functions.[20] GQ structures can form from a single DNA strand (intramolecular) or multiple strands (intermolecular), resulting in different arrangements: parallel, antiparallel, or hybrid (3+1) (Figure 1B).[21] The orientation of these strands influences the types of loops connecting the G-tetrad layers, which are essential to GQ functionality.[22] Loop structure within GQs adds another layer of complexity. Depending on their positioning, loops can bridge adjacent guanines to form lateral loops, connect opposing guanines in diagonal loops, or link top and bottom stacks in double-chain reversal loops.[23] Parallel GQs, where all strands align in the same direction, typically feature double-chain reversal loops.[24a] In antiparallel arrangements, where strands run in opposite directions, lateral or diagonal loops are more common.[24b] The hybrid (3+1)

arrangement combines three strands in one direction with a single opposing strand, producing a mix of loop types (Figure 1C).[24c]

GQs are also characterized by four grooves, each separated by the sugar-phosphate backbone of the four guanines. When the guanines are aligned in parallel and share the same glycosidic orientation within a G-tetrad, these grooves are uniformly medium in width.[25] In contrast, an antiparallel arrangement of guanines with mixed *syn* and *anti*-glycosidic conformations creates grooves of varying sizes, including both narrow and wide grooves.

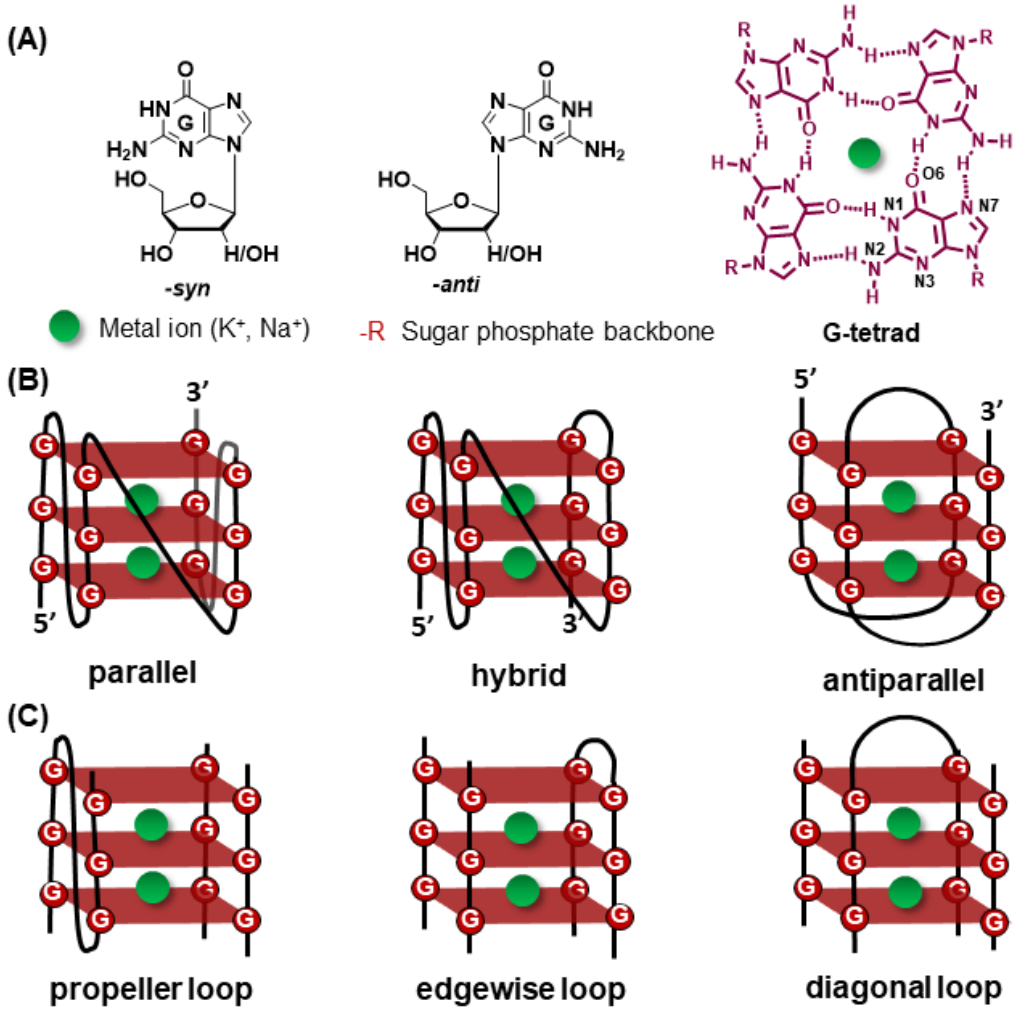


Figure 1. (**A**) Structures of basic units of GQ. *syn*- and *anti*- glycosidic conformations of guanosine and planar G-tetrad. (**B**) common folding topologies of the GQ structure. (**C**) different loops connecting the GQ tetrads.

1.2.2 Biological function of GQs

1.2.2.1 DNA GQ

Replication: During DNA replication, single-stranded regions often form on the lagging strand, which is particularly prone to GQ formation, especially under conditions of replication

stress or slowing.[26] These GQ structures, whether pre-existing or newly formed during replication, must be unwound by specific helicases such as BLM, WRN, and FANCJ.[27] These helicases are crucial because folded GQ structures cannot serve as templates for DNA replication. The unwinding and resolution of GQs are crucial for ensuring uninterrupted replication and preventing errors such as deletions and strand breaks, thereby safeguarding genomic stability and integrity. Additionally, by influencing replication timing and fork progression, GQs may indirectly regulate the expression of nearby genes (Figure 2).[28] Thus, the careful management of GQ structures during replication not only ensures genomic stability but also plays a potential role in regulating gene expression through its impact on replication dynamics.

Transcription: Approximately 50% of human genes contain predicted GQ sequences near their promoters, with a significant enrichment in oncogenes and regulatory genes compared to housekeeping or tumor suppressor genes. GQs play a dual role in cancer biology, as highlighted by Carvalho et al., who described them as a "double-edged sword": they can serve as "friends" when targeted as therapeutic agents but act as "foes" when their misregulation promotes cancer progression.[29a] The *MYC* gene, a proto-oncogene involved in cell proliferation and growth, exemplifies this duality. Its promoter region contains sequences capable of forming GQ structures that regulate its transcription. Stabilization of these GQs by small molecules, such as BRACO-19 and pyridostatin, can suppress *MYC* expression, reducing tumor cell proliferation.[29b] Conversely, disruption or destabilization of GQs, either due to mutations or altered cellular conditions, can enhance transcriptional activation of oncogenes like *MYC* or *KRAS*.[29c] This misregulation can drive tumor progression and contribute to therapy resistance, underscoring the critical regulatory role of GQs in oncogene transcription.[30] In oncogenes like *MYC*, *KRAS*, and *BCL-2*, GQs often act as molecular switches, toggling gene expression on or off based on their structural stability.[29a,30,31] Stabilizing these GQs suppresses oncogene transcription, while destabilization can lead to aberrant activation, further promoting tumorigenesis. The evolutionary selection of GQs as regulatory elements is particularly evident in gene promoters near transcription start sites. Promoter-associated GQs are correlated with open chromatin regions and nuclease hypersensitivity, facilitating greater accessibility for transcription machinery. Specific loop structures within GQs contribute to their stability, maintaining defined folds that influence gene activity.[29b,31] Given their prevalence in oncogene promoters and their regulatory potential, GQs are attractive targets for cancer therapies.

Strategies aimed at either stabilizing or destabilizing these structures offer promising approaches to modulate oncogene expression and counter tumor progression.[32]

Human telomeric GQ: Telomeres, positioned at the ends of linear chromosomes, are specialized DNA-protein complexes essential for maintaining chromosomal stability. These structures protect chromosome ends from degradation, prevent inappropriate fusion, and protect genomic integrity.[33a] Telomeres consist of a double-stranded DNA region and a G-rich single-stranded overhang. The single-stranded overhang is particularly significant as it can fold into stable GQ structures *in vitro*, providing an additional layer of protection against nuclease activity and unwanted repair mechanisms.[33b,c] This structural stability contributes to telomeres' ability to prevent chromosome ends from being mistaken as double-strand breaks, which could otherwise activate cellular DNA repair pathways, potentially leading to erroneous repair or fusion events.[34] Additionally, telomeres address the "end replication problem," a limitation of DNA polymerases that prevents the complete replication of chromosome ends during DNA synthesis. Without telomeres, this gradual loss of genetic material would compromise genomic stability and cellular viability, highlighting their critical role in cellular longevity and genome maintenance. The formation of GQ structures further enhances telomere function, ensuring their stability and protecting chromosome ends from replication stress and genomic instability.[34]

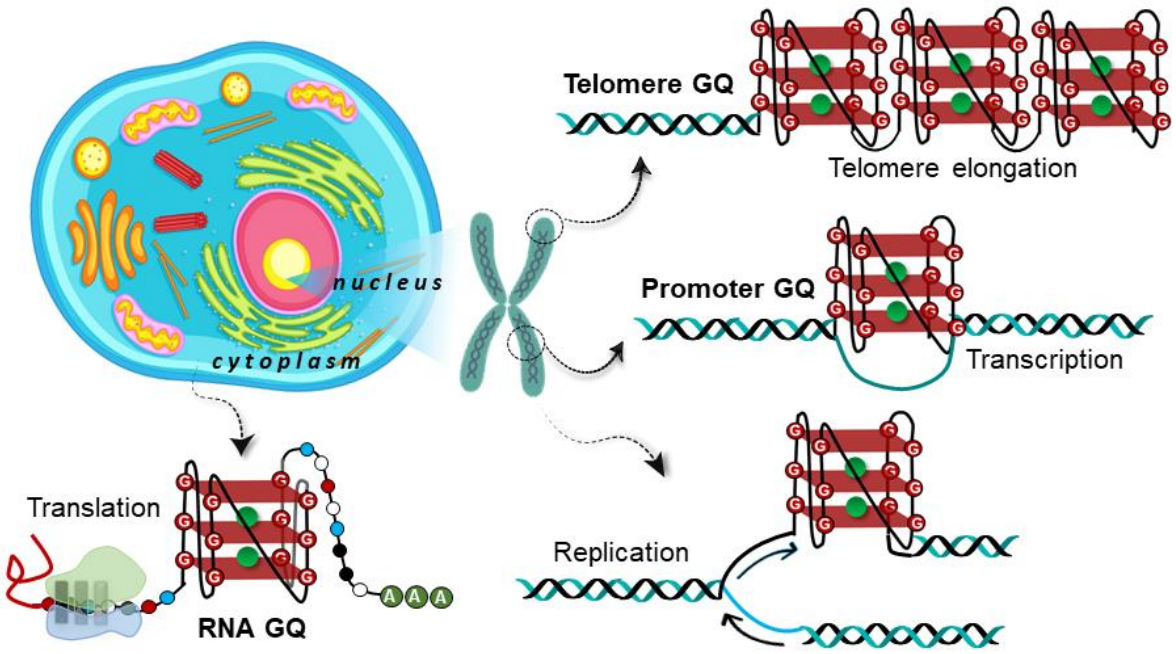


Figure 2. Different biological functions of GQs. Cell image was adopted from https://www.geeksforgeeks.org/overview-of-cell-biology/

1.2.2.2 RNA GQs

While DNA primarily exists in a double-stranded form, RNA is generally single-stranded and often adopts diverse secondary structures, such as hairpins, loops, bulges, and pseudoknots.[35a] In eukaryotic cells, DNA is confined to the nucleus, where it associates with histones and other regulatory factors, while RNA is distributed across both nuclear and cytoplasmic compartments, where it interacts with various protein partners.[35b] This distribution and functional versatility of RNA are reflected in its structural properties, where the 2′-OH group influences the sugar pucker and promotes O2′–O4′ hydrogen bonding between nucleotides, creating a scaffold for a water network within the GQ grooves.[36a,b] This network stabilizes the sugar pucker and locks the sugar-phosphate backbone in the *anti*-conformation, resulting in a single, highly stable RNA GQ structure. Also, the 2′-OH, therefore, plays a pivotal role in RNA's unique structural rigidity compared to DNA, making RNA GQs less flexible and more stable.[36] They are predominantly present in mRNAs, noncoding RNAs, mitochondrial RNAs, and microRNAs.[37] They participate in multiple cellular processes, illustrating the complex nature of GQ biology. Within mRNAs, they are found in the 5′-UTR, 3′-UTR, and coding regions, where they generally inhibit translation by impeding the formation of the pre-initiation complex.[38] This inhibitory effect correlates with the stability of the GQ structure, though RNA helicases such as DHX9 and DHX36 can unfold these structures, relieving the block and enabling translation.[39] Overall, RNA GQs are critical regulators of gene expression and cellular processes. Their stability, along with the unwinding activity of RNA helicases, plays a key role in controlling translation and other essential functions, making them promising targets for therapeutic intervention.

1.2.3 Multimeric GQs

GQs employ two main mechanisms for multimerization. In the first mechanism, interfaces are created from two separate GQs. The individual nucleic acid strands initially form monomeric GQs, where each G-tetrad is composed entirely of guanines from the same strand. These monomeric GQ units then stack on top of one another through π–π stacking interactions between their terminal interfaces, leading to the formation of higher-order GQ structures.[40] The GQs can stack in three possible orientations: 5′ to 3′ (head-to-tail), 5′ to 5′ (head-to-head), and 3′ to 3′ (tail-to-tail).[40b] Research and molecular dynamics simulations show that the 5′ to 5′ stacking orientation is the most commonly observed, owing to its optimal stacking geometry (Figure 3, top right).[41] This stacking process is influenced by the topology and loop orientation

of GQ. In parallel GQs, where the strands run in the same direction and loops are positioned alongside the tetrads, stacking is more favorable. Similarly, the hybrid GQ topology is conducive to multimeric folding due to its loop arrangements.[42] In contrast, antiparallel GQs, where strands run in opposite directions and loops extend away from the GQ core, may hinder stacking interactions between terminal G-tetrads or interfere with strand bending, limiting the formation of subsequent unimolecular GQ structures.[43] Similarly, in an alternate mechanism, the multimeric GQs form through G-tetrads, which incorporate guanine residues from distinct nucleic acid strands and are referred to as intermolecular GQs.[44] Intermolecular GQs can adopt dimeric, trimeric, or tetrameric forms, with the participating strands originating either solely from DNA or from both DNA and RNA molecules, forming what is known as a DNA hybrid (Figure 3, top and bottom left).[44] Higher-order GQs have become specific targets in the promoters of thousands of genes throughout the human genome.[45] Although these multimeric GQs share some of the same protein-binding properties as monomeric GQs, they also seem to regulate transcriptional activity, potentially functioning as adjustable "dimmer switches" for gene expression.[45] Structurally, they provide a more complex drug-targeting landscape, with GQ-GQ interaction sites and protein-like binding pockets formed by unique loop sequences. A notable example is the hTERT core promoter higher-order GQ, which has been the focus of over a decade of research, leading to the discovery of a specialized receptor and the first drug-like small molecule that selectively targets a higher-order GQ.[46]

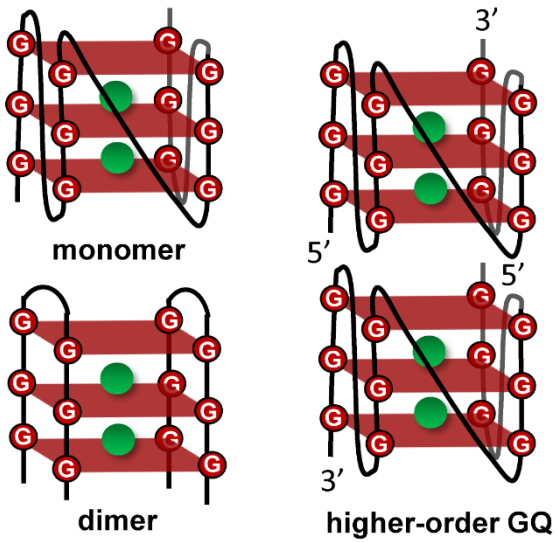


Figure 3. The top left shows a monomeric GQ. The bottom left illustrates a dimeric GQ, where GQs are connected through two strands. The top right shows a stacked intermolecular GQ having 5'–5' head-to-head interaction, where multiple (two) GQ monomers stack on top of each other, connected by loops and stabilized by stacking interactions between the guanine quartets. The orientations are indicated to show the directionality of the strands.

1.2.4 GQ in viruses

Viruses are intracellular parasites that replicate by hijacking the host's replication and protein synthesis machinery.[47a] According to the Baltimore classification, human viruses are classified into seven groups based on their genome type and replication mechanism: 1) DNA viruses with double-stranded (ds) or single-stranded (ss) genomes, 2) RNA viruses with ds or ss genomes of positive (ssRNA (+)) or negative (ssRNA (-)) polarity, and 3) retroviruses that convert RNA into DNA during replication via reverse transcription.[47b] These viruses exhibit diverse genetic material and can infect all domains of life—bacteria, archaea, and eukaryotes.[48] Despite advancements in antiviral treatments, such as direct-acting antivirals for hepatitis C and antiretroviral therapy for HIV, current therapies often fall short of fully eradicating most viruses or achieving long-term control. This limitation, compounded by a shortage of drugs targeting many human viruses, underscores the critical need for identifying new antiviral targets.[49] The prevalence of putative GQ-forming sequences across viral genomes, despite variations between virus families, suggests a conserved and significant role in viral function.[50] Targeting these GQs with specific ligands could not only provide novel antiviral treatments but also help enhance our understanding of viral mechanisms, especially in the context of drug resistance.

1.2.5 GQ ligands

As GQs are highly polymorphic and differ significantly from canonical duplexes or single-stranded configurations, they present an encouraging avenue for pharmacological intervention, particularly in combating viruses.[51,52] Their diverse structural characteristics make them suitable for specific targeting with small molecules. Ligand design and in silico screening efforts have yielded thousands of GQ ligands catalogued in the GQ Ligands Database (http://www.g4ldb.com), with a few undergoing clinical trials for cancer therapy and showing promise for virological applications.[53,54] Importantly, GQ ligands from different structural families, demonstrate the ability to hinder the replication of both DNA viruses such as herpesviruses and hepatitis-B virus, and RNA viruses such as HIV-1 and hepatitis C virus.[55] Despite the considerable evidence pointing to the significant involvement of GQs in viral functions, their mechanisms of action remain poorly understood. Also, the structural polymorphism of GQs and their shared structural framework make it challenging to design ligands with high specificity for a particular GQ topology.[56] Significant efforts have been directed to enhance the selectivity and binding strengths of ligands, that can specifically

interact with distinct GQ targets.[57] While many molecules exhibit promising performance in cell-free assays, they often demonstrate limited efficacy in cellular systems or *in vivo* applications.[58] Detailed insights into their absorption, distribution, metabolism, and excretion profiles are therefore essential for optimizing these ligands, refining treatment strategies, and enhancing their therapeutic viability.[57,58]

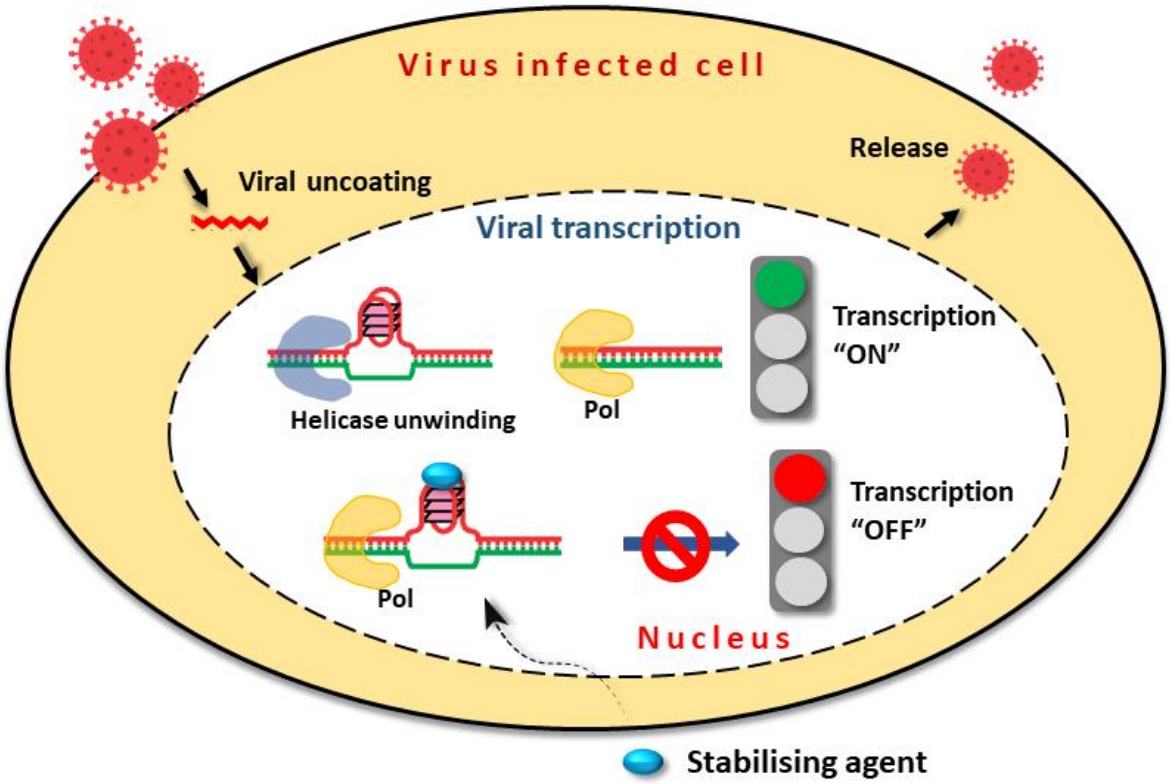


Figure 4. The representative figure illustrates the process of viral infection and transcription regulation within a host cell. Upon viral uncoating (top left), the viral genome enters the host cell. Initially, helicase unwinds the viral genome (top right), allowing polymerase (Pol) to initiate viral transcription (green traffic light, transcription "ON"). Later, the addition of a stabilizing agent (blue circle) halts transcription by preventing further unwinding of the GQ, thus stopping viral transcription (red traffic light, transcription "OFF"). The viral particles are eventually released from the host cell (right). This diagram highlights the dynamic regulation of viral transcription within the infected cell.

1.3 Methods for investigating GQ structures and their interactions with ligands

1.3.1 CD spectroscopy

CD is a well-established biophysical technique widely used to analyze the secondary structure, folding dynamics, and binding properties of chiral molecules.[59] The unique characteristics of CD make it particularly valuable for studying the secondary structures and conformations adopted by nucleic acids.[60] The CD spectral signature of nucleic acids arises from several

factors, including the asymmetric backbone of sugars, the helicity of secondary nucleic acid structures, and the long-range tertiary ordering of nucleic acids under specific solution conditions.[61] This makes CD an effective tool for characterizing GQ topology, exploring the role of cations, examining GQ-ligand interactions, and understanding thermal stabilization induced by ligands. For example, B-DNA displays a positive band between 260 nm and 280 nm and a negative band around 245 nm.[60] The CD spectral features of GQ structures serve as indicators of their topology: parallel GQs exhibit a positive band around ~260 nm and a negative dip at ~240 nm; antiparallel GQs show a positive band at ~295 nm and a negative band at ~260 nm, while hybrid GQs display two positive bands (~295 nm and ~260 nm) and a negative one at ~245 nm.[62a,b] These distinctive spectral features make CD the method of choice for studying GQ structures.[62a] Furthermore, the CD provides a rapid means of assessing GQ/ligand interactions, especially for interactions occurring on short timescales.[63] Typically, GQ ligands lack inherent chirality and do not exhibit a CD signal in solution. However, when these ligands interact with GQ structures, they can develop an induced CD (ICD) signal due to the alignment of electric transition moments between the ligand and the GQ bases.[64] The presence of an ICD signal indicates GQ/ligand interactions and can even alter the characteristic signature of the GQ topology itself. These advantages have led to an increased use of CD for studying the effects of ligands on GQs. Despite its many benefits, CD often provides qualitative rather than quantitative data due to its relatively low signal-to-noise ratio. While CD is effective in observing the impact of ligands on GQ structures, it has limitations, such as its inability to determine stoichiometry or calculate binding constants, which typically require complementary methods. Additionally, CD struggles to differentiate between multiple coexisting structures in solution and has difficulty distinguishing between duplex and GQ competition.[65]

1.3.2 UV-spectroscopy - Thermal melting

Nucleic acids absorb strongly around 260 nm due to the presence of heterocyclic purine and pyrimidine rings. However, this absorbance diminishes in structured forms like duplex DNA, because of π-stacking and hydrogen bonding.[66] Heating disrupts these interactions, exposing the bases to the solvent and causing a rise in absorbance at 260 nm—a principle, widely used in thermal melting experiments to evaluate the stabilization or destabilization of DNA structures.[67] Interestingly, certain secondary structures, such as GQs and i-motifs, show distinctive absorbance changes at 295 nm, driven by the n→π* transition unique to their

tetraplex stacking interactions.[67b] Also, a ligand that stabilizes the folded structure will increase the observed melting temperature in both duplex and quadruplex nucleic acids, indicating preferential binding to the structured form.[66] However, for mixtures of multiple quadruplex conformations in equilibrium, UV thermal melting may not resolve individual melting temperatures.

1.3.3 TDS

Thermal Difference Spectroscopy (TDS) relies on the fact that nucleic acid absorbance depends on both its sequence and structure. Differences in absorbance are often observed between the structured and unstructured forms of the same oligonucleotide.[68a] By measuring absorbance across wavelengths from 220 to 340 nm at two distinct temperatures, TDS generates a differential absorbance curve.[68] This curve is created by subtracting the absorbance spectrum at high temperature, where the oligonucleotide is fully unfolded, from that at low temperature, where it is fully folded. TDS assumes that high temperatures lead to complete unfolding, although certain structures, like some GQs, may retain structure even at elevated temperatures. This technique is useful for identifying specific GQs as well as other DNA structures, including B- and Z-DNA, duplexes, triplexes, and i-motif structures.[68a] A typical GQ TDS profile features two positive peaks at 249 and 270 nm and a negative peak at 295 nm. TDS is a straightforward and complementary method for identifying the formation of GQ, however, it cannot differentiate between various GQ topologies.

1.3.4 Polymerase stop assay

The *Taq* polymerase assay is a method to study GQ structures by leveraging the polymerase's inability to efficiently replicate through stable GQ regions.[69a] By incorporating GQ-forming sequences into template DNA, the assay detects stalling of *Taq* polymerase, which is visualized as truncated or reduced amplification products. It is used to assess GQ stability, study ligand-induced stabilization, and explore GQ impact on replication.[69] GQ structures in DNA or RNA can stall DNA polymerase or reverse transcriptase activity.[70a] By comparing polymerase pause sites under GQ-stabilizing conditions (such as in the presence of K^+ or a stabilizing ligand) versus non-stabilizing conditions (such as Li^+), researchers can pinpoint the 5′ ends of GQs *in vitro*. Originally, this technique involved sequence analysis on polyacrylamide gels, but it has since evolved into a genome-wide DNA polymerase-stalling assay followed by high-throughput sequencing (GQ-seq). GQ-seq revealed over 700,000 DNA GQ sites in the human

genome, including non-canonical GQ structures that are challenging to predict computationally. Some GQs predicted by algorithms were not observed, underscoring the value of experimental mapping.[70b] GQ-seq reference maps are now available for model organisms. A similar approach, reverse transcriptase stalling (rGQ-seq), was applied to poly(A)-enriched RNAs, mapping RNA GQ structures in over 3,000 human mRNAs.[70a,c] While these techniques provide valuable *in vitro* GQ maps, they reflect population averages and do not account for potential protein influences on GQ formation.

1.3.5 NMR

NMR spectroscopy is an exceptionally powerful technique for studying nucleic acid structures, offering atomic-level, high-resolution insights in solution. Beyond structural details, it allows examination of nucleic acid folding kinetics, dynamics, and interactions with proteins and small molecules.[71] During the folding process, nucleic acids form distinct hydrogen bonds—such as Watson-Crick (WC), Hoogsteen, and hemiprotonated C-C^+ bonds—which are identifiable by unique 1D NMR imino proton signals. Duplex structures with WC hydrogen bonds show imino proton signals in the 12−14 ppm range, while GQs with Hoogsteen bonds exhibit signals from 10−12 ppm. Likewise, i-motif structures with hemiprotonated C-C^+ pairs are characterized by signals between 15−17 ppm.[72] In some nucleic acid sequences, structural complexity results in broad or overlapping peaks due to signal convergence caused by dynamic equilibria among multiple conformers.[72c] Promoter regions of oncogenes, for example, can adopt various GQ conformations, creating a challenging structural equilibrium. To achieve well-resolved imino proton peaks, researchers often use mutated sequences engineered to favor a single structural topology, facilitating more precise structural analysis. Traditionally, GQ structures were explored using NOESY spectroscopy to detect proton-proton interactions.[71d] However, the model-based approach of NOESY could introduce ambiguity in structure assignment. This issue has been addressed by site-specific isotope labeling with low-level ^{13}C enrichment, allowing for more accurate assignment of imino protons and other resonances. For determining the solution structure of GQ, a combination of NOESY, COSY, TOCSY, ^{13}C-^{1}H HMBC, and ^{13}C-^{1}H and ^{15}N-^{1}H HSQC experiments is typically used, either with native or isotope-labeled oligonucleotides.[73] However, this approach is generally limited to sequences under 100 nucleotides, as larger nucleic acid structures exhibit extensive resonance overlap. NMR spectroscopy, recognized for its non-invasive approach and sensitivity to nucleic acid structural changes, is widely used to study nucleic acid interactions *in vivo*. Traditionally, *Xenopus laevis* oocytes, about 1 mm in diameter, have been the preferred model for eukaryotic

in-cell NMR, allowing the direct microinjection of exogenous nucleic acids.[74,75] Preparing samples typically involves injecting 150–200 oocytes with 30–50 nL of a 3–10 mM nucleic acid solution, yielding intracellular concentrations of approximately 150–250 μM.[76] Measurements are taken at 18 °C, the physiological temperature for these oocytes. However, the NMR spectra generally exhibit signal broadening and reduced resolution due to the intracellular viscosity and sample variability within the oocytes.[76]

The Trantírek and Dötsch teams developed a novel in-cell NMR method to study nucleic acid structures in *Xenopus laevis* oocytes. Using 1D NMR spectroscopy, they compared the structural formation of the human telomeric GQ across *in vitro, ex-vivo* (within PEG, Ficol 70, and *Xenopus laevis* egg extract), and in-cell environments. Their results revealed that, in cellular conditions, telomeric sequences could exhibit a mix of antiparallel, hybrid, and parallel topologies.[77] Trantírek's group recently expanded this method to observe DNA-ligand interactions within the nuclei of human cells.[78] Nonetheless, in-cell NMR faces challenges, such as the limited stability of oligonucleotides (ONs) in cellular environments and low resolution of imino proton signals in *ex-vivo* and *in vivo* conditions. Consequently, it may not be suitable for analyzing highly complex structures within cells. Additionally, because promoter regions of certain oncogenes may equilibrate among various GQ forms, NMR is less effective for tracking these dynamic equilibria.

^{1}H and ^{15}N NMR spectroscopy, while extensively employed for studying GQ architectures in living *Xenopus laevis* oocytes, also face obstacles such as the limited resolution of ^{1}H spectra and the high cost of ^{15}N labeling, which hinder their broader use.[79a] To address these limitations, ^{19}F NMR has emerged as a powerful alternative. This technique offers enhanced sensitivity and resolution for studying nucleic acid structures, tracking their dynamics, and examining interactions with proteins and ligands, making it a promising tool for advancing in-cell studies (Figure 5, **6**–**10**).[79] Although it does not provide the atomic-level detail characteristic of ^{1}H and ^{13}C NMR, ^{19}F NMR is especially useful for assessing structural dynamics and molecular interactions under various conditions. To facilitate this, numerous ^{19}F-labeled nucleoside modifications have been designed and integrated into DNA and RNA using either solid-phase synthesis or enzymatic methods.[79]

In our lab, we synthesized dual-labeled 5-fluoro-benzofuran derivatives of deoxyuridine and uridine to explore nucleic acid interactions via ^{19}F NMR. Using a modified deoxyuridine analog, we examined the human telomeric GQ structure and ligand interactions, which revealed a hybrid-type 2 GQ topology within live cells. We also used 5-fluoro-benzofuran-modified uridine analog to study RNA conformational variations.[80] To enhance

dual-probe sensitivity, a 5-trifluoromethyl-benzofuran ring was added to modified nucleosides at the C5 position of 2'- deoxyurindine and uridine nucleosides. This was used in the G-rich sequence of the EGFR oncogene promoter to examine GQ structure populations and in the Braf C-rich sequence to identify the main i-motif structure.[81] The modified RNA analog also facilitated the study of aminoglycoside-induced changes in bacterial ribosomal RNA.[81c]

Additionally, Xu and colleagues introduced 3,5-bis(trifluoromethyl)-benzene-labeled deoxyguanosine and deoxyuridine probes to investigate the human telomeric GQ structure and the thrombin-binding aptamer.[82] They further attached the 3,5-bis(trifluoromethyl)-benzene core at the 5' position of H-Telo DNA and RNA sequences to detect GQ topologies within cells. Their live-cell ^{19}F NMR analysis showed that TERRA RNA adopted a parallel GQ structure in cellular environments.[83] Also, Li group has developed a new ^{19}F NMR fingerprinting approach designed to quickly identify and quantify the topology of H-Telo GQs. This technique involves systematically analyzing the ^{19}F chemical shifts of a commercially available 2′F-rG probe at specific positions in 12 H-Telo sequences, each of which is known to adopt distinct intramolecular GQ topologies.[84]

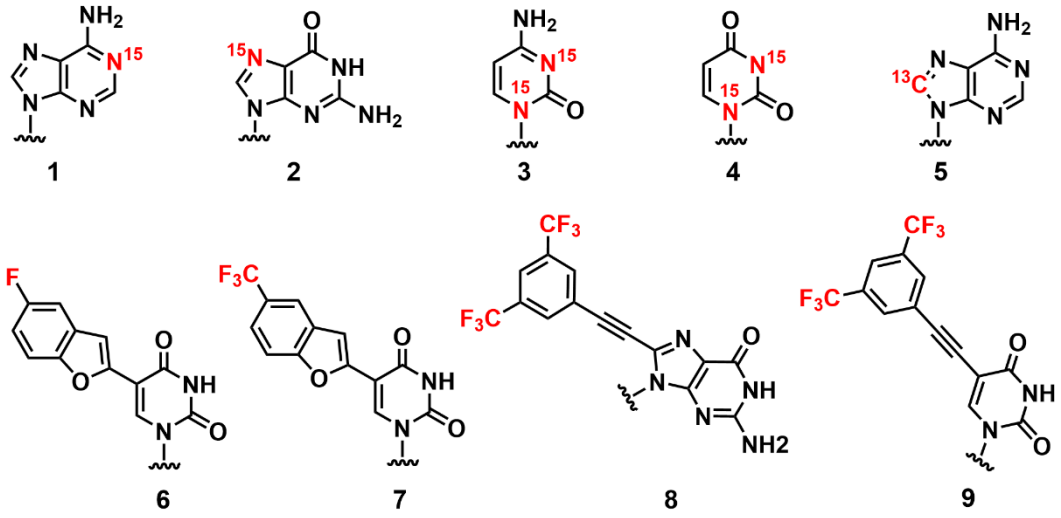


Figure 5. NMR-active labels (^{15}N, ^{13}C, and ^{19}F) are used to investigate nucleic acid structures through NMR spectroscopy

1.3.6 EPR

EPR spectroscopy is a highly sensitive technique for studying nucleic acid structures and dynamics, leveraging the magnetic properties of unpaired electrons.[85] Unlike NMR, which detects nuclear spins, EPR uses stable spin labels—typically nitroxide radicals—introduced into nucleic acids to avoid background interference from their natural diamagnetic state.[85a] These spin-labeled nucleosides must remain chemically inert and stable under experimental

conditions, ensuring their suitability for studies in both *in vitro* and *in vivo* environments (Figure 5). Various labeling strategies enable the attachment of nitroxide radicals to nucleobases, sugars, or phosphate backbones, facilitating analysis of nucleic acid structure, dynamics, and interactions with ligands or proteins.[86]

Advanced EPR techniques, such as pulsed electron double resonance (PELDOR) and double electron-electron resonance (DEER), measure inter-spin distances and reveal conformational changes.[87] PELDOR has been particularly useful for monitoring RNA dynamics, RNA-ligand binding, and protein-nucleic acid interactions. For example, the Drescher group used DEER to characterize GQ conformations in the H-Telo sequence in K^+ conditions, distinguishing between parallel and antiparallel topologies. Similarly, the Clever group employed rigid metal complexes as spin labels to investigate GQ dimer structures, identifying head-to-head versus tail-to-tail arrangements through precise distance measurements.[88a] The Qin group's work further extended this approach by using nitroxide-labeled phosphates to study H-Telo GQ folding in various environments (e.g., different salts, molecular crowding, and ligand presence), uncovering a mixture of parallel and antiparallel forms.[88b]

The ability of EPR to detect long-range dipolar couplings and analyze coexisting conformations makes it an exceptional tool for structural studies in cellular contexts.[87] Unconstrained by molecule size, EPR has enabled the investigation of complex structures, such as the H-Telo GQ sequence in recent in-cell studies, which revealed both parallel and three-tetrad antiparallel conformations within cells.[87] However, a notable limitation of EPR is the need for cryogenic temperatures and shock-frozen cells, as free radicals degrade rapidly in live cells, which can hinder the analysis of dynamic processes in their native environment.[88c]

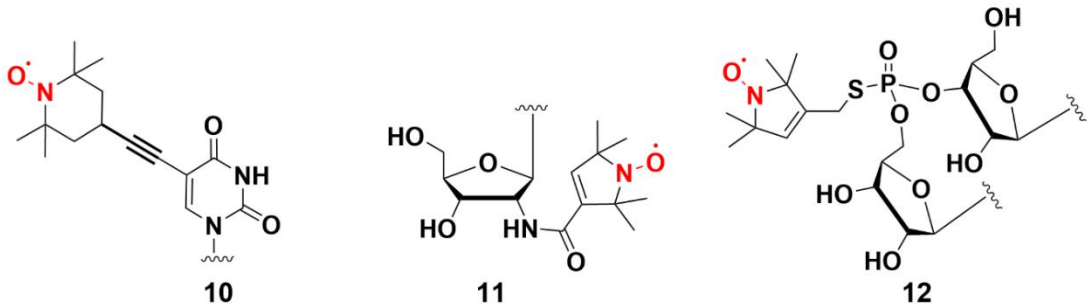


Figure 5. Nucleoside analogs with spin labels for structural analysis of nucleic acids using EPR spectroscopy.

1.3.7 X-ray crystallography

X-ray crystallography, a vital technique for obtaining detailed three-dimensional structural information about nucleic acids, has successfully produced high-resolution structures of GQs, some with resolutions finer than 1 Å.[89] The crystal structure is determined by non-covalent interactions among nucleic acid molecules, with water molecules occupying the spaces between them.[90] Mapping electron density in these structures can be challenging, as similar atoms (hydrogen, carbon, nitrogen, and oxygen) can produce indistinguishable diffraction patterns. To address this, heavy atom derivatization is commonly employed to enhance electron density mapping.[90] Bromine, iodine, and selenium are often used as heavy atom modifications along with cationic forms of alkali metals, transition metals, and lanthanides, which can be incorporated into the crystal structure through soaking methods (Figure 6).[90–92] It is crucial that the introduction of heavy atoms does not disrupt the native structure of the nucleic acids. Heavy elements, which contain a higher number of electrons, significantly alter the diffraction pattern compared to the native crystal.[93] For instance, selenium-modified nucleosides have been incorporated into nucleic acids for crystallization purposes (Figure 6). However, due to the structural polymorphism exhibited by nucleic acids, predicting whether they will adopt the same topology under crystallization conditions—such as synthetic molecular crowding and dehydration conditions—remains challenging, particularly in complex cellular environments.[94–100]

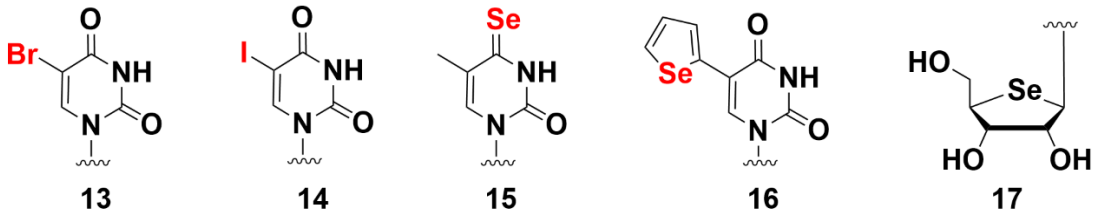


Figure 6. Nucleoside analogs containing heavy elements (Br, I, Se) for studying nucleic acid structures through X-ray crystallography.

1.3.8 Fluorescence spectroscopy

Fluorescence-based techniques offer powerful tools for studying nucleic acid structures, folding dynamics, and molecular recognition in real-time, with high sensitivity. Since native nucleosides lack inherent fluorescence, researchers either employ non-covalent fluorescent dyes or modify ONs by incorporating fluorescent nucleoside analogs to monitor nucleic acid behaviour. These fluorescent probes respond to conformational changes in nucleic acids by altering their emission properties, such as emission wavelength, intensity, quantum yield,

anisotropy, and lifetime.[101] Techniques like Fluorescence Resonance Energy Transfer (FRET), structure-specific probes, and fluorescent nucleoside analogs are particularly useful for investigating various nucleic acid architectures, structural transitions, and molecular recognition events.[102] Specifically, covalently attached, environment-sensitive fluorescent nucleoside analogs have proven highly effective for distinguishing between different nucleic acid structures and quantifying ligand-binding interactions. These fluorescence-based strategies are invaluable for probing the intricate structural dynamics of nucleic acids, including complex non-canonical forms like GQs. Herein, we have discussed the various fluorescent labeling techniques for the detection of GQ structures and their recognition properties *in vitro* and *in vivo.*

1.3.8.1 External fluorescent probes

FRET is a valuable technique for gathering structural information about GQs *in vivo*. It measures changes in distances within or between macromolecules and has been utilized to reveal the diverse conformations that GQs can adopt.[103a] In FRET experiments, GQ ONs are tagged with a donor fluorophore at the 5′-end and an acceptor fluorophore at the 3′-end. When the donor is excited, its energy is transferred to the acceptor through dipole–dipole interactions.[103b] Variations in the distance between the labeled sites lead to detectable changes in energy transfer. Oligonucleotides that can fold into intramolecular quadruplexes are labeled at the 3′- and 5′-ends with a fluorophore and a quencher (for example, 5′-fluorescein and 3′-tetramethylrhodamine). When a quadruplex forms, the fluorescence is quenched due to the proximity of the quencher to the fluorophore. Conversely, during thermal denaturation of the quadruplex, the fluorophore and quencher separate, leading to a restoration of fluorescence. This change in fluorescence is measured to generate melting curves.[104] Nevertheless, this method offers considerable advantages and has been extensively utilized to study the binding of various types of GQ ligands, including phenanthrolines, acridines, indoloquinolines, and pyridine-2,6-dicarboxamide derivatives, to different GQ structures.[104]

While external fluorescent probes serve as powerful tools for studying GQ conformations, they also pose inherent challenges that may influence native structures and lead to potential artifacts. The introduction of large fluorophores, such as FRET donor-acceptor pairs, can introduce steric hindrance, altering GQ folding topology or stability. For instance, studies have shown that bulky fluorophore modifications at loop or terminal positions can shift the equilibrium between parallel and antiparallel topologies.[104c] Similarly, extending

conjugation in nitrogenous bases to enhance fluorescence properties can impact base stacking interactions, which are critical for GQ stability. A study by Shirude et al. demonstrated that the incorporation of large chromophoric groups could alter cation coordination within the GQ core, leading to changes in folding kinetics and stability.[104d]

Furthermore, some fluorophores exhibit intrinsic interactions with nucleic acids, potentially leading to additional stabilization or unwarranted structural rearrangements. For example, Cy3 and Cy5 fluorophores, commonly used in FRET studies, have been reported to interact with G-rich sequences, thereby influencing structural transitions.[104e] To minimize such perturbations, alternative labeling strategies, such as minimalistic fluorophores or site-specific modifications that do not interfere with key structural elements, should be considered.[104f] Therefore, the use of internal labels at positions less likely to alter GQ folding or the adoption of environment-sensitive probes that do not require large structural modifications may provide more reliable structural insights.

Internal fluorescent probes

An ideal probe for GQ detection should possess several important characteristics, including a high luminescence quantum yield, significant emission enhancement upon binding to GQs, strong photostability, and chemical robustness.[101] It should also have high selectivity for GQ DNA over other DNA forms, such as single-stranded or duplex DNA, and exhibit good solubility. For cellular imaging applications, additional attributes are beneficial (discussed below). Although no universally recognized gold standard exists for GQ probes, several molecules with remarkable GQ-binding capabilities have been identified.

1.3.8.2 Fluorescent nucleosides

Using fluorescent nucleoside analogs positioned at specific sites within GQ structures can reveal key photophysical information about their organization and interactions with proteins and ligands. This approach supports the design of binding assays for identifying ligands that target GQ with high specificity. A range of fluorescent nucleoside analogs have been synthesized for this purpose, and these can be grouped into distinct categories based on their structural and functional properties.

Isomorphic fluorescent nucleoside probes

Isomorphic nucleoside analogs differ from their natural counterparts primarily by the presence of minor substituents or by changes in the number and positioning of heteroatoms within the heterocyclic ring structure.[105] 2-Aminopurine (2AP) is frequently used as a probe to examine the structure, dynamics, and interactions of nucleic acids with proteins and ligands. Being structurally similar to adenine, it forms WC hydrogen bonds with thymine and exhibits strong fluorescence emission.[105] The photophysical properties of 2AP are highly responsive to changes in solvent polarity. Majima and colleagues demonstrated the use of 2AP to track human telomeric GQ formation by replacing a loop adenosine in the H-Telo GQ sequence with 2AP (Figure 7, **18**).[106] This substitution did not disrupt the native GQ structure and resulted in higher fluorescence and a longer fluorescence lifetime in the antiparallel GQ structure compared to the duplex form. In the duplex, 2AP undergoes fluorescence quenching due to stacking interactions with adjacent guanine residues, which facilitate electron transfer.[106] However, in the GQ structure, loop residue distortions reduce these stacking interactions, exposing 2AP to the solvent and enhancing its fluorescence. With excitation and emission wavelengths near 303 nm and 370 nm, respectively, and a quantum yield of 0.68 in water, 2AP-riboside has proven exceptionally valuable due to its remarkable sensitivity to environmental changes.[107]

Similarly, Sugiyama and colleagues applied a 2AP analog to detect GQ and i-motif formation at the 5' end of the retinoblastoma (Rb) gene, showing enhanced fluorescence and longer lifetime in the tetraplex structures compared to the duplex form.[108a] This fluorescence enhancement in tetraplexes likely arises from reduced stacking with neighboring bases. Despite its effectiveness as a fluorescent probe, 2AP has limitations: its excitation and emission fall within the UV range, and its quantum yield significantly decreases when incorporated into oligonucleotides, thereby limiting its sensitivity. In a recent study, Kankai and colleagues utilized 2AP to examine the folding of thrombin and G3T aptamers, which adopt antiparallel and parallel GQ conformations, respectively. They introduced 2AP into the loop regions of the thrombin aptamer sequence and the G3T aptamer sequence, known for its affinity to HIV-1 integrase. In their unfolded forms, both aptamers displayed low fluorescence. However, upon GQ formation in the presence of K^+ ions, fluorescence intensity significantly increased. Notably, the 2AP-labeled G3T aptamer showed a pronounced fluorescence enhancement, reaching levels comparable to that of the isolated nucleoside analog.[108b]

Expanded fluorescent nucleosides

Expanded nucleosides feature extra aromatic rings fused to the purine or pyrimidine core. For example, thiophene-fused uridine derivative (Figure 7, **19**) known for its fluorescence solvatochromism and high quantum yield has been applied to identify mismatches in duplexes and detect abasic sites in RNA.[109] Meanwhile, benzo-fused fluorescent nucleoside analogs (Figure 7, **20**, **21**) demonstrated their potential in creating an orthogonal genetic system distinct from the natural one.[110] Luedtke and collaborators synthesized modified thymidine (**22**) and cytidine (**23**) analogs, using fluorescence measurements to explore metal base pair formations and their associated thermodynamics.[111] Additionally, tricyclic 2′-deoxycytidine derivatives (tC, tC^O, and tC_{nitro}, **24**–**26**) were co-developed by Wilhelmsson and Albinsson . Among these, tC and tC^O analogs, both as free nucleosides and incorporated into DNA, exhibit high quantum yields, while tC_{nitro} is non-fluorescent.[112] Due to spectral overlap between tC^O's emission and tC_{nitro}'s absorption maxima, these analogs have been utilized as a FRET pair to measure distances between DNA bases at varied positions.[113]

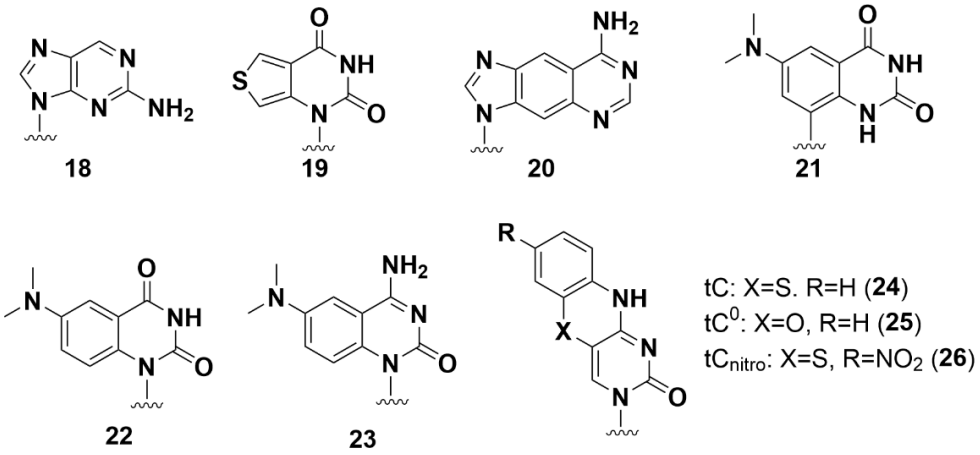


Figure 7. Structural representations of isomorphic and size-expanded fluorescent nucleoside analogs.

Extended conjugation based fluorescent nucleosides

Tor's team developed fluorescent nucleoside analogs by attaching various heterocyclic rings to nitrogenous purine and pyrimidine bases. Among the rings tested—such as furan, thiophene, oxazole, and thiazole—they found that furan-modified pyrimidine nucleosides displayed probe-like properties, with emission peaks within the visible spectrum and sensitivity variations in solvent polarity.[114] These properties made the probe useful for applications like

detecting abasic sites, monitoring interactions between RNA and aminoglycosides, and studying RNA-protein interactions.[115]

Similarly, Srivatsan and colleagues synthesized fluorescent nucleosides by linking heterocycles like selenophene, N-methyl indole, benzothiophene, benzofuran, and 5-fluorobenzofuran to the pyrimidine bases (Figure 8).[114b] Among these, modified uridine and deoxyuridine derivatives containing selenophene, benzothiophene, benzofuran, and 5-fluorobenzofuran displayed characteristics suitable for probing nucleic acid structures.[116] These modified nucleosides were employed to decipher the abasic sites in both DNA and RNA.[116cd] In addition, benzofuran-modified deoxyuridine was used to differentiate between various nucleic acid structures—such as duplexes, i-motifs, and GQs—through fluorescence and lifetime measurements. However, the probe's emission in the UV range posed a limitation for cellular studies. This issue was overcome by designing a 5-fluorobenzofuran-modified deoxyuridine nculeoside, which enabled the identification of the H-Telo GQ topology in live cells through in-cell ^{19}F NMR.[80a]

In parallel, Manderville and colleagues synthesized a series of fluorescent guanosine analogs by attaching a variety of heterocyclic groups, including furan, pyrrole, thiophene, benzofuran, indole, benzothiophene, and benzonitrile, to the guanosine base.[114b] These analogs demonstrated moderate to high quantum yields and exhibited solvatochromic and viscochromic behaviors. They applied these fluorescent nucleosides to study GQ structures in the human telomeric sequence and the thrombin-binding aptamer.[114b]

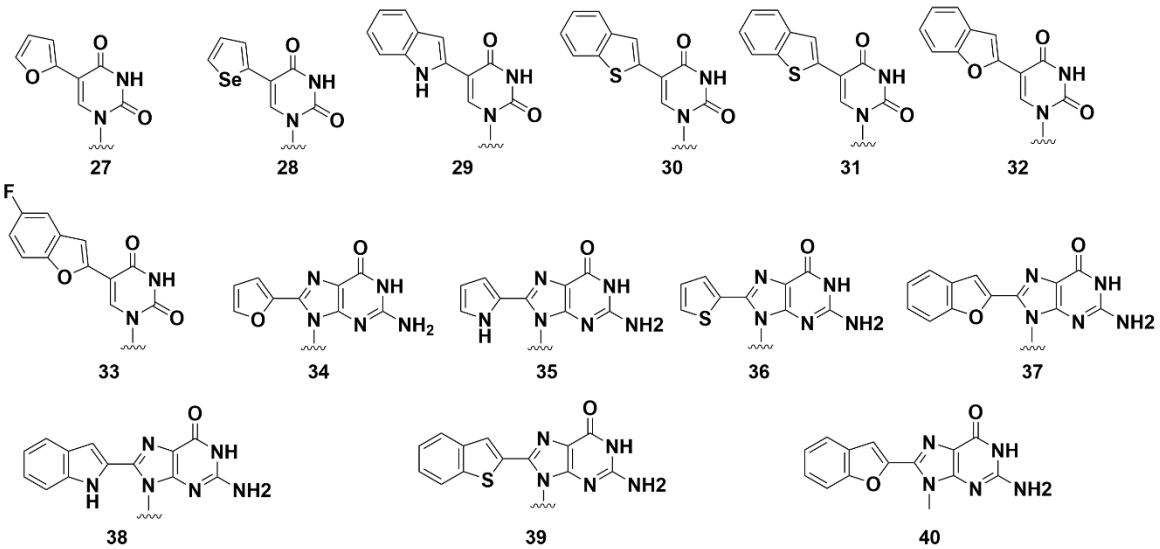


Figure 8. Structural representations of fluorescent nucleoside analogs featuring extended conjugation.

1.3.8.3 Non-covalent probes

Fluorescent small-molecule probes are invaluable for studying GQ structures due to their high sensitivity, spatiotemporal precision, and cost-effectiveness. These non-covalent binders selectively target GQ, triggering fluorescence changes—such as intensity, wavelength, or lifetime shifts—that enable real-time visualization of GQ formation and localization in living cells.[115] Techniques like fluorescence lifetime imaging microscopy further enhance the study of these interactions.[116] An ideal fluorescent probe for GQ analysis should combine high specificity, fluorescence activation upon binding, strong binding affinity, water solubility, membrane permeability, high quantum yield, and low toxicity.[117,118] Advances in probe design have significantly improved our ability to investigate GQ structures and their functional roles in cellular contexts.[116] Among the commonly used probes for GQ detection are cationic dyes such as thiazole orange (**TO**) and thioflavin T (**ThT**) (Figure 9, **41**, **42**), which are known to turn on fluorescence upon binding to GQ structures. These dyes are selective, binding preferentially to GQs over single- and double-stranded nucleic acids.[117a]

In their native aqueous state, these dyes exhibit minimal fluorescence, however, when bound to a GQ structure, they show a significant increase in fluorescence.[119] Recently, Shangguan and collaborators developed a series of TO derivatives, including **TO-5a** and **TO-5b** (Figure 9, **43, 44**). The **TO-5a** derivative demonstrated improved specificity for GQ structures, while **TO-5b** showed a stronger affinity for antiparallel GQs. Notably, using live-cell imaging, they revealed that **TO-5b** successfully enters cells and stains GQ structures in both DNA and RNA located within the nuclei and nucleoli of living cells.[120]

In 2016, Zhou and colleagues created a fluorescent probe, **PDP-Cy5**, by combining the GQ ligand PDS with the Cy5 fluorophore (Figure 9, **45**).[120b,c] This probe binds effectively to both DNA and RNA GQs, enabling visualization in cells or on PAGE gels. Additionally, it stabilizes GQ structures, significantly raising their melting temperature. This effect is particularly pronounced in RNA, which retains its structure even at high temperatures. The design of GQ probes that integrate a GQ-binding ligand with a fluorophore takes advantage of the unique benefits of both components, presenting an effective approach for creating fluorescent probes with excellent photophysical properties specific to GQ structures. Similarly, Balasubramanian and his team developed a GQ probe named **SiR-PyPDS** by combining silicon-rhodamine with the GQ-binding ligand pyridostatin (Figure 9, **46**).[121d] They employed **46** to visualize individual GQs at the single-molecule level in live cells using fluorescence

imaging. Their study enabled the monitoring of both the overall GQ population and the dynamic processes of folding and unfolding within live cells.

Vilar and collaborators have created a unique fluorescent probe, **DAOTA-M2**, which includes a triangulenium core (Figure 11A).[121e] Due to its cell-permeability and low toxicity, **DAOTA-M2** can be utilized for live-cell imaging in real time. One remarkable feature of this probe is its ability to emit distinct lifetimes when bound to different nucleic acid forms, which can be visualized through fluorescence-lifetime imaging microscopy (FLIM). This property allows **DAOTA-M2** to differentiate DNA GQs from both double- and single-stranded DNA based on emission lifetime. Despite its efficacy with DNA, the authors highlight that **DAOTA-M2** could be further optimized, as it currently demonstrates limited ability to selectively identify RNA GQ. Consequently, the development of probes specifically targeting RNA GQs has become a key direction for advancing GQ research (Figure 11B).

Jia-Heng Tan group developed a strategy to detect *NRAS* GQs using click chemistry.[123] By combining the fluorescent GQ probe **ISCH-1** with an antisense oligonucleotide targeting an adjacent G-rich tail, they created a GQ-triggered fluorogenic hybridization (GTFH) probe called **ISCH-nras1**. This probe selectively binds to the G4T25 sequence in *NRAS* mRNA. In cell staining, **ISCH-nras1** produced intense fluorescence foci with *NRAS* mRNA, but no signal was observed with mutated or deleted sequences, indicating its specificity for visualizing NRAS GQs at the cellular level.[123]

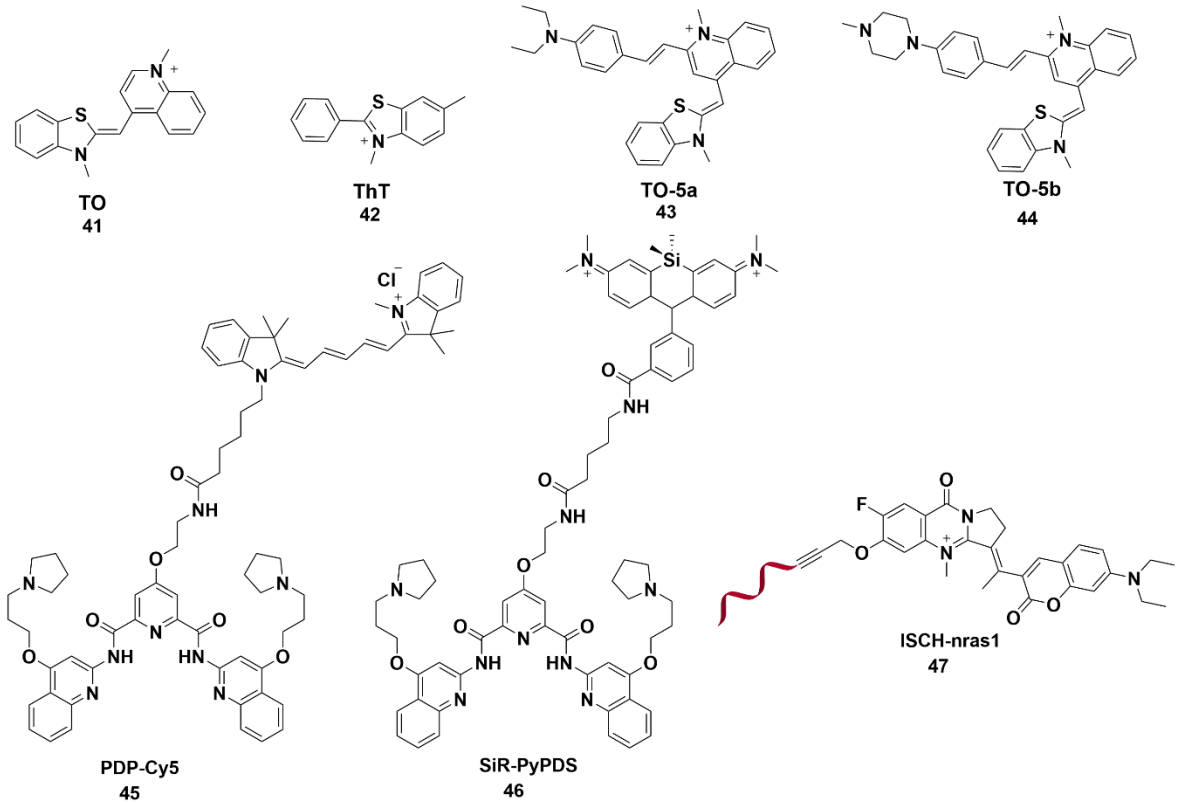


Figure 9. Non-covalent fluorescence light-up probes specific to GQ structures.

In another advancement, Monchaud and his team developed **N-TASQ**, a probe specifically designed for detecting RNA GQs. This probe, which mimics natural GQ structures through synthetic G-quartets, lights up only upon binding to GQs.[122] When unbound, the naphthalene fluorescence in **N-TASQ** is largely suppressed due to quenching by its guanine groups, making it nearly undetectable in its free state. Upon binding with GQ structures, however, **N-TASQ** displays fluorescence peaks at 396 nm and 416 nm, along with a broad emission range from 380 nm to 600 nm. In live cells without prior treatment, **N-TASQ** staining produces prominent cytoplasmic foci, demonstrating its effectiveness in identifying RNA GQs directly within cellular environments (Figure 10).[122]

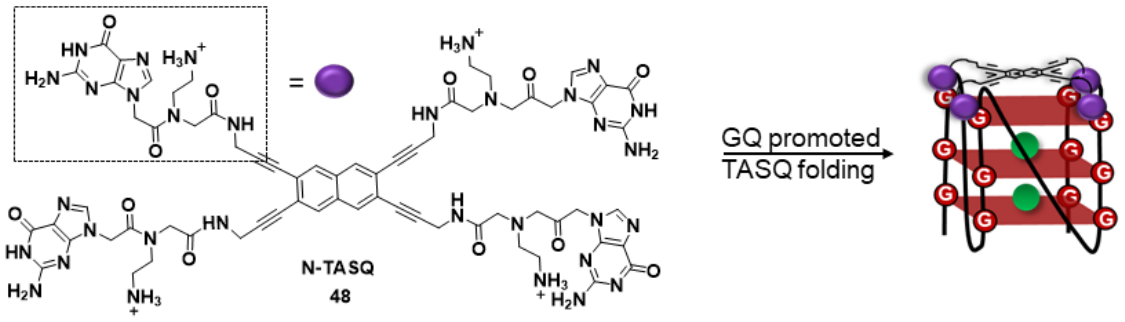


Figure 10. GQ-mediated folding of the **N-TASQ** probe. The **N-TASQ** probe self-assembles into a G-quartet via stacking interactions upon binding to a GQ motif, enabling targeted recognition of GQ structures.

1.3.9 Antibody based tools

A major advancement in identifying the presence and specific locations of GQs *in vivo* occurred over a decade ago with the creation of antibodies targeting telomeric DNA GQs. These enabled the first direct visualization of GQs through *in situ* immunostaining in the micronuclei of *Stylonychia lemnae.*[12,124] Subsequent studies expanded their use to map GQ locations across the human genome via immunoprecipitation and deep sequencing, revealing their prevalence in promoters, untranslated regions (UTRs), introns, and subtelomeric regions, implicating their roles in transcription, termination, and gene regulation.[124] However, these findings were limited by the possibility that some GQs may have formed during the analysis of fragmented DNA rather than reflecting their native states.

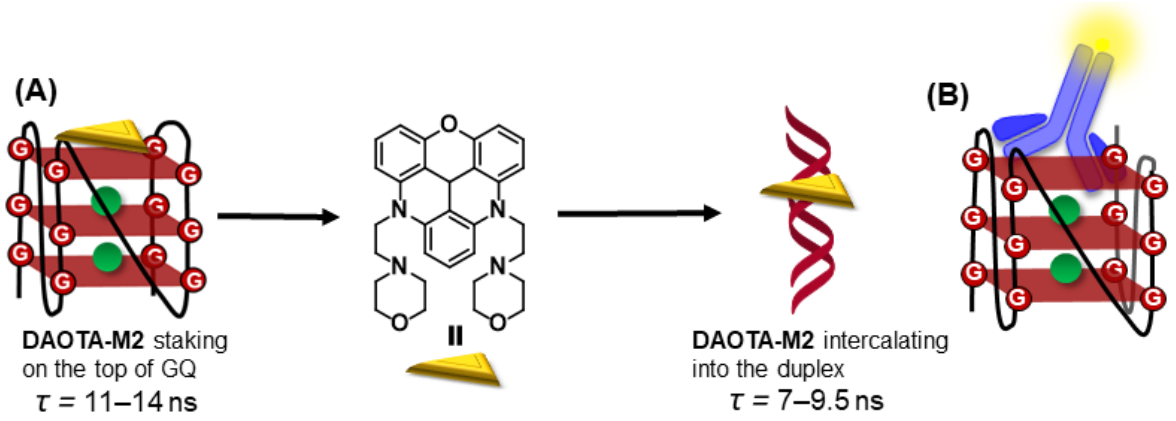


Figure 10. (**A**) **DAOTA-M2** stacking on top of the GQ increases its lifetime, while interaction with the duplex reduces the lifetime. (**B**) Dye-labeled antibody detecting GQ inside cells.

Building on these early insights, researchers have developed a range of advanced tools to detect and study GQs, including antibodies and single-chain variable fragments (scFvs). For instance, BG4, identified through phage display, has been instrumental in visualizing GQs in human cells and mapping their genomic distributions using ChIP-seq.[129a] BG4's ability to detect nuclear foci and cytoplasmic RNA GQs, particularly in combination with GQ stabilizers like pyridostatin, has significantly advanced the field (Figure 11B). Similarly, other scFvs like D1 and antibodies such as 1H6 have provided complementary approaches for studying GQ conformations.[129b–d] Despite the progress, challenges remain in capturing the full diversity of GQ structures and understanding their dynamic formation across different cellular contexts. Emerging tools, including GQ-specific nanobodies and modifications to existing probes for single-cell analyses, offer promising avenues for addressing these gaps and deepening our understanding of GQs' functional roles in both chromatin and RNA regulation.[130]

1.4 Statement of the research problem

The structure-function relationship of secondary nucleic acid structures, particularly GQs, is critical for advancing nucleic acid-based diagnostics and therapeutics. While significant progress has been made using biophysical tools, the structural polymorphism and dynamic transitions of GQs complicate their characterization in cellular environments. Moreover, even with advanced structural techniques, selectively targeting a specific viral GQ amidst the vast repertoire of human genomic GQs remains a significant challenge. Although a range of biophysical methods exist, most available probes have limitations in detecting GQ topologies both *in vitro* and in cellular systems, highlighting the need for more sophisticated platforms to analyze GQ structures comprehensively.

Current methods for analyzing GQ structures predominantly rely on single-label, single-technique approaches, which restrict the depth of insight into these dynamic and complex motifs. This research addresses the limitations of such approaches by introducing a dual-labeled nucleoside analog that integrates both fluorescence and ^{19}F NMR. This dual-label strategy allows for more detailed and sensitive detection of GQ structures, particularly in the context of viral genomes, where G-rich regions play essential roles in viral replication and function. These regions are not only critical for biological processes but some also serve as unique binding domains for ligands, offering possibilities for dual-domain interactions.

This thesis revolves around developing and applying nucleoside-based probe system to probe GQ structures. Central to this work is the usage of a dual-labeled nucleoside analog, 5-fluorobenzofuran-2′-deoxyuridine (FBFdU), which integrates both fluorescence and ^{19}F NMR tags. When used in combination with 5′-fluoro-deoxyuridine, this probe system provides a highly sensitive and minimally invasive approach for detecting and analyzing secondary structures of HIV-1 long terminal repeat (LTR) region. These probes were employed to map the druggable spaces of GQ structures, offering a detailed understanding of the binding modes and interactions of small-molecule ligands with G-rich regions. Furthermore, the RNA analog of the dual-labeled probe (FBFU) enabled the detection of conserved G-rich regions within the SARS-CoV-2 genome by utilizing complementary fluorescence and ^{19}F NMR signatures. This dual-label strategy not only advances structural studies but also holds promise for identifying antiviral drug candidates capable of targeting these viral regions and their corresponding mutants effectively. The minimally perturbing nature of these modified nucleosides, coupled with their distinct and complementary spectral properties, positions them as a powerful tool for advancing the structural analysis of nucleic acids. This work significantly contributes to elucidating the structure-function relationships of nucleic acids and lays the foundation for novel therapeutic approaches targeting viral pathogens.

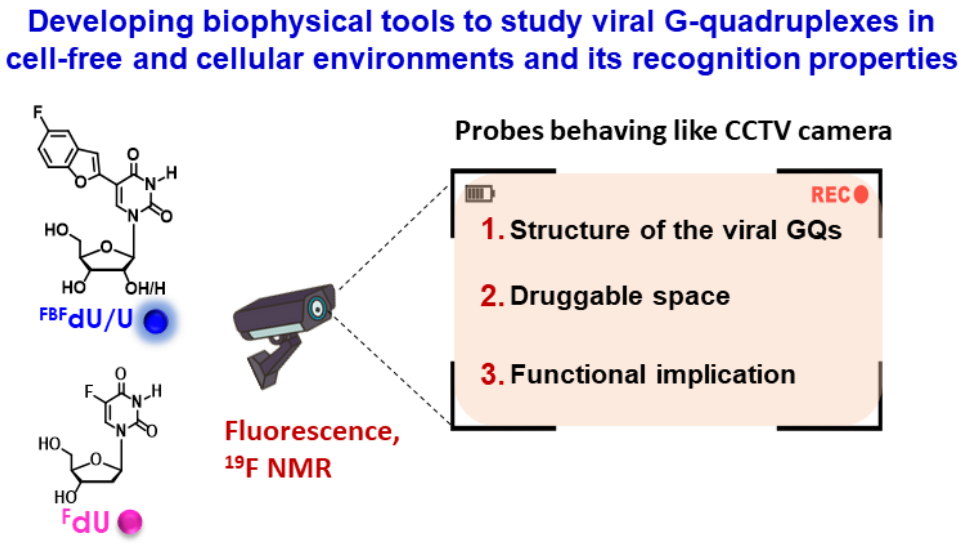


Figure 12. Thesis objectives: Probe systems and their application in examining the secondary structures of viral nucleic acids, their polymorphism, and their interaction with ligands.

1.5 References

1. (a) Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature,* **1953**, *171*, 737–738. (b) Smith, J.; Doe, A. Functions of nucleic acids beyond genetic information. *Nat. Rev. Mol. Cell Biol.* **2020** *21*, 234–245.
2. (a) Lilley, D. M. J. Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.,* **1995**, *28*, 253–300 (b) Tinoco, I., & Bustamante, C. How RNA folds. *J. Mol. Biol.,* **1999,** *293*, 271–281. (c) Cech, T. R.; Steitz, J. A. The noncoding RNA revolution—Trashing old rules to forge new ones. *Cell,* **2014**, *157*, 77–94 (d) Gesteland, R. F.; Cech, T. R., Atkins, J. F. The RNA world: The nature of modern RNA suggests a prebiotic RNA world. **2006** Cold Spring Harbor Laboratory Press. (e) Noller, H. F. RNA structure: Reading the ribosome. *Science,* **2005**, *309*, 1508–1514.
3. (a) Choi J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893–5909. (b) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **2010**, 67, 43–62. (c) Wang, G.; Vasquez, K. M. Modulation of DNA structure formation using small molecules. *Mutat. Res.* **2006**, *598*, 103–119. (c) Bacolla, A.; Wells, R. D. Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.* **2009**, *48*, 273–285. (d) Wells, R. D.; Dere, R.; Hebert, M. L., Napierala, M.; Son, L. S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **2005**, *33*, 3785–3798. (e) Rich, A.; Nordheim, A.; Wang, A. H. The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **1984**, *53*, 791–846.
4. (a) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 474–489. (b) Buske, F. A.; Mattick, J. S.; Bailey, T. L. Potential *in vivo* roles of nucleic acid triple-helices. *RNA Biol.* **2011**, *8*, 427–439. (c) Butcher, S. E.; Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **2011**, *44*, 1302–1311.

5. Lim, K. W.; Alberti, P.; Guédin, A. et al. Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res*. **2009**, *37*, 6239-6248.
6. (a) Tanner, N.K. *FEMS Microbiol. Rev.* **1999**, *23*, 257–275 (b) Puerta-Fernández, E.; Romero-López, C.; Barroso-delJesus, A.; Berzal-Herranz, A. Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiol Rev.* **2003**, *27*, 75–97.
7. (a) Meisburger, S. P.; Pabit, S. A.; Pollack, L. Determining the Locations of Ions and Water around DNA from X-Ray Scattering Measurements. *Biophys J.* **2015**,*108*, 2886–2895. (b) Harshman, S. W.; Young, N. L.; Parthun, M. R.; Freitas, M. A. H1 histones: current perspectives and challenges. *Nucleic Acids Res.* **2013**, *41*, 9593–9609. (c) Castro, C.; Smidansky, E. D.; Arnold, J. J.; Maksimchuk, K. R.; Moustafa, I.; Uchida, A.; Götte, M.; Konigsberg, W.; Cameron, C. E. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.*, **2009**, *16*, 212–218.
8. (a) Le Vay, K.; Salibi, E.; Song, E. Y.; Mutschler, H. Nucleic Acid Catalysis under Potential Prebiotic Conditions. *Chem Asian J.* **2020**, *15*, 214–230. (b) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid-ion interactions. *Annu Rev Biochem.* **2014**, *83*, 813 (c) Galer, P.; Wang, B.; Plavec, J.; Sket, P. Unveiling the structural mechanism of a G-quadruplex pH-Driven switch. *Biochimie* **2023**, *214*, 73– 78.
9. (a) Chen, K.; Zhao, B. S.; He, C. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem. Biol.*, **2016**, *23*, 74–85. (b) Cui, L., Ma, R., Cai, J., Guo, C., Chen, Z., Yao, L., Wang, Y., Fan, R., Wang, X., & Shi, Y. RNA modifications: importance in immune cell biology and related diseases. *Signal transduct. target. ther.* **2022**, *7*, 334.
10. (a) Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *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. Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. (d) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **2002**, *417*, 876–880. (e) Wachowius, F.; Höbartner, C. Chemical RNA Modifications for Studies of RNA Structure and Dynamics. *ChemBioChem* **2010**, *11*, 469–480.
11. Saliba, A. E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **2014**, *42*, 8845–8860.
12. (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *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. Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018**, *10*, 631−637.
13. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.
14. (a) Stollar, B. D.; Raso, V. Antibodies recognise specific structures of triple-helical polynucleotides built on poly(A) or poly(dA). *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. G-quadruplex DNA:

a novel target for drug design. *Nat. Commun.* **2021**, *12*, 162–172. (d) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011**, *39*, 5768–5775 (e) Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantírek, L.; Dötsch, V. In-Cell NMR and EPR Spectroscopy of Biomacromolecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 10300–10314.
15. Fracchioni, G.; Vailati, S.; Grazioli, M.; Pirota, V. Structural unfolding of G-quadruplexes: From Small Molecules to Antisense Strategies. *Molecules.* **2024**, *29*, 3488.
16. (a) Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893– 5909. (b) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure *Nucleic Acids Res.,* **2006**, *34*, 5402–5415. (c) Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci.* **1962**, *48*, 2013–2018.
17. Williamson, J. R. G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol Struct* **1994**, *23*, 703–730.
18. Jing You, Hui Li, Xi-Ming Lu, Wei Li, Peng-Ye Wang, Shuo-Xing Dou, Xu-Guang Xi; Effects of monovalent cations on folding kinetics of G-quadruplexes. *Biosci Rep* 31 August **2017**; 37 (4): BSR20170771.
19. (a) Raghuraman M.K., Cech T.R. Effect of monovalent cation-induced telomeric DNA structure on the binding of Oxytricha telomeric protein. *Nucleic Acids Res.* **1990**, *18*, 4543–4552. (b) Kankia B.I., Marky L.A. Folding of the thrombin aptamer into a G-quadruplex with Sr^{2+}: Stability, heat, and hydration. *J. Am. Chem. Soc.* **2001**, *123*, 10799–10804. (c) Luo, Y.; Živković, M. L.; Wang, J.; Ryneš, J.; Foldynová-Trantírková, S.; Trantírek, L.; Verga, D.; Mergny, J.-L. A sodium/potassium switch for G4-prone G/C-rich sequences, *Nucleic Acids Res.*, **2024**, *52*, 448–461.
20. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int Ed Engl*. **2004**, *43*, 668–698.
21. Yang D. G-Quadruplex DNA and RNA. *Methods in molecular biology (Clifton, N.J.)* **2019**, *2035*, 1–24.
22. Chen, Y.; Yang, D. Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Protoc. Nucleic Acid Chem.* **2012**, Chapter 17: Unit17.5.
23. Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J*. **2010**, *277*, 1107–1117.
24. (a) Tong, X.; Lan, W.; Zhang, X.; Wu, H.; Liu, M.; Cao, C. Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Res.,* **2011**, *39*, 6753–6763. (b) Parkinson, G. N.; Lee, M. P.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. **2002**, *417*, 876–880. (c) Phan, A. T.; Luu, K. L.; Patel, D. J.; *Nucleic Acids Res,* **2006**,
25. Li, K.; Yatsunyk, L.; Neidle, S. Water spines and networks in G-quadruplex structures, *Nucleic Acids Res.*, **2021**, *49*, 519–528.
26. Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*. **2012**, *13*, 770–780.
27. Paeschke, K.; Capra, J. A.; Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*. **2011,** *145*, 678–691.
28. (a) Sarkies, P.; Reams, C.; Simpson, L. J.; Sale, J. E. Epigenetic instability due to defective replication of structured DNA. *Mol Cell*. **2010**, *40*, 703–713. (b) Sarkies, P.; Murat, P.; Phillips, L. G.; Patel, K. J.; Balasubramanian, S.; Sale, J. E. FANCJ

coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res*. **2012**, *40*, 1485–1498.
29. Huppert, J. L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007,** *35* 406–413 (b) Bahls, B.; Aljnadi, I. M.; Emídio, R.; Mendes, E.; Paulo, A. G-quadruplexes in c-MYC Promoter as Targets for Cancer Therapy. *Biomedicines*. **2023**, *11*, 969. (c) Brooks, T. A.; Hurley, L. H.; Targeting MYC Expression through G-quadruplexes. *Genes Cancer.* **2010**, *1*, 641–649.
30. (a) Dexheimer, T.S.; Sun, D.; Hurley, L.H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–5415. (b) Rankin, S.; Reszka, A.P.; Huppert, J.; Zloh, M.; Parkinson, G.N.; Todd, A.K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589.
31. (a) Ross, J.; Miron, C. E.; Plescia, J, et al. Targeting MYC: From understanding its biology to drug discovery. *Eur J Med Chem*. **2021**, *213*, 113137. (b) Marquevielle, J.; Robert, C.; Lagrabette, O.; Wahid, M.; Bourdoncle, A.; Xodo, L. E.; Mergny, J.-L.; Salgado, G. F. Structure of two G-quadruplexes in equilibrium in the KRAS promoter, *Nucleic Acids Res.*, **2020**, *48*, 9336–9345
32. Tian, T.; Chen, Y.-Q.; Wang, S.-R.; Zhou, X. G-quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem.*, **2018**, *4*, 1314 —1344
33. (a) Dahse, R.; Fiedler, W.; Ernst, G. Telomeres and telomerase: biological and clinical importance. *Clin. Chem.*, **1997**, *43*, 708–714. **(b)** Chan, S. W.; Blackburn, E.H. New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin *Oncogene* **2002**, *21*, 553–563 (c) Neidle, S.; Parkinson, G.N. The structure of telomeric DNA *Curr. Opin. Struct. Biol.* **2003,** *13*, 275–283
34. (a) McClintock, B. The Stability of Broken Ends of Chromosomes in Zea mays. *Genetics*, **1941** *26*, 234–282. (b) Muller, H. J. The Measurement of the Effect of X-rays on the Genetic Material of the Drosophila melanogaster. *Science*, **1938**, *87*, 112–113.
35. (a) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429*, 2127–2147. (b) Lyu, K.; Chow, Eugene Y.-C.; Mou, X.; Chan, T-F.; Kwok, C. K. RNA G-quadruplexes (rG4s): genomics and biological functions, *Nucleic Acids Res.*, **2021**, *49*, 5426–5450,
36. (a) Auffinger, P.; Westhof, E. Water and ion binding around r(UpA)12 and d(TpA)12 oligomers: Comparison with RNA and DNA (CpG)12 duplexes. *J. Mol. Biol.,* **2001**, *305*, 1057–1072. (b) Egli, M.; Portmann, S.; Usman, N. RNA hydration: A detailed look. *Biochemistry* **1996**, *35*, 8489–8494. (c) Zhang, D. H.; Fujimoto, T.; Saxena, S.; Yu, H. Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*. **2010**, *49*, 4554–4563.
37. Arora, A.; Dutkiewicz, M. Scaria, V.; Hariharan, M.; Maiti, S.; Kurreck, J. Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* **2008**, *14*, 1290
38. (a) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nat. Chem. Biol.* An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. **2007**, *3*, 218. (b) Shahid, R.; Bugaut, A.; Balasubramanian, S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **2010**, *49*, 8300.
39. (a) Murat, P.; Marsico, G.; Herdy, B.; Ghanbarian, A. T.; Portella, G.; Balasubramanian, S. *Genome Biol.* RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human

mRNAs. **2018**, *19*, 229. (b) Sauer, M.; Juranek, S. A.; Marks, J.; De Magis, A.; Kazemier, H. G.; Hilbig, D.; Benhalevy, D.; Wang, X.; Hafner, M.; Paeschke, K. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nat. Commun.* 2019, **10**, 2421. (c) Wolfe, A. L.; Singh, K. Y.; Zhong, P.; Drewe, V. K.; Rajasekhar, V. R.; Sanghvi, K. J.; Mavrakis, M.; Jiang, J. E.; Roderick, J.; Van der Meulen, J. H.; Schatz, C. M.; Rodrigo, C.; Zhao, P.; Rondou, E.; de Stanchina, J.; Teruya-Feldstein, M. A.; Kelliher, F.; Speleman, J. A.; Porco Jr., J.; Pelletier, G.; Ratsch, Wendel, H. G. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **2014**, *513*, 65.

40. (a) Chen, Y.; Yang, D. Sequence, stability, structure of G-quadruplexes and their drug interactions. *Curr. Protoc. Nucleic Acid Chem.* **2012**; Chapter 17: Unit17.5. (b) Do, N.Q.; Lim, K.W.; Teo, M. H.; Heddi, B.; Phan, A.T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.,* 2011, 39, 9448–9457.
41. Lech, C. J.; Heddi, B.; Phan, A. T. Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Res.* **2013**, *41*, 2034–2046.
42. Gao, C.; Liu, Z.; Hou, H.; Ding, J.; Chen, X.; Xie, C.; Song, Z.; Hu, Z.; Feng, M.; Mohamed, H. I.; Xu, S.; Parkinson, G. N.; Haider, S.; Wei. D. BMPQ-1 binds selectively to (3+1) hybrid topologies in human telomeric G-quadruplex multimers. *Nucleic Acids Res.,* **2020**, *48*, 11259–11269.
43. Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M.T.; De Pauw, E.; Gabelica. V. G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.,* **2008**, *130*, 10208–10216.
44. Frasson, I.; Pirota, V.; Richter, S. N; Doria, F. Multimeric G-quadruplexes: A review on their biological roles and targeting. *Int J Biol Macromol.* **2022**, *204*, 89–102.
45. Kolesnikova, S.; Curtis, E. A. Structure and function of multimeric G-quadruplexes. *Mol. Basel Switz.,***2019**, *24*, Article E3074.
46. Monsen, R. C. Higher-order G-quadruplexes in promoters are untapped drug targets. *Front Chem*. **2023**, *11*, 1211512.
47. Hassan, Z.; Kumar, N. D.; Reggiori, F.; Khan, G. How Viruses Hijack and Modify the Secretory Transport Pathway. *Cells.* **2021**, *10*, 2535 (b) Gelderblom, H, R. Structure and Classification of Viruses. *In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston;* **1996**. Chapter 41. Available from:
48. Dávila-Ramos, S.; Castelán-Sánchez, H. G.; Martínez-Ávila, L.; Sánchez-Carbente, M. D. R.; Peralta, R.; Hernández-Mendoza, A.; Dobson, A. D. W.; Gonzalez, R. A.; Pastor, N.; Batista-García, R. A. A Review on Viral Metagenomics in Extreme Environments. *Frontiers in microbiology,* **2019**, *10*, 2403.
49. (a) Ellebedy, A. H.; Ahmed, R. Antiviral Vaccines: Challenges and Advances. (2016). *The Vaccine Book*, 283–310. (b)
50. Amrane, S.; Jaubert, C.; Bedrat, A. et al. Deciphering RNA G-quadruplex function during the early steps of HIV-1 infection. *Nucleic Acids Res.* **2022**, *50*, 12328–12343.
51. Neidle S. Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.* **2016**, *59*, 5987–6011.
52. Abiri, A.; Lavigne, M.; Rezaei, M.; Nikzad, S.; Zare, P.; Mergny, J.-L.; Rahimi, H.-R. Unlocking G-quadruplexes as antiviral targets. *Pharmacol. Rev.* **2021**, *73*, 897–923.
53. (a) Wang, Y.-H.; Yang, Q.-F.; Lin, X. et al. G4LDB 2.2: a database for discovering and studying G-quadruplex and i-motif ligands. *Nucleic Acids Res.* **2021**, *50*, D150–D160. (b) Luo J., Wei W., Waldispühl J., Moitessier N. Challenges and current status of

computational methods for docking small molecules to nucleic acids. *Eur. J. Med. Chem.* **2019**, *168*, 414–425.
54. Ruggiero, E.; Richter, S. N. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* **2020**, *54*, 101–131.
55. (a) Jaubert, C.; Bedrat, A.; Bartolucci, L.; Di Primo, C.; Ventura, M.; Mergny, J.-L.; Amrane, S.; Andreola, M.-L. RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep.* **2018**, *8*, 8120 (b) Ruggiero, E.; Richter S. N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.* **2018**, *46*, 3270–3283.
56. Kosiol, N.; Juranek, S.; Brossart, P. et al. G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* **2021**, *20*, 40–57.
57. Asamitsu, S.; Bando, T.; Sugiyama, H. *Chem. Eur. J.* **2019**, *25*, 417–430.
58. Iachettini, S.; Biroccio, A.; Zizza, P. Therapeutic Use of G4-Ligands in Cancer: State-of-the-Art and Future Perspectives. *Pharmaceuticals (Basel).* **2024**, *17*, 771.
59. Miles, A. J.; Janes, R. W.; Wallace, B. A. Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chem Soc Rev.* **2021**, *7*, 8400–8413
60. Vorlíčková, M.; Kejnovská, I.; Bednářová, K.; Renčiuk, D.; Kypr, J. Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. *Chirality* **2012**, *24*, 691–698.
61. Cappannini, A.; Mosca, K.; Mukherjee, S. et al. NACDDB: Nucleic Acid Circular Dichroism Database. *Nucleic Acids Res*. **2023**, *51*(D1):D226–D231.
62. (a) Carvalho, J.; Queiroz, J. A.; Cruz, C. Circular dichroism of G-Quadruplex: A laboratory experiment for the study of topology and ligand binding. *J. Chem. Educ.* **2017**, *94*, 1547–1551. (b) Villar-Guerra, R.D.; Trent, J.O.; Chaires, J.B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed*. **2018**, *57*, 7171–7175.
63. Garbett, N.C.; Ragazzon, P.A.; Chaires, J.O.B. Circular dichroism to determine binding mode and affinity of ligand-dna interactions. *Nat. Protoc.* **2007**, *2*, 3166–3172.
64. Šmidlehner, T.; Piantanida, I.; Pescitelli, G. Polarization spectroscopy methods in the determination of interactions of small molecules with nucleic acids-Tutorial. *Beilstein J. Org. Chem.* **2017**, *14*, 84–105.
65. Kong, D. M.; Ma, Y. E.; Wu, J.; Shen, H. X. Discrimination of G-quadruplexes from duplex and single-stranded DNAs with fluorescence and energy-transfer fluorescence spectra of crystal violet. *Chemistry*. **2009**, 15, 901–909.
66. El-Yazbi.; Amira, F.; G. R. Loppnow. Detecting UV-induced nucleic-acid damage. *Trends in Analytical Chemistry,* **2014**, *61*, 83–91.
67. (a) Jr. Tinoco, I. Hypochromism in Polynucleotides *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.
68. (a) Luo, Y.; Granzhan, A.; Marquevielle, J. et al. Guidelines for G-quadruplexes: I. *in vitro* characterization. *Biochimie*. **2023,** 214(Pt A):5-23. (b) Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **2005**, *33*, e138–e144. (c) Evans, L.; Kotar, A.; Valentini, M.; Filloux, A.; Jamshidi, S.; Plavec, J.; Rahman, K. M.; Vilar, R. Identification and characterisation of G-quadruplex DNA-forming sequences in the Pseudomonas aeruginosa genome. *RSC Chem. Biol.* **2023**, *4*, 94–100
69. (a) Wu, G.; Han, H. A DNA Polymerase Stop Assay for Characterization of G-Quadruplex Formation and Identification of G-Quadruplex-Interactive Compounds. In: Yang, D., Lin, C. (eds) *G-Quadruplex Nucleic Acids. Methods in Molecular Biology*, **2019**, *2035*. Humana, New York, NY. (b) Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Hurley, L. H. Inhibition of human telomerase by a G-

quadruplex-interactive compound. *J. Med. Chem.,* **1997**, *40*, 2113–2116. (b) Woodford, K. J.; Howell, R. M.; Usdin, K. A novel K^+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.* **1994**, *269*, 27029–27035. (c) Han, H., Hurley, L. H.; Salazar, M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.* **1999**, *27*, 537–542.

70. (a) Kwok, C. K.; Balasubramanian, S. Targeted detection of G-quadruplexes in cellular RNAs. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 6751–6754 (b) Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **2015**, *33*, 877–881. (c) Guo, J. U.; Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **2016**, *353*, 5371.
71. (a) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **2003**, *4*, 936–962. (b) Adrian, M.; Heddi, B.; Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods* **2012**, *57*, 11–24. (c) Yamaoki, Y.; Nagata, T.; Sakamoto, T.; Katahira, M. Recent progress of in-cell NMR of nucleic acids in living human cells. *Biophys. Rev.* **2020**, *12*, 411–417. (d) Dayie, T. K.; Olenginski, L. T.; Taiwo, K. M. Isotope Labels Combined with Solution NMR Spectroscopy Make Visible the Invisible Conformations of Small-to-Large RNAs. *Chem. Rev.* **2022**, *122*, 10, 9357–9394.
72. (a) Feigon, J.; Koshlap, K.; Smith, M. F. W. ^{1}H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **1995**, *261*, 225–255. (b) Phan, A. T.; Mergny, J.-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **2002**, *30*, 4618–4625. (c) Duszczyk, M. M.; Zanier, K.; Sattler, M. A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. *Nucleic Acids Res.* **2008**, 36, 7068–7077.
73. Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **2005**, *33*, e138–e144.
74. Hansel, R.; Foldynova-Trantirkova, S.; Lohr, F.; Buck, J.; Bongartz, E.; Bamberg, E.; Schwalbe, H.; Dotsch, V.; Trantirek, L. *J. Am. Chem. Soc.,* **2009**, *131*, 15761-15768
75. Salgado GF, Cazenave C, Kerkour A, Mergny JL. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem Sci.* **2015**, *6*, 3314–3320.
76. (a) Bao, H. L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of human telomere RNA G-quadruplex structures *in vitro* and in living cells. *Nucleic Acids Res.* **2017**, 45, 5501–5511.
77. (a) Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions *Nucleic Acids Res.* **2011**, *39*, 5768–5775. (b) Hänsel, R.; Löhr, F.; Trantírek, L.; Dötsch, V. High-resolution insight into G-overhang architecture. *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.
78. 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.
79. (a) Wang, C.; Xu, G.; Liu, X. et al. ^{19}F Nuclear Magnetic Resonance Fingerprinting Technique for Identifying and Quantifying G-Quadruplex Topology in Human Telomeric Overhangs. *J Am Chem Soc*. **2024**, *146*, 4741–4751. (b) Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. ^{19}F NMR: a valuable tool for studying biological events. *Chem. Soc. Rev.* **2013**, *42*, 7971−7982. (b) Himmelstoß, M.; Erharter, K.; Renard, E.; Ennifar, E.; Kreutz, C.; Micura, R. 2′-O-Trifluoromethylated RNA – a powerful modification for RNA chemistry and NMR spectroscopy. *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. Locked 2′-Deoxy-2′,4′-Difluororibo Modified Nucleic Acids: Thermal Stability, Structural Studies, and siRNA Activity. *J. Am. Chem. Soc.* **2020**, *142*, 4739−4748. (d) Baranowski, M. R.; Warminski, M.; Jemielity, J.; Kowalska, J. 5'-fluoro(di)phosphate-labeled oligonucleotides are versatile molecular probes for studying nucleic acid secondary structure and interactions by ^{19}F NMR. *Nucleic Acids Res.* **2020**, *48*, 8209−8224. (e) Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. Synthesis of nucleoside mono- and triphosphates bearing oligopyridine ligands, their incorporation into DNA and complexation with transition metals. *J. Org. Chem.* **2012**, *77*, 1026−1044. (f) Sakamoto, T.; Hasegawa, D.; Fujimoto, K. Simultaneous Amino Acid Analysis Based on ^{19}F NMR Using a Modified OPA-Derivatization Method. *Chem. Commun.* **2015**, *51*, 8749−8752.

80. (a) Manna, S.; Sarkar, D.; Srivatsan, S. G. A Dual-App Nucleoside Probe Provides Structural Insights into the Human Telomeric Overhang in Live Cells. *J. Am. Chem. Soc.* **2018**, *140*, 12622−12633. (b) Manna, S.; Sontakke, V. A.; Srivatsan, S. G. Incorporation and Utility of a Responsive Ribonucleoside Analogue in Probing the Conformation of a Viral RNA Motif by Fluorescence and ^{19} F NMR Spectroscopy. *ChemBioChem* **2022**, *23*, e202100601.
81. (a) 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**,*14*, 5627−5637 (b) Khatik, S. Y.; Srivatsan, S. G. Environment-Sensitive Nucleoside Probe Unravels the Complex Structural Dynamics of i-Motif DNAs. *Bioconjug Chem.* **2022**, *33*, 1515−1526. (c) Khatik, S. Y.; Roy, S.; Srivatsan, S. G. Synthesis and Enzymatic Incorporation of a Dual-App Nucleotide Probe That Reports Antibiotics-Induced Conformational Change in the Bacterial Ribosomal Decoding Site RNA. *ACS Chem Biol*. **2024**, *19*, 687−695.
82. (a) Ishizuka, T.; Yamashita, A.; Asada, Y.; Xu, Y. Studying DNA G-Quadruplex Aptamer by ^{19}F NMR. *ACS Omega* **2017**, *2*, 8843–8848. (b) Ishizuka, T.; Zhao, P.-Y.; Bao, H.-L.; Xu, Y. A multi-functional guanine derivative for studying the DNA G-quadruplex structure. *Analyst* **2017**, *142*, 4083–4088.
83. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of human telomere RNA G-quadruplex structures *in vitro* and in living cells using 19F NMR spectroscopy. *Nucleic Acids Res.* **2017**, *45*, 5501–5511. (d) Bao, H.-L.; Liu, H.-S.; Xu, Y. Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.* **2019**, *47*, 4940−4947.
84. Bao, H.-L.; Liu, H.-S.; Xu, Y. Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.* **2019**, *47*, 4940−4947.
85. (a) Qin, P. Z.; Dieckmann, T. Application of NMR and EPR methods to the study of RNA. *Curr. Opin. Struc. Biol.* **2004**, *14*, 350–359. (b) Wachowius, F.; Höbartner, C. Chemical RNA Modifications for Studies of RNA Structure and Dynamics *ChemBioChem* **2010**, *11*, 469–480.
86. Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M. Pulsed Electron Double Resonance in Structural Studies of Spin-Labeled Nucleic Acids. *Angew. Chem. Int. Ed.* **2009**, *48*, 9728–9730.
87. (a) Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantírek, L.; Dötsch, V. In-Cell NMR and EPR Spectroscopy of Biomacromolecules. *Angew. Chem. Int. Ed.* **2014**, *53*, 10300–10314. (b) Manna, S.; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model by Using Responsive Fluorescent Nucleoside Probes. *Chembiochem*. **2017**, *17*, 1604-1615.
88. (a) Stratmann, L. M.; Kutin, Y.; Kasanmascheff, M.; Clever, G. H. Precise Distance Measurements in DNA G-Quadruplex Dimers and Sandwich Complexes by Pulsed

Dipolar EPR Spectroscopy. *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. Conformations of Human Telomeric G-Quadruplex Studied Using a Nucleotide-Independent Nitroxide Label. *Biochemistry* **2016**, *55*, 360–372. (c) Schiemann, O., & Prisner, T. F. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q. Rev. Biophys*, **2007**, *40*, 1–53. (d) Kang, C., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature*, **1992**, *356*, 126–131.

89. (a) Mooers, B. H. Crystallographic studies of DNA and RNA. Methods **2009**, *47*, 168–176. (b) Pike, A. C. W.; Garman, E. F.; Krojer, T.; Delft, F.; Carpenter, E. P.; Simplified heavy-atom derivatization of protein structures via co-crystallization with the MAD tetragon tetrabromoterephthalic acid. Acta. Crystallogr. D. *Struct. Biol.* 2016, 72, 303–318.
90. (a) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *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. The crystal structure of an antiparallel chair-type G-quadruplex formed by Bromo-substituted human telomeric DNA. *Nucleic Acids Res.* **2019**, *47*, 5395. (c) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2012**, *40*, 4691–700. 48.
91. (a) Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z. Synthesis of a 2′-Se-Uridine Phosphoramidite and Its Incorporation into Oligonucleotides for Structural Study. *J. Am. Chem. Soc.* **2002**, *124*, 24−25.
92. (a) Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M. Selenium-Assisted Nucleic Acid Crystallography: Use of Phosphoroselenoates for MAD Phasing of a DNA Structure. *J. Am. Chem. Soc.* **2002**, *124*, 14910–14916.
93. Dauter, M.; Dauter, Z. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Methods Mol. Biol.* **2017**, *1607*, 349–356.
94. Heddi, B.; Cheong, V. V.; Schmitt, E.; Mechulam, Y.; Phan, A. T. Recognition of different base tetrads by RHAU (DHX36): X-ray crystal structure of the G4 recognition motif bound to the 3'-end tetrad of a DNA G-quadruplex. *J Struct Biol.* **2020**, *209*, 107399.
95. Terrell, J. R.; Le, T. T.; Paul, A. et al. Structure of an RNA G-quadruplex from the West Nile virus genome. *Nat Commun.,* **2024**, *15*, 5428.
96. Parkinson, G. N.; Collie, G. W. X-ray Crystallographic Studies of G-Quadruplex Structures. *In Methods in Molecular Biology; Springer: Berlin, Germany*, **2019**; *2035*, 131–155.
97. Clark, G.R.; Pytel, P.D.; Squire, C.J.; Neidle, S. Structure of the first parallel DNA quadruplex-drug complex. *J. Am. Chem. Soc.* **2003**, *125*, 4066–4067.
98. Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Höbartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R.; et al. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat. Struct. Mol. Biol.* **2005**, *12*, 218−224.
99. Höbartner, C.; Rieder, R.; Kreutz, C.; Puffer, B.; Lang, K.; Polonskaia, A.; Serganov, A.; Micura, R. *J. Am. Chem. Soc.* 2005, 127, 12035−12045.
100. Bazzicalupi, C.; Ferraroni, M.; Bilia, A.R.; Scheggi, F.; Gratteri, P. The crystal structure of human telomeric DNA complexed with berberine: An interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res*. **2013**, *41*, 632–638.
101. (a) Kruger, A. C.; Birkedal, V. Single molecule FRET data analysis procedures for FRET efficiency determination: Probing the conformations of nucleic acid structures.

Methods **2013**, *64*, 36–42. (b) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18938–18943.

102. Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 14629–14634.
103. Juskowiak, B. Nucleic acid-based fluorescent probes and their analytical potential. *Anal Bioanal Chem.* **2011**, 399, 3157–3176.
104. (a) Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals (Basel).* **2021**, *14*, 769. (b) Wang, L.; Gaigalas, A. K.; Blasic, J.; Holden, M. J.; Gallagher, D. T.; Pires, R. Fluorescence resonance energy transfer between donor-acceptor pair on two oligonucleotides hybridized adjacently to DNA template. *Biopolymers*. **2003**, *72*, 401–412. (c) Sidhu, P. K.; Chaires, J. B.; Mergny, J. L. Fluorescent probes for G-quadruplex structures: A review. *Biochimie*, **2021**, *185*, 178–195. (d) Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA of the proto-oncogene c-MYC. *J. Am. Chem. Soc.,* **2007,** *129*, 7484–7485. (e) Pond, S. J. K.; Tsukamoto, T.; Mack, K. W.; Shank, N. I. Cy3 and Cy5 dyes attached to G-quadruplexes modulate structural dynamics and ligand interactions. *Nucleic Acids Res.,* 2018, *46*, 2315–2325. (f) Mishra, A.; Panda, D.; Suresh Kumar, G. Fluorescent labeling strategies for studying nucleic acid structures and interactions: Challenges and recent advances. *Chem. Soc. Rev.,* **2022,** *51,* 3546–3575.
105. Dziuba, D.; Didier, P.; Ciaco, S.; Barth, A.; Seidel, C. A. M.; Mély, Y. Fundamental photophysics of isomorphic and expanded fluorescent nucleoside analogues. *Chem. Soc. Rev.* **2021***, 50,* 7062−7107.
106. T. Kimura, K. Kawai, M. Fujitsuka and T. Majima , Fluorescence properties of 2-aminopurine in human telomeric DNA. ***C****hem. Commun.,* **2004**, 1438–1439.
107. Kirk, S. R.; Luedtke, N. W.; Tor, Y. 2-Aminopurine as a Real-Time Probe of Enzymatic Cleavage and Inhibition of Hammerhead Ribozymes. *Bioorg. Med. Chem.* **2001**, *9*, 2295– 2301.
108. (a) Xu Y.; Sugiyama, H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb) *Nucleic Acids Res.*, **2006**, *34*, 949 —954). (b) Johnson, J.; Okyere, R.; Joseph, A.; Musier-Forsyth, K.; Kankia, B. Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res.,* **2013**, *41*, 220–228.
109. Srivatsan, S. G.; Greco, N. J.; Tor, Y. A highly emissive fluorescent nucleoside that signals the activity of toxic ribosome-inactivating proteins. *Angew. Chem. Int. Ed.* **2008**, *47*, 6661–6665.
110. Saitoa, Y.; Hudson, R. H. E. J. Base-modified fluorescent purine nucleosides and nucleotides for use in oligonucleotide probes. *Photochem. Photobiol. C.* **2018**, *36*, 48–73.
111. Schmidt, O. P.; Benz, A. S.; Mata, G; Luedtke, N. W. Hg^{II} binds to C–T mismatches with high affinity. *Nucleic Acids Res.* **2018**, *46*, 6470–6479.
112. (a) Preus, S.; Kilså, K.; Wilhelmssonb, L. M.; Albinsson, B. Photophysical and structural properties of the fluorescent nucleobase analogues of the tricyclic cytosine (tC) family. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8881–8892. (b) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B.

Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res.* **2008**, *36*, 157–167.
113. Börjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems. *J. Am. Chem. Soc.* **2009**, *131*, 4288–4293.
114. (a) Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. Fluorescent Nucleoside Analogs: Probes for Investigating Nucleic Acid Structure and Function. *Isr. J. Chem.* **2013**, *53*, 366–378. (b) Manna, S.; Srivatsan, S. G. Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.* **2018**, *8*, 25673–25694.
115. (a) Greco, N. J.; Tor, Y. Simple Fluorescent Pyrimidine Analogues Detect the Presence of DNA Abasic Sites. *J. Am. Chem. Soc.* **2005**, *127*, 10784–10785. (b) Srivatsan, S. G.; Tor, Y. Fluorescent pyrimidine ribonucleotide: synthesis, enzymatic incorporation, and utilization. *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053. (c) Srivatsan, S. G.; Tor, Y. Using an emissive uridine analogue for assembling fluorescent HIV-1 TAR constructs. *Tetrahedron* **2007**, *63*, 3601–3607.
116. (a) Pawar, M. G.; Srivatsan, S. G. Synthesis, photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog *Org. Lett.* **2011**, *13*, 1114–1117. (b) Pawar, M. G.; Srivatsan S. G. Probing the competition between duplex and G-quadruplex/i-motif structures using a conformation-sensitive fluorescent nucleoside probe. *J. Phys. Chem. B.* **2013**, *117*, 14273–14282. (c) Tanpure, A. A.; Srivatsan, S. G. A Microenvironment-Sensitive Fluorescent Pyrimidine Ribonucleoside Analogue: Synthesis, Enzymatic Incorporation, and Fluorescence Detection of a DNA Abasic Site. *Chem. Eur. J.* **2011**, *17*, 12820–12827. (d) Synthesis and Photophysical Characterisation of a Fluorescent Nucleoside Analogue that Signals the Presence of an Abasic Site in RNA. Tanpure, A. A.; Srivatsan, S. G. *Chembiochem.* **2012**, *13*, 2392–9156. (e) Tanpure, A. A.; Srivatsan, S. G. Conformation-sensitive nucleoside analogues as topology-specific fluorescence turn-on probes for DNA and RNA G-quadruplexes. *Nucleic Acids Res.* **2015**, *43*, e149. (f) Sabale, P. M.; Tanpure, A. A.; Srivatsan, S. G. Probing the competition between duplex and G-quadruplex/i-motif structures using a conformation-sensitive fluorescent nucleoside probe. *Org. Biomol. Chem.* **2018**, *16*, 4141–4150.
117. (a) Pandith, A.; Luo, Y.; Jang, Y.; Bae, J.; Kim, Y. Self-Assembled Peptidyl Aggregates for the Fluorogenic Recognition of Mitochondrial DNA G-Quadruplexes. *Angew. Chem., Int. Ed.* **2023**, *62*, e202215049. (b) Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **2014**, *10*, 686−691. (c) Zhang, Z.-H.; Qian, S. H.; Wei, D. G.; Chen, Z.-X. In vivo dynamics and regulation of DNA G-quadruplex structures in mammals. *Cell Biosci.* **2023**, *13*, 117. (d) Shivalingam, A.; Izquierdo, M. A.; Marois, A. L.; Vysniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. *Nat. Commun.* **2015**, *6*, 8178.
118. (a) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an Efficient Inducer and Selective Fluorescent Sensor for the Human Telomeric G-Quadruplex DNA. *J. Am. Chem. Soc.* **2013**, *135*, 367−376. (b) Di Antonio, M.; Ponjavic, A.; Radzevicius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L. M.; Lee, S. F.; Klenerman, D. et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **2020**, *12*, 832−837. (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. Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **2021**, *12*, 162. (d) Tang, G.;

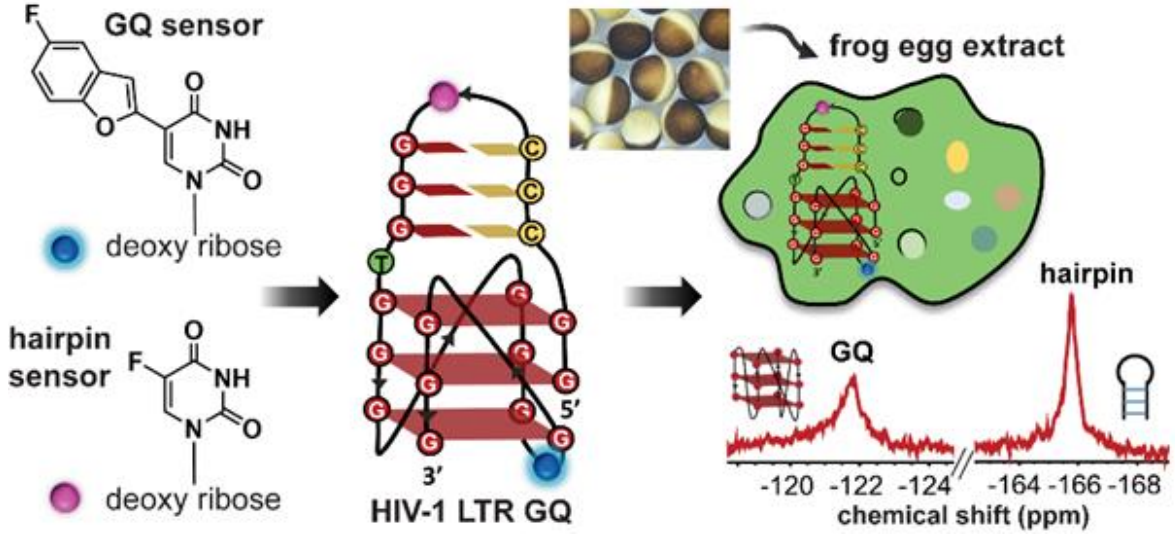
Wang, X.; Huang, H. et al. Small Molecule-Induced Post-Translational Acetylation of Catalytic Lysine of Kinases in Mammalian Cells. *J Am Chem Soc.* **2024**, *146*, 23978−23988. (e) Liu, L. Y.; Ma, T. Z.; Zeng, Y. L.; Liu, W.; Zhang, H.; Mao, Z. W. Organic-Platinum Hybrids for Covalent Binding of G-Quadruplexes: Structural Basis and Application to Cancer Immunotherapy. *Angew. Chem., Int. Ed.* **2023**, *62*, e202305645.

119. (a) Bhasikuttan, A. C.; Mohanty, J. Targeting G-quadruplex structures with extrinsic fluorogenic dyes: promising fluorescence sensors. *Chem. Commun*. **2015**, *51*, 7581−7597. (b) Suseela, V.; Narayanaswamy, N.; Pratihar, S.; Govindaraju, T. Far-red fluorescent probes for canonical and non-canonical nucleic acid structures: current progress and future implications. *Chem. Soc. Rev.* **2018**, *47*, 1098−1131. (c) Yuan, J.-H.; Shao, W.; Chen, S.-B.; Huang, Z.-S.; Tan, J.-H. Recent advances in fluorescent probes for G-quadruplex nucleic acids. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 18−24.
120. (a) Zhang, L.; Liu, X.; Lu, S.; Liu, J.; Zhong, S.; Wei, Y.; Bing, T.; Zhang, N.; Shangguan, D. Thiazole Orange Styryl Derivatives as Fluorescent Probes for G-Quadruplex DNA. *ACS Appl. Bio. Mater.* **2020**, *3*, 2643–2650. (b) Rodriguez, M.S.;. Yeoman, J. A.; Trentesaux, C.; Riou, J.F.; Balasubramanian, S. A novel small molecule that alters shelterin integrity and triggers a DNA damage response at telomeres, *J. Am. Chem. Soc*., **2008**, *130*,15785. (c) Wu, F.; Liu, C.; Chen, Y.; Yang, S.; Xu, J.; Huang, R.; Wang, X.; Li, M.; Liu, W.; Mao, W.; Zhou, X.; Visualization of G-quadruplexes in gel and in live cells by a near-infrared fluorescent probe, *Sens. Actuators, B* , **2016**, *236*.
121. (a) 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. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem*. **2020**, *12*, 832–837. (b) Chen, S.-B.; Hu, M.-H.; Liu, G.-C.; Wang, J.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Tan, J.-H. Visualization of NRAS RNA G-Quadruplex Structures in Cells with an Engineered Fluorogenic Hybridization Probe *J. Am. Chem. Soc.* **2016**, *138*, 10382−10385. (c) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. Fluorescence properties of 2-aminopurine in human telomeric DNA. *Chem. Commun.* **2004**, 1438–1439. (d) 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. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **2020**, *12*, 832–837. (e) Shivalingam, A.; Izquierdo, M.A.; Marois, A.L.; Vysniauskas, A.; Suhling, K.; Kuimova, M.K.; Vilar, R. The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy, *Nat. Commun.* **2015**, *6,* 8178.
122. Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G. Pirrotta, M.; Perrier-Cornet, J. M.; Wong, J. M.; Monchaud, D. Visualization of RNA-quadruplexes in live cells, *J. Am. Chem. Soc.* **2015**, *137*, 8521–85.
123. Chen, S. B.; Hu, M. H.; Liu, G. C.; Wang, J.; Ou, T. M.; Gu, L. Q.; Huang, Z. S.; Tan, J. H. Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe, *J. Am. Chem. Soc.* **2016**, *138*, 10382–10385.
124. Lam, E. Y.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **2013,** *4*, 1796–1780.
125. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R., Nikan, M.; Britton S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S. Jackson S.P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol*. **2012**, *8,* 301–310.

126. (a) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Jr, Sen, D. Lansdorp, P. M. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.,* **2014**, *42*, 860–869.
127. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H.J.; Plückthun, A. *In Vitro* Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with Stylonychia lemnae Macronuclei. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8572–8577.
128. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere End-Binding Proteins Control the Formation of G-Quadruplex DNA Structures In Vivo. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–854.
129. (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5*, 182–186. (b) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, *6*, 75–80. (c) 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. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **2016**, *23*, 1261–1270. (d) Henderson, A.; Wu, Y.; Huang, Y.C.; Chavez, E.A.; Platt, J.; Johnson, F.B.; Brosh, R.M.; Sen, D.; Lansdorp, P. M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Res.* **2014**, *42*, 860–869.
130. (a) Li, C.; Wang, H.; Yin, Z.; Fang, P.; Xiao, R.; Xiang, Y.; Wang, W.; Li, Q.; Huang, B.; Huang, J.; et al. Ligand-Induced Native G-Quadruplex Stabilization Impairs Transcription Initiation. *Genome Res.* **2021**, *31*, 1546–1560. (b) Lyu, J.; Shao, R.; Kwong Yung, P.Y.; Elsässer, S.J. Genome-Wide Mapping of G-Quadruplex Structures with CUT&Tag. *Nucleic Acids Res.* **2022**, *50*, e13.

Chapter 2

Structural elucidation of HIV-1 LTR G-quadruplexes in cellular environment with nucleoside analogs



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

Roy, S.; Majee, P.; Sudhkar, S.; Misra, S.; Kalia, J; Pradeepkumar, P. I.; Srivatsan, S. G. *Chem Sci.,* **2024**, *15,* 7982–7991.

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

2.1 Introduction

HIV-1 is one of the most lethal retroviruses and is responsible for the immunocompromised disease known as AIDS. The virus induces a chronic infection by etching the host cell DNA genome with a proviral DNA that is reverse transcribed from its RNA genome. Established treatments use a cocktail of drugs having different modes of action to control the disease progression.[1,2] However, due to persistence of latent reservoirs, drug-resistance and promiscuity of the viral polymerase, it is very hard to eradicate the virus completely from the host system.[3,4] One of the current ways to counter latency involves awakening the dormant viruses and simultaneously inhibiting viral replication with antiviral agents.[5–7] It is also hypothesized that targeting certain structural and functional segments of the integrated viral DNA genome could complement the above strategy and help in curing the disease.[8] An important and a highly conserved gene segment that could be suitable for this purpose is the long terminal repeat (LTR) of the HIV-1 promoter region.[9,10]

The HIV-1 LTR drives the transcription of viral genes to produce the full-length genome. The initiation of transcription is navigated by the promoter region 5′-LTR, which is composed of U3, R and U5 regions.[11] The U3 region consists of three functional segments including the highly conserved core-binding site of NF-κB and Sp1 transcription factors[12], which harbors contiguous G-rich tracts capable of forming G-quadruplex (GQ) structures namely LTR-II, LTR-III and LTR-IV (Figure 1A).[13] A part of this G-rich region has also been shown to form *in vitro* a two-tetrad antiparallel and a three-tetrad hybrid GQ structure called HIVpro1 and HIVpro2.[14,15] Notably, LTR-III and LTR-IV form GQs *in vitro* in a mutually exclusive manner. LTR-III adopts a unique architecture made of a hybrid-type GQ juxtaposed with a three G-C paired hairpin motif (Figure 1B).[6] On the other hand, LTR-IV attains a parallel GQ topology.[17] However, the entire G-rich region majorly forms the GQ-hairpin form like the LTR-III motif and the parallel form of LTR-IV is induced when it binds to ligands or protein factors.[11,13] Functionally, these structures are orthogonal wherein LTR-III GQ represses the transcription activity, whereas the LTR-IV GQ enhances the activity.[17] Several endogenous protein binders have been identified that modulate the viral transcription by either stabilizing (e.g., nucleolin and liposarcoma) or destabilizing (e.g., hnRNP A2/B1 and NCp7) the LTR GQs.[18–21] Together, these studies reveal certain important aspects of LTR GQ-forming motifs that could be exploited in devising novel therapeutic strategies. The LTR GQ region represent an evolutionary conserved element across all primate lentivirus and the balance between different GQ structures is implicated in the propagation and latency of the virus.[22] As

these GQs adopt unique architectures and show opposite functional effects, we envision that the virus status can be selectively controlled in the host cell by using structure-specific binders. In this direction, it is important to gain a comprehensive understanding of the structural polymorphism, dynamics and druggable space of LTR GQs in cellular environment to advocate a viable screening platform.

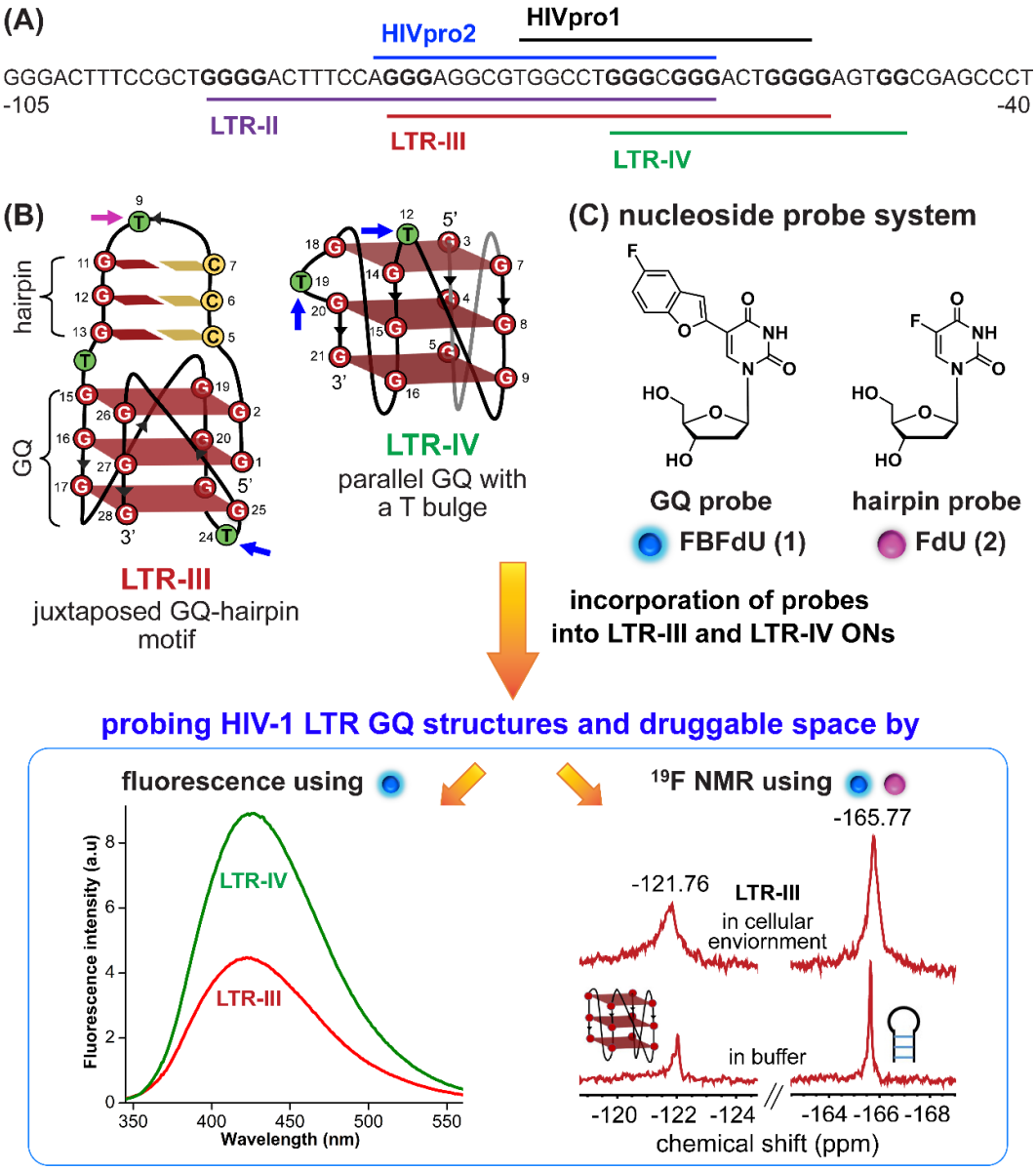


Figure 1. (**A**) Schematic representation of the HIV-1 LTR G-rich region that forms different GQ structures. (**B**) Secondary structures of LTR-III and LTR-IV GQs are depicted using respective NMR structures (PDB: 6H1K and 2N4Y). (**C**) Environment-sensitive nucleoside system designed to probe the structural polymorphism and druggable space of LTR GQs by fluorescence and ^{19}F NMR techniques in cell-free and cellular environments. Potential sites for incorporation of GQ probe (FBFdU **1**) and hairpin probe (FdU **2**) into the loop region of LTR-III and LTR-IV G-rich sequences are shown in blue and magenta arrows, respectively.

Multitude of biophysical and biochemical techniques including CD, UV thermal melting, fluorescence, NMR and X-ray crystallography are used for characterizing GQs.[23–30] Every technique adds a piece of valuable information, which helps in defining the topology, stability and recognition properties of a G-rich motif *in vitro.*[31] Further, antibodies[32–35] and chemical probes have been developed to detect ensembles of DNA and RNA GQs in cells.[36–39] However, majority of tools often fall short when evaluating co-existing structures, as the challenges are twofold. One, except for a very few,[40–45] most tools do not efficiently differentiate GQ topologies. Two, most tools that provide valuable information *in vitro* cannot be used to assess structures in cellular environment. The problem is compounded by the fact that sequences with multiple G stretches exhibit a high degree of structural polymorphism and dynamics,[46,47] which can vary in cell-free and cellular environments. Therefore, we sought to devise a probe platform that would provide spectral signatures for different GQ structures formed by the LTR G-rich region, which then could be used to determine the preferred GQ structure in cells.

We develop microenvironment-sensitive dual-functional nucleoside probes that immensely aid in studying nucleic acid conformations and topology-specific ligand/drug interaction.[48,49] In particular, 5-fluorobenzofuran-modified 2′-deoxyuridine (FBFdU) serves as an excellent two-channel readout system to detect different GQ conformations of the human telomeric repeat *in vitro* and in cellular milieu by using fluorescence and ^{19}F NMR techniques.[48] We envisaged that this nucleoside probe could be used to study the complexity and dynamics of LTR GQ structures. However, one of the GQs, namely LTR-III forms a scaffold made of a juxtaposed GQ-hairpin motif that is unique for the HIV-1 virus. Hence, we realized that by using a single GQ sensing probe it would not be possible to survey the landscape of LTR. In this regard, here we report the development of a probe system that uses FBFdU as a GQ sensor and 5-fluoro-2′-deoxyuridine (FdU) as a hairpin sensor (Figure 1C). The probes judiciously placed in loop positions of the LTR G-rich region are minimally perturbing, and importantly, produce distinct and resolved spectral signatures for the GQ-hairpin structure formed by LTR-III and a parallel-stranded GQ structure formed by LTR-IV. Rewardingly, we deduced the GQ structure adopted by the LTR promoter region in an *ex vivo* model (*Xenopus laevis* oocytes extract) by using ^{19}F NMR signatures obtained *in vitro.*

2.2 Results and Discussion

2.2.1 Design of nucleoside probe platform

The loop composition, orientation and loop residue interaction with neighboring bases are different amongst the GQs.[51,52] Environment-sensitive nucleoside analogs capable of sensing these differences act as good GQ probes.[53,54] Therefore, to configure a probe system, we decided to exploit the differences in the structural features of LTR GQs, particularly at the loop nucleoside level. In case of HIV-1 LTR, the functionally important LTR-III and IV regions form two GQs with unique topological features. The 28-mer LTR-III ON (**3**) adopts a hybrid-type parallel-antiparallel GQ structure comprising of three stacked tetrads connected by four loops (Figure 1B, Table 1). The 12-nucleotide diagonal loop (3–14 residues) forms a hairpin motif wherein three G-C base pairs are capped by a loop formed by G_8-T_9-G_{10} residues. This hairpin juxtaposed with the GQ is an unusual architecture.[16] We envisioned that T_9, part of the hairpin loop and T_{24}, part of a 3-nt lateral loop (A_{22}-C_{23}-T_{24}) connecting the G-tetrads would be potential sites for placing nucleoside probes to distinguish hairpin and GQ motifs. On the other hand, LTR-IV ON **8** adopts an all-parallel stranded GQ structure in K^+ solution, which includes a bulge containing T_{19} that stacks with A_{17} of the nearby 1-nt propeller loop (Figure 1B, Table 1). It is shown that T_{19} is not mandatory for GQ formation, and hence, placing a GQ sensor at this position should not affect the native GQ structure. Further, a 4-nt propeller loop formed by A_{10}-C_{11}-T_{12}-G_{13} bases is also envisioned as a good location to place the probe.[17] Based on this key information, we used a combination of two highly conformation-sensitive nucleosides probes, FBFdU (GQ sensor)[48] and FdU (hairpin/duplex sensor)[55,56] to distinguish different topologies adopted by the LTR G-rich region (Figure 1B and 1C). A foreseeable advantage of this probe combination is that the chemical shift region of FBFdU (around -122 ppm) and FdU (around -165 ppm) are significantly different so that the individual topologies can be unequivocally distinguished, which otherwise is difficult by other currently available tools (*vide infra*).

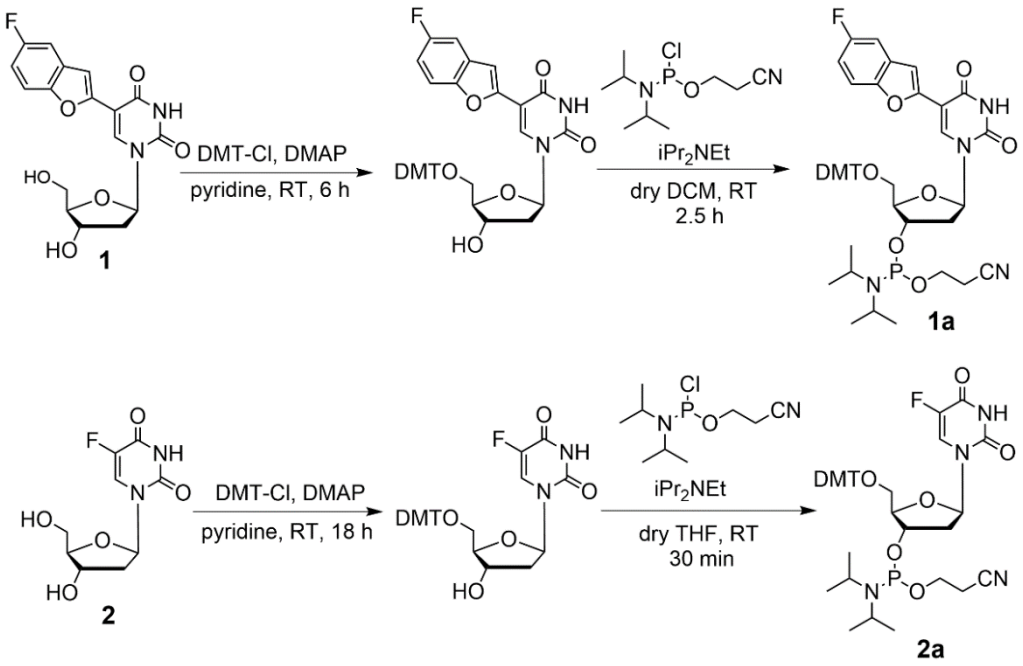
Table 1. Native and labeled LTR G-rich ONs.

G-rich domain	ON[a]	5′----------3′
LTR-III	**3**	GGG AGG CGT GGC CTG GGC GGG ACT GGG G
	4	GGG AGG CG**2** GGC CTG GGC GGG AC**1** GGG G
	5	GGG AGG CGT GGC CTG GGC GGG AC**1** GGG G
	6	GGG AGG CG**2** GGC CTG GGC GGG ACT GGG G
	7	CCC CAG TCC CGC CCA GGC CAC GCC TCC C
LTR-IV	**8**	CTG GGC GGG ACT GGG GAG TGG T
	9	CTG GGC GGG AC**1** GGG GAG TGG T
	10	CTG GGC GGG ACT GGG GAG **1**GG T
	11	ACC ACT CCC CAG TCC CGC CCA G
LTR-III + IV	**12**	AGG GAG GCG TGG CCT GGG CGG GAC TGG GGA GTG GT
	13	AGG GAG GCG TGG CCT GGG CGG GAC **1**GG GGA GTG GT
	14	AGG GAG GCG TGG CCT GGG CGG GAC TGG GGA G**1**G GT
	15	AGG GAG GCG **2**GG CCT GGG CGG GAC TGG GGA GTG GT
	16	AGG GAG GCG **2**GG CCT GGG CGG GAC TGG GGA G**1**G GT

[a]ONs **3**, **8** and **12** are native unmodified ONs of LTR-III, LTR-IV and LTR-(III+IV), respectively. ONs **4**–**6** are native LTR-III modified with FBFdU (**1**) and or FdU (**2**) at T_{24} and T_9, respectively. ONs **9** and **10** are native LTR-IV modified with FBFdU (**1**) at T_{12} and T_{19}, respectively. ONs **13**–**16** are native LTR-(III+IV) modified with FBFdU (**1**) and or FdU (**2**) at T_{25} or T_{32} and T_{10}, respectively. **7** and **11** are complementary ONs of **4** and **8**, respectively.

2.2.2 Probing LTR-III GQ-hairpin motif

Based on the structural considerations, ON **4** was synthesized wherein FBFdU was incorporated at the GQ domain (T_{24}) and FdU was incorporated at the hairpin domain (T_9) using phosphoramidites **1a** and **2a**, respectively (Scheme S1). The placement of these labels was strategically done to avoid any peak overlap that might occur due to the presence of two fluorine labels, ensuring accurate analysis of the structures. The ON was purified by gel electrophoresis and characterized by mass analysis (Table A1, Appendix-I). CD spectra of control unmodified ON **3** and modified ON **4** were found to be similar depicting the formation of a hybrid GQ topology with positive bands at ~265 nm and ~285 nm (Figure 2A).[16] GQ form of ON **4** exhibited a slightly higher T_m value as compared to the native ON **3** (Figure 2B and Table 2). These results indicate that the incorporation of FBFdU and FdU has only a minor impact on the formation and stability of ON **4** GQ structure.



Scheme S1. FBFdU phosphoramidite (**1a**) and FdU phosphoramidite (**2a**) were synthesized following reported procedures.[48,57,58] ^{31}P NMR spectra of phosphoramidites (**1a**) and (**2a**) are provided in Appendix-I.

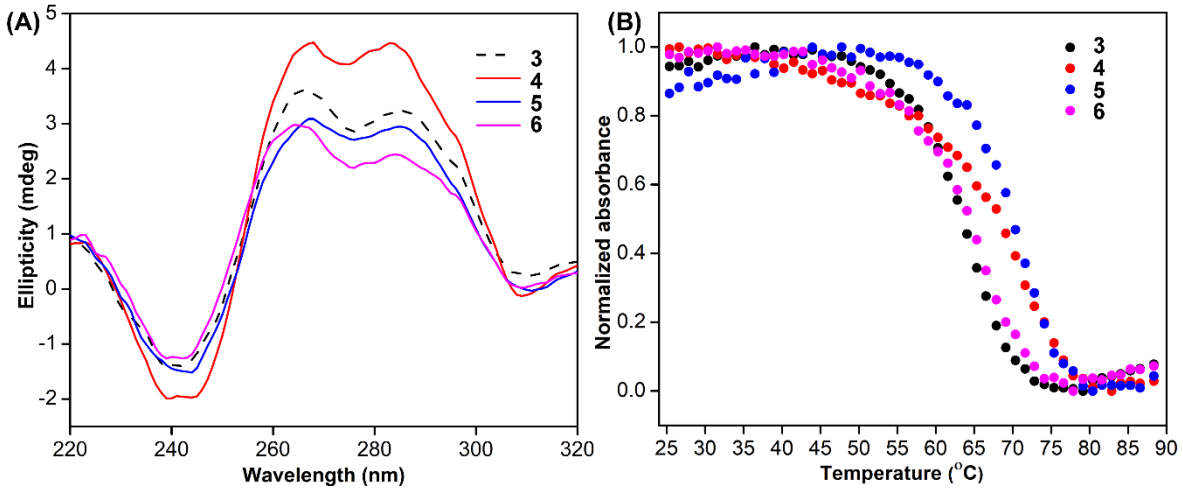


Figure 2. (**A**) CD spectra of control LTR-III ON **3** (5 µM), modified LTR-III ONs **4**–**6** (5 µM). (**B**) UV-thermal melting profiles for the same at 295 nm (2 µM).

Table 2: T_m °C of LTR-III native and modified ONs.

ONs	**3**	**4**	**5**	**6**
T_m °C	65.0 ± 0.7	69.4 ± 1.6	69.0 ± 1.7	64.8 ± 0.7

The ability of FBFdU to serve as a GQ reporter was evaluated by recording the fluorescence of LTR-III ON **4** and its corresponding duplex (**4•7**) in a buffer containing K^+

ions (Figure 3A and 3B). The GQ form of **4** displayed a discernibly lower fluorescence intensity and a slightly red-shifted emission band (417 nm) as compared to its perfect duplex (**4•7**, 414 nm). In support of our probe system design, ^{19}F NMR spectrum of ON **4** exhibited two distinct peaks at -122.51 ppm and -165.73 ppm arising from FBFdU and FdU, respectively (Figure 3C, blue line). ^{1}H NMR spectra of ON **4** and **3** revealed imino proton signals for both GQ and hairpin domains (Figure 4). In order to assign signals in the ^{19}F NMR spectrum, two singly modified ONs **5**, containing FBFdU at T_{24} (GQ domain) and **6**, containing FdU at T_9 (hairpin domain), were synthesized (Table 1, Figure 3A, Appendix-I). CD profiles and T_m values of these ONs indicated the formation of a stable hybrid GQ structure like the native and modified ONs **3** and **4** (Figure 2A and 2B, Table 2). ON **5** produced a single peak at -122.51 ppm from the GQ sensor matching the signal obtained using ON **4** (Figure 3C, red line). ON **6** produced a signal at -165.71 ppm from the hairpin sensor similar to ON **4** (Figure 3C, green line). Hence, signals emanating from FBFdU (-122.51 ppm) and FdU (-165.73 ppm) of ON **4** are assigned to GQ and hairpin domains, respectively. When ON **4** was annealed to a complementary ON **7**, the duplex structure produced new peaks at -121.54 ppm associated to FBFdU and -165.62 ppm associated to FdU (Figure 5). ^{1}H NMR also validated the formation of a duplex structure where characteristic peaks for Hoogsteen H-bonded GQ imino protons (10–12 ppm) disappeared and Watson-Crick H-bonded imino protons appeared between 12–14 ppm (Figure 5). Henceforth, the probe system provides a simplified ^{19}F NMR spectrum to detect the two domains of LTR-III ON **4** simultaneously.

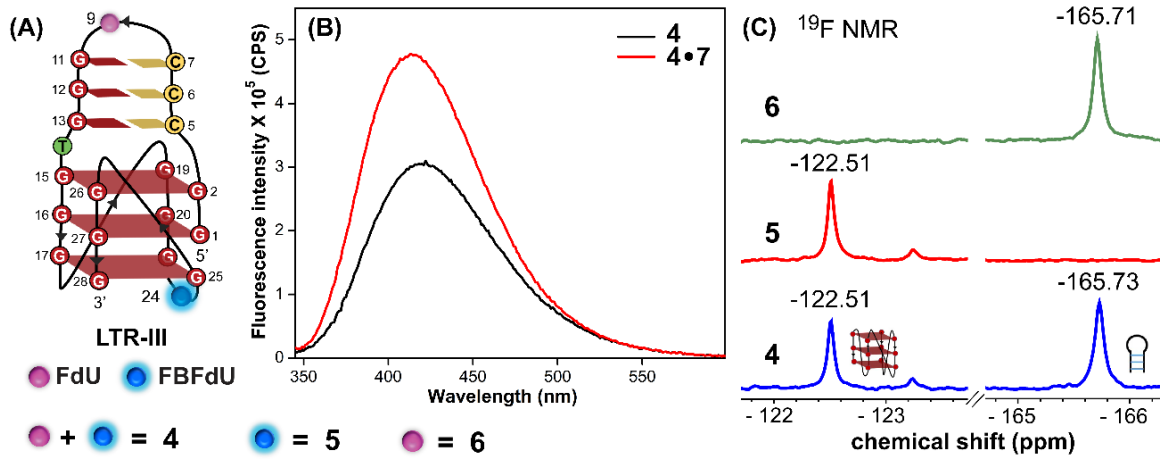


Figure 3. (**A**) Schematic representation of juxtaposed GQ-hairpin structure of LTR-III. ON **4** is modified with FdU at T_9 and FBFdU at T_{24} positions. ON **5** contains FBFdU at T_{24} position. ON **6** contains FdU at T_9 position. (**B**) Fluorescence spectra of ON **4** (GQ) and its duplex **4•7**. The samples were excited at 330 nm with excitation and emission slit widths of 7 nm and 9 nm, respectively. (**C**) ^{19}F NMR spectra of ONs **4**–**6**.

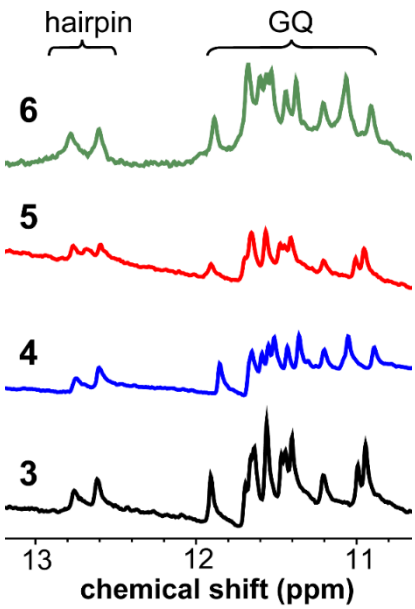


Figure 4. Partial ^{1}H NMR spectra (45 μM) of control ON **3** and modified ONs **4**–**6**.

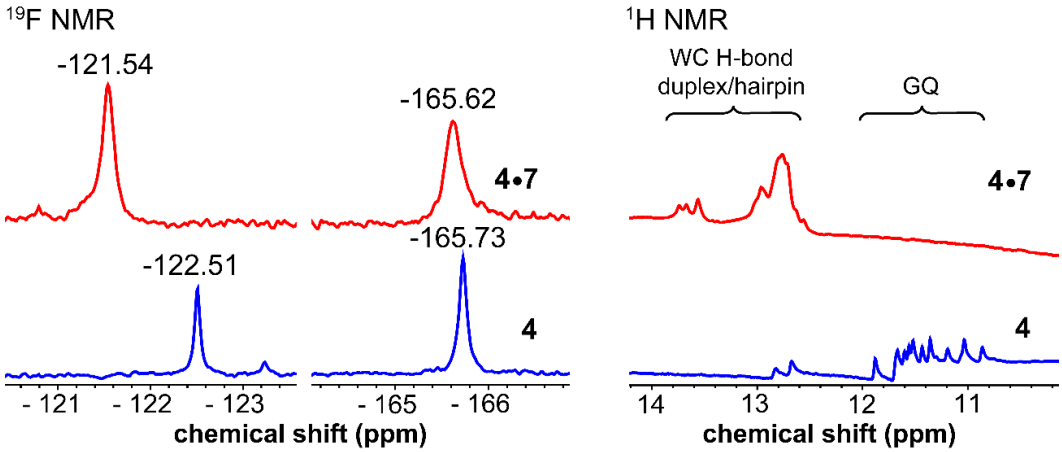


Figure 5. ^{19}F and ^{1}H NMR spectra (45 μM) of ON **4** and its duplex **4•7**.

2.2.3 Probing LTR-IV GQs

Next, we studied the LTR-IV G-rich region by incorporating FBFdU at T_{12} (ON **9**) and T_{19} (ON **10**) positions using the modified phosphoramidite **1a** (Table 1, Figure 6A, Appendix-I). To our presumption, FBFdU placed at these positions did not affect the formation and stability of the native parallel GQ topology as ascertained by CD and UV-thermal melting experiments (Figure 6A and 6B, Table 3). ONs **9** and **10** fluorescently reported the formation of a GQ structure with an intense band centered around 428 nm (Figure 7B). However, corresponding duplexes showed reduced emission. In particular, duplex **10•11** displayed a significant

reduction in fluorescence intensity. Notably, ^{19}F NMR spectra of ON **9** (75 µM) exhibited multiple peaks revealing the formation of different GQ structures (Figure 7B, blue line). Imino proton signals appearing between 10–12 ppm supported the formation of GQs (Figure 8). GQs can stack on top of each other by 5′-5′ end-to-end stacking interaction resulting in higher ordered GQs and such structures are usually observed for sequences capable of forming a parallel topology.[59] Hence, in consensus with a literature report,[17] observed ^{19}F signals could be associated to higher ordered structures originating from the monomeric parallel GQ motif. To further evaluate the ^{19}F signals, NMR spectra of ON **9** were recorded at a much lower concentration and in the presence of a synthetic crowding agent (PEG). It has been shown that low concentrations of the ON favour monomeric structure, whereas synthetic crowding agents like PEG favour higher ordered structures.[60,61] ^{19}F NMR recorded at a lower concentration of the ON **9** (10 µM) produced a major peak at -120.42 ppm, which was also present at a higher concentration of the ON (Figure 7C, purple line). While this peak was assigned to the monomeric GQ structure, a peak at -121.74 ppm in PEG 200 (40% v/v) is likely due to the formation of a higher ordered GQ structure (green line). The formation of GQ structures under these conditions was confirmed by ^{1}H NMR (Figure 8). ON **9** hybridized to it's complementary ON **11** gave a ^{19}F peak for the duplex form, which was also ascertained by ^{1}H NMR (Figure 7C, red line and Figure 8). ON **10** containing the modification at T_{19} position though gave multiple ^{19}F peaks, the spectrum was not well resolved (Figure 9). Peak broadening was also observed in ^{1}H NMR spectrum, and hence, this sequence was not used in further studies. Taken together, these results endorse that FBFdU is a useful GQ tool that allows access to study sequences forming multiple structures.

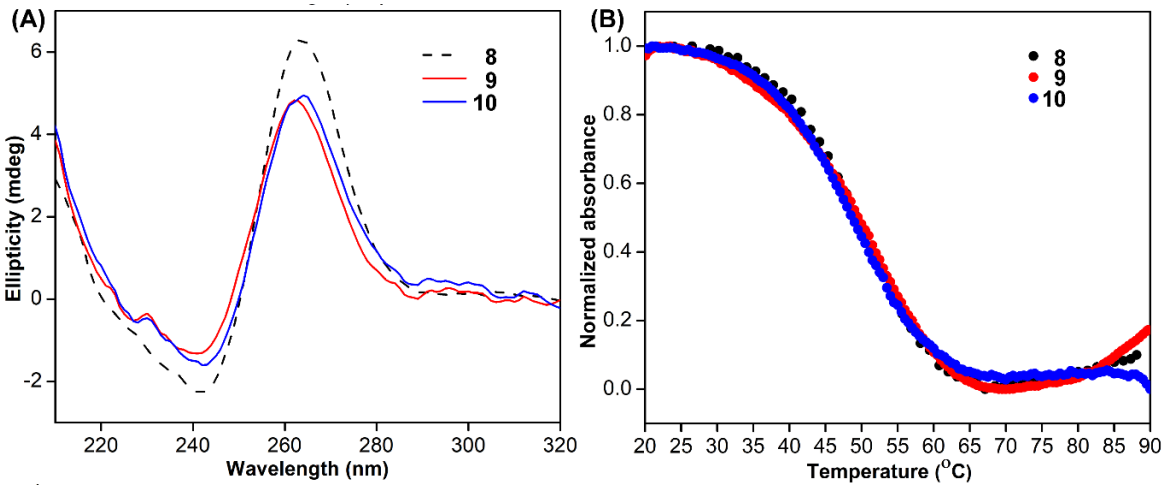


Figure 6. (**A**) CD spectra of control LTR-IV ON **8** (8 µM), modified LTR-IV ONs **9** and **10** (8 µM). (**B**) UV-thermal melting profiles for the same at 295 nm (5 µM). See Table 3 for T_m values.

Table 3. T_m °C of LTR-IV native and modified ONs.

ONs	8	9	10
T_m °C	51.6 ± 0.3	52.2 ± 0.7	51.6 ± 1.2

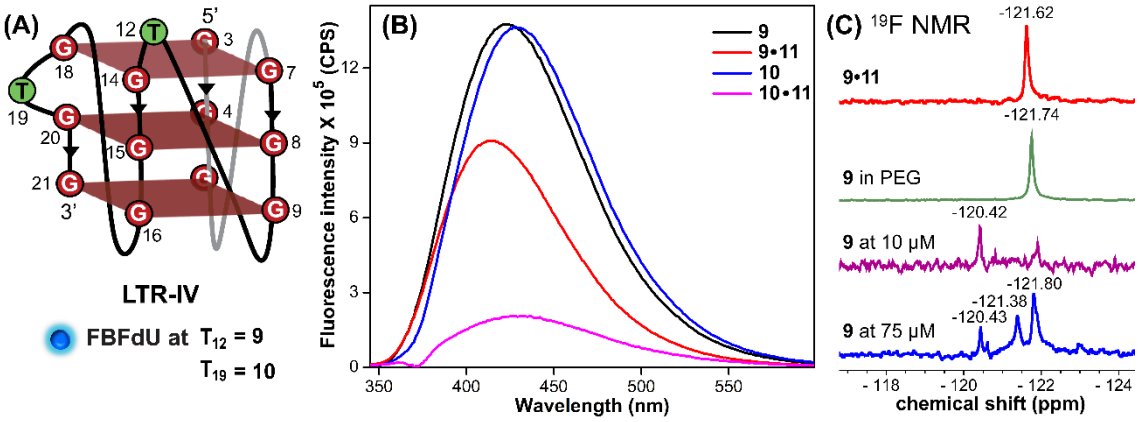


Figure 7. FBFdU reports the formation of LTR-IV GQs. (**A**) Schematic representation of the parallel GQ structure of LTR-IV ON. ON **9** contains FBFdU at T_{12} position. ON **10** contains FBFdU at T_{19} position. **9•11** and **10•11** are corresponding duplexes. (**B**) Fluorescence spectra (1 µM) of ONs **9** and **10** (GQ) and their duplexes. The samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively. (**C**) ^{19}F NMR spectra of ON **9** under different conditions.

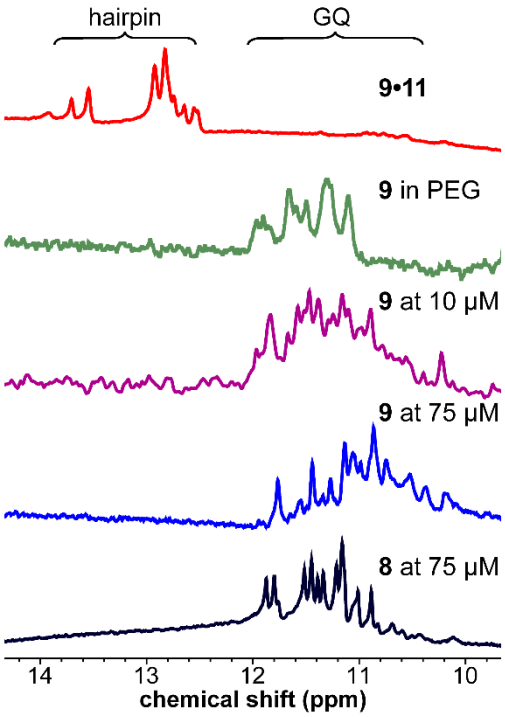


Figure 8. Partial ^{1}H NMR spectra of control ON **8** (75 µM) and ON **9** under different conditions.

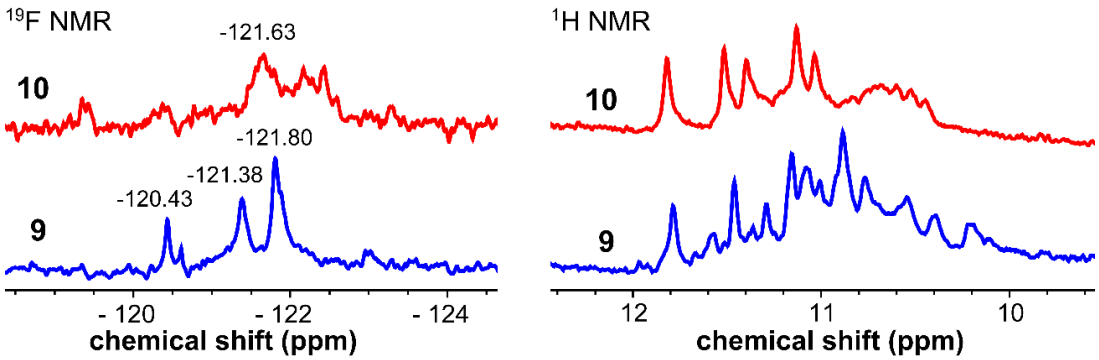


Figure 9. ^{19}F and ^{1}H NMR spectra of ONs **9** and **10.**

2.2.4 Computational models provide insights on how the probe system senses LTR GQs

To gain structural insights on the GQ sensing ability of FBFdU, we compared the photophysical properties of the nucleoside analog and structural models of labeled LTR ONs. FBFdU is highly environment-sensitive, and hence, when incorporated into ONs its photophysical properties are influenced by stacking interaction, solvation-desolvation and rigidification-derigidification effects, and electron transfer process with adjacent guanosine bases (Table 4).[48,62] While a polar environment favours high and red-shifted fluorescence, stacking interaction and electron transfer process quench the fluorescence of the analog. Further, due to the presence of a molecular rotor element in the nucleoside, the orientation of the FBF ring with respect to the uracil ring influences its fluorescence intensity.[48] Therefore, observed differences in fluorescence intensity of GQ and duplex structures are due to differences in the microenvironment around the probe.

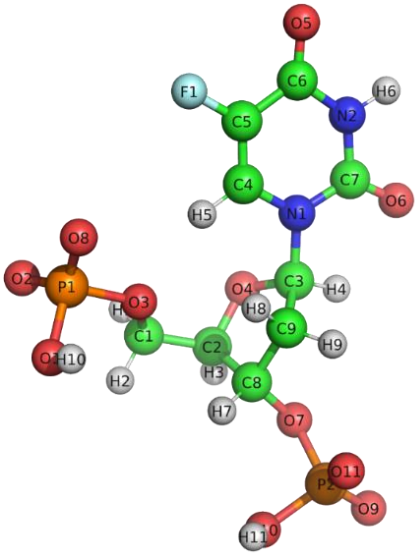
To examine the probe environment and its interaction with neighboring bases, structural models of labeled LTR-III ON **4** and LTR-IV ONs **9** and **10** were generated. Force field parameters of FBFdU and FdU (Figure 10 and 11) were first calculated and were incorporated into the templates with PDB ID: 6H1K and 2N4Y.[16,17] Two control systems, LTR-III ON **3** and LTR-IV ON **8**, were also generated. All the complexes were subjected to 500 ns MD simulations and the root mean square deviation (RMSD) revealed that the simulations are well equilibrated (Figure 11). Superimposition of the major cluster of ON **4** and the native ON **3** is almost identical (Figure 13). This is further visible from the root mean square fluctuation (RMSF, Figure 13A). In ON **4**, FBFdU placed at T_{24} strongly stacks below the tetrad formed by G_{25}•G_{28}•G_{17}•G_{21} and experiences a hydrophobic environment (Figure 14A–14C). The plots representing the center of mass (COM) distance and angle defined between the normal to FBFdU and G_{28} showed steady values of ~5 Å and 0−45° (~82% stacking), respectively, which are the defined parameters for proper stacking interaction (Figure 15).[63] This major

conformation exists for ~86% of the simulation, which manifests in the form of a low intense band around an emission maximum of methanol (417 nm, Figure 3B). In case of duplex **4•7**, C5-modified FBFdU flanked between C_{23} and G_{25} would be projected in the major groove and is likely to experience less stacking interaction as compared to in the GQ structure. Hence, the duplex structure display higher fluorescence intensity with no apparent change in emission maximum. A similar conformation has been observed for C5-heterocycle-modified pyrimidine nucleoside analogs in duplexes.[64,65]

Note: Computational analysis was conducted by Prof. P. I. Pradeepkumar and his students, Priyasa Majee and Sruthi Sudhakar.

Table 4. Absorbance and emission wavelengths of nucleoside FBFdU (**1**) in different solvents.[48] Data reported from *J. Am. Chem. Soc.* **2018**, *140*, 12622–12633.[48]

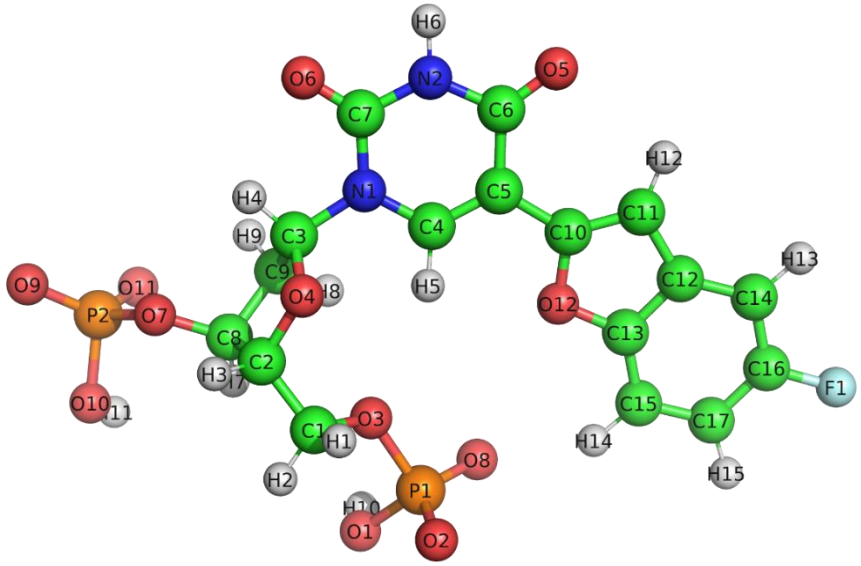
solvent	λ_{max} (nm)	λ_{em} (nm)	Φ
water	322	437	0.11
methanol	322	418	0.04
dioxane	324	400	0.03
ethylene glycol	325	420	0.20
glycerol	326	424	0.52




```
     0  0              2
  This is a remark line
molecule.res
  FDU            INT         0
  CORRECT        OMIT        DU       BEG
        0.0000
           1   DUMM        DU        M        0        -1        -2         0         0         0         0
           2   DUMM        DU        M        1         0        -1     1.449         0         0         0
```

3	DUMM	DU	M	2	1	0	1.523	111.21	0	0
4	P1	P	M	3	2	1	1.54	111.208	-180	1.294462
5	O2	O2	E	4	3	2	1.469	81.307	-120.503	-0.909793
6	O8	O2	E	4	3	2	1.472	46.311	76.917	-0.909793
7	O3	OS	M	4	3	2	1.646	110.787	-13.801	-0.640503
8	C1	CT	M	7	4	3	1.398	120.517	-120.842	0.144315
9	H1	H1	E	8	7	4	1.083	110.027	63.971	0.038448
10	H2	H1	E	8	7	4	1.086	110.759	-55.331	0.038448
11	C2	CT	M	8	7	4	1.516	109.506	-175.355	0.10031
12	H3	H1	E	11	8	7	1.083	108.884	179.331	0.068969
13	O4	OS	S	11	8	7	1.426	109.53	-62.744	-0.471067
14	C3	CT	3	13	11	8	1.392	111.377	121.413	0.280157
15	H4	H2	E	14	13	11	1.075	110.454	103.214	0.089924
16	N1	N*	S	14	13	11	1.474	108.136	-141.827	-0.385478
17	C4	CM	B	16	14	13	1.383	119.469	52.773	0.047923
18	H5	H4	E	17	16	14	1.075	116.654	3.002	0.233236
19	C5	CM	B	17	16	14	1.328	121.991	-176.865	-0.094409
20	C6	C	B	19	17	16	1.45	121.491	0.422	0.642038
21	O5	O	E	20	19	17	1.2	127.066	179.948	-0.707254
22	N2	NA	B	20	19	17	1.38	111.743	0.057	-0.646274
23	H6	H	E	22	20	19	0.997	116.47	179.661	0.304756
24	C7	C	S	22	20	19	1.379	128.863	0.555	0.73023
25	O6	O	E	24	22	20	1.2	119.961	178.871	-0.727943
26	F1	F	E	19	17	16	1.327	122.327	179.93	-0.148638
27	C9	CT	B	14	13	11	1.524	106.529	-18.129	-0.132195
28	H8	HC	E	27	14	13	1.082	110.583	-84.598	0.063958
29	H9	HC	E	27	14	13	1.08	113.265	151.624	0.063958
30	C8	CT	M	11	8	7	1.537	115.49	56.62	0.179574
31	H7	H1	E	30	11	8	1.081	111.369	22.658	0.078991
32	O7	OS	M	30	11	8	1.402	107.933	144.829	-0.626347
LOOP										
C7	N1									
C8	C9									
IMPROPER										
C7	C4	N1	C3							
C5	H5	C4	N1							
C6	C4	C5	F1							
C5	N2	C6	O5							
C6	C7	N2	H6							
N1	N2	C7	O6							
DONE										
STOP										

Figure 10. Cartesian coordinates and RESP charges calculated for the FdU adduct generated using Gaussian 16. Carbons atoms are represented by green, nitrogen atoms by blue, oxygen atoms by red, fluorine by cyan and hydrogen atoms by white respectively.



0 0 2
This is a remark line
molecule.res
FBFDU INT 0
CORRECT OMIT DU BEG
0.0000

1	DUMM	DU	M	0	-1	-2	0	0	0	0
2	DUMM	DU	M	1	0	-1	1.449	0	0	0
3	DUMM	DU	M	2	1	0	1.523	111.21	0	0
4	P1	P	M	3	2	1	1.54	111.208	-180	1.329951
5	O2	O2	E	4	3	2	1.469	102.542	-69.037	-0.907294
6	O8	O2	E	4	3	2	1.477	133.025	105.246	-0.907294
7	O3	OS	M	4	3	2	1.644	57.297	-173.44	-0.629415
8	C1	CT	M	7	4	3	1.402	118.04	54.604	0.143944
9	H1	H1	E	8	7	4	1.084	109.87	71.765	0.042331
10	H2	H1	E	8	7	4	1.086	110.559	-47.174	0.042331
11	C2	CT	M	8	7	4	1.514	110.845	-167.033	0.101983
12	H3	H1	E	11	8	7	1.083	108.375	176.116	0.071046
13	O4	OS	S	11	8	7	1.429	109.832	-66.203	-0.463418
14	C3	CT	3	13	11	8	1.391	110.667	128.117	0.279273
15	H4	H2	E	14	13	11	1.077	111.028	96.849	0.094203
16	N1	N*	S	14	13	11	1.475	107.506	-147.989	-0.379024
17	C4	CM	B	16	14	13	1.364	119.403	52.105	0.123751
18	H5	H4	E	17	16	14	1.075	114.89	2.014	0.233422
19	C5	CM	B	17	16	14	1.35	123.996	-177.596	-0.298709
20	C6	C	B	19	17	16	1.455	118.481	1.281	0.680537
21	O5	O	E	20	19	17	1.202	127.194	179.279	-0.708119
22	N2	NA	B	20	19	17	1.381	113.242	-0.56	-0.638327
23	H6	H	E	22	20	19	0.997	116.23	179.821	0.311693
24	C7	C	S	22	20	19	1.373	128.831	0.532	0.743339
25	O6	O	E	24	22	20	1.197	120.755	179.188	-0.707045
26	C10	C*	S	19	17	16	1.461	121.17	-178.289	0.120509
27	C11	C*	B	26	19	17	1.347	133.367	-175.809	-0.216033
28	C12	CB	S	27	26	19	1.446	105.519	-179.349	-0.092125
29	C13	CB	B	28	27	26	1.389	105.35	0.092	0.052891
30	C15	CA	B	29	28	27	1.38	124.347	179.52	-0.068718
31	C17	CA	B	30	29	28	1.38	116.104	0.291	-0.19327
32	C16	CA	B	31	30	29	1.39	119.966	-0.033	0.074936
33	C14	CA	S	32	31	30	1.372	123.833	-0.186	-0.165353
34	H13	HA	E	33	32	31	1.075	120.429	-179.75	0.122269
35	F1	F	E	32	31	30	1.346	117.813	-179.963	-0.18672
36	H15	HA	E	31	30	29	1.074	121.136	179.921	0.1269

37	H14	HA	E	30	29	28	1.078	119.977	-179.04	0.228791
38	O12	OS	E	29	28	27	1.35	109.737	-0.162	-0.17201
39	H12	HA	E	27	26	19	1.066	125.988	0.284	0.160246
40	C9	CT	B	14	13	11	1.522	105.968	-24.509	-0.132497
41	H8	HC	E	40	14	13	1.083	110.505	-79.957	0.065488
42	H9	HC	E	40	14	13	1.079	113.784	155.703	0.065488
43	C8	CT	M	11	8	7	1.538	115.921	53.869	0.183496
44	H7	H1	E	43	11	8	1.081	111.413	19.034	0.083087
45	O7	OS	M	43	11	8	1.401	107.861	141.357	-0.616531

LOOP

C7	N1
O12	C10
C14	C12
C8	C9

IMPROPER

C7	C4	N1	C3
C5	H5	C4	N1
C10	C6	C5	C4
C5	N2	C6	O5
C6	C7	N2	H6
N1	N2	C7	O6
C11	C5	C10	O12
C10	C12	C11	H12
C11	C14	C12	C13
C15	C12	C13	O12
C17	C13	C15	H14
C15	C16	C17	H15
C17	C14	C16	F1
C16	C12	C14	H13

DONE

STOP

Figure 11. Cartesian coordinates and RESP charges calculated for the FBFdU adduct generated using Gaussian 16. Carbons atoms are represented by green, nitrogen atoms by blue, oxygen atoms by red, fluorine by cyan and hydrogen atoms by white respectively.

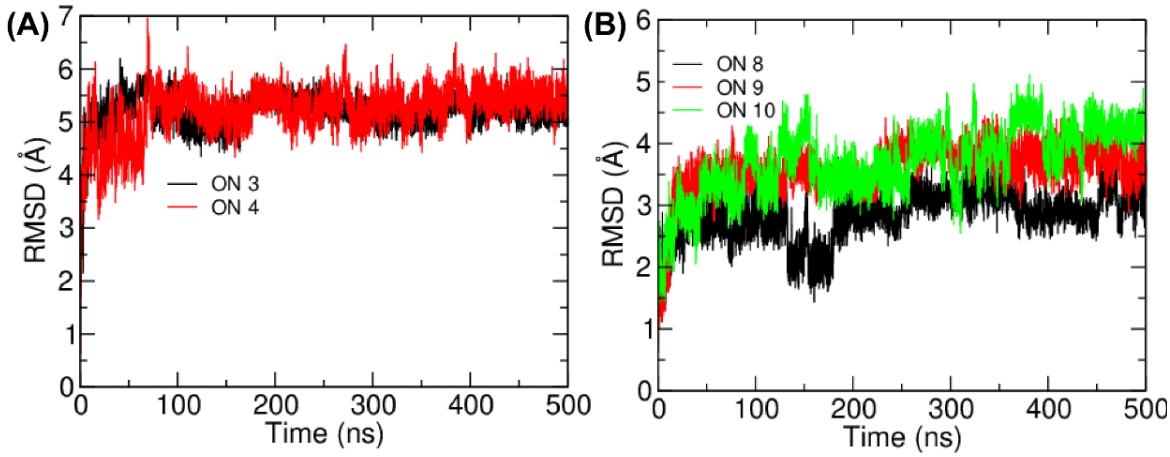


Figure 12. RMSD plot of (**A**) LTR-III ON **3** and ON **4** (**B**) LTR-IV ON **8**, ON **9** and ON **10**. RMSD values were calculated from the 500 ns MD simulations.

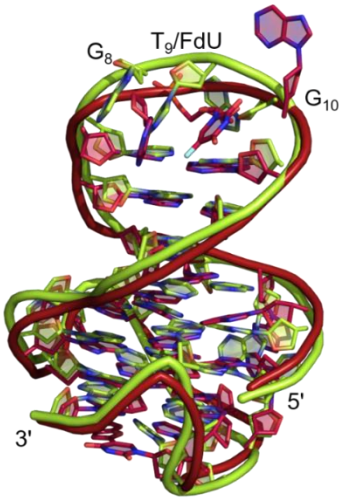


Figure 13. Superimposed images of the major clusters of LTR-III native ON **3** and modified ON **4**. ON **3** and **4** are represented in green and maroon, respectively. The bases, which show maximum changes in the orientation are labeled in the Figure. The clusters have been obtained from the 500 ns MD simulation. The incorporation of FdU results in the partial stacking of FdU with G_8 and alters the orientation slightly from the native form while FBFdU remains the same.

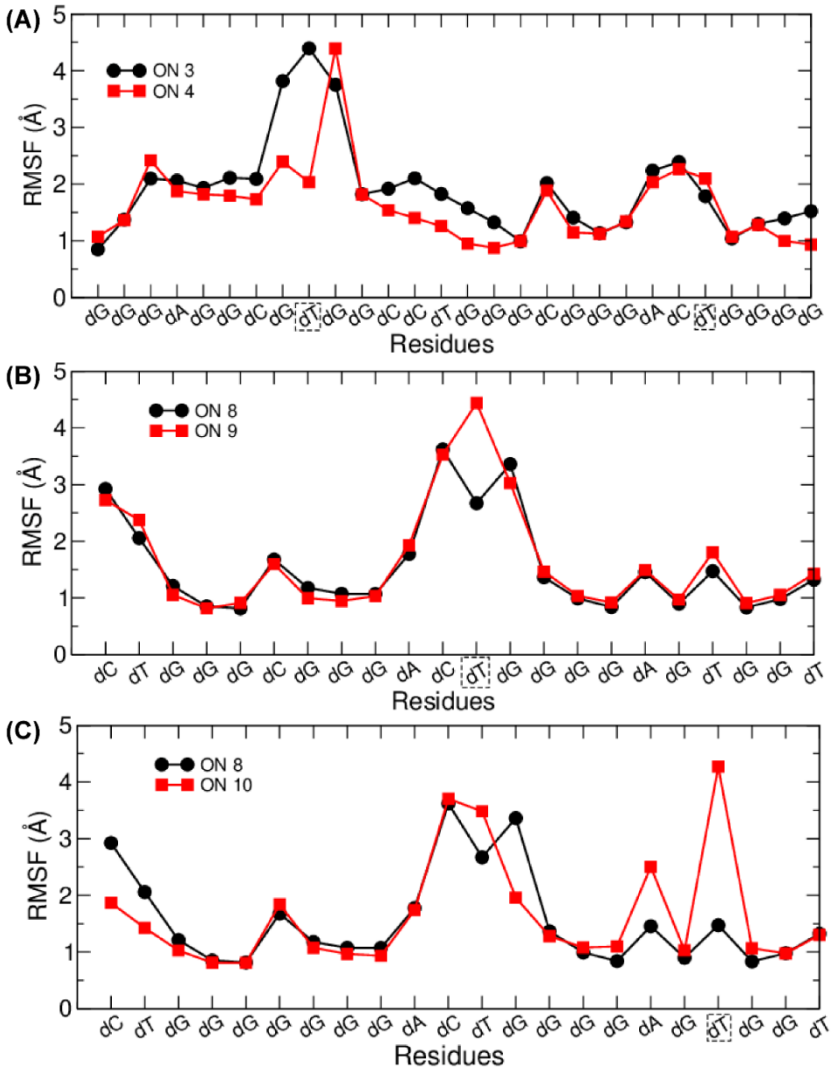


Figure 14. RMSF plots of (**A**) LTR-III ON **3** and ON **4** with variation in the G_8 and FdU (**B**) LTR-IV ON **8** and ON **9** (**C**) LTR-IV ON **8** and ON **10** with variation in the probe. The nucleotides are represented from 5′ to 3′. The position of the probe has been indicated by a dashed box around the nucleotide RMSF values were calculated from the 500 ns MD simulations.

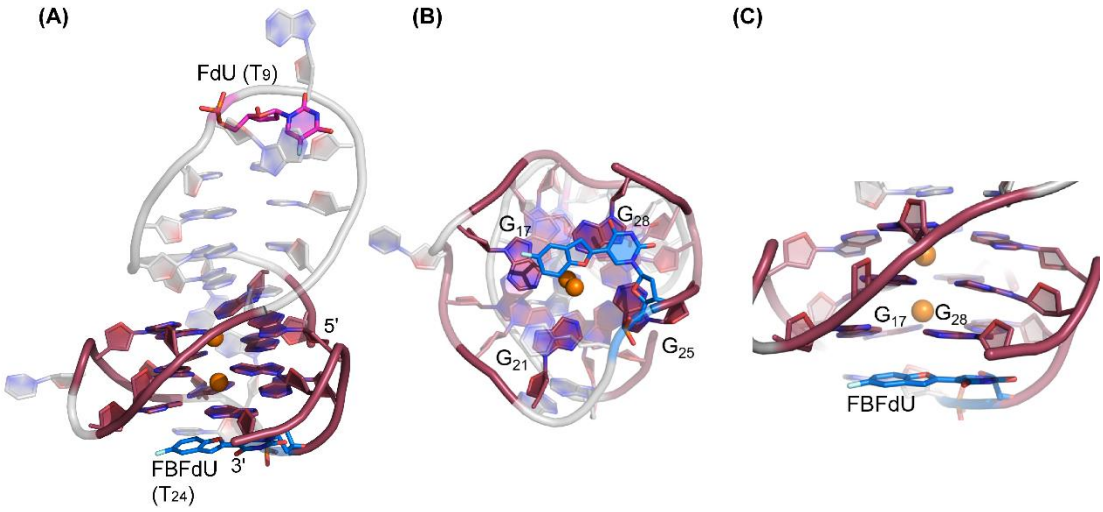


Figure 15. Representative images of major clusters of LTR-III ON **4**. (**A**) Overall structure with FdU and FBFdU in ON **4**. (**B**) Axial view showing the stacking of FBFdU over the bottom quartet. (**C**) Zoomed-in image showing the perpendicular orientation of FBFdU stacked with G_{17} and G_{28}. GQ bases are represented in maroon, FdU in magenta and FBFdU in blue. K^+ ions are represented as orange spheres. The clusters have been obtained from the 500 ns MD simulation.

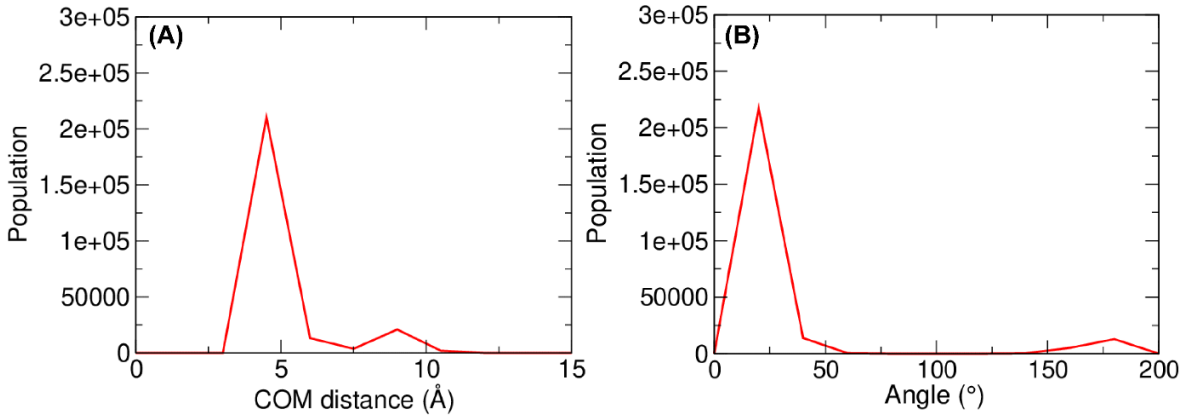


Figure 16. Population distribution of the (**A**) centre of mass distance (COM) distance between G_{28} and FBFdU in ON **4**, (**B**) angle between the normal to G_{28} and FBFdU. The values were calculated from the 500 ns MD simulations.

Simulations of LTR-IV ON **9** revealed 3 clusters of ~30, 30, and 26% each, and ON **10** revealed 2 clusters accounting for ~85% of the population (Figure 17 and 18). The modified base adopts an alternate conformation without affecting the GQ topology (Figure 19 and 20). Models of **9** and **10** revealed that the probe placed at T_{12} and T_{19} positions, respectively, is flipped out (solvent exposed) and is away from the G-tetrad core (Figure 17 and 18). Hence, these ONs exhibit high fluorescence intensity with emission maximum more towards water polarity, which is in line with the photophysical properties of the nucleoside (423 nm and 428 nm, Table 4). In case of duplex **9•11**, FBFdU flanked between C_{11} and G_{13} would experience partial stacking interaction, and hence, shows lower fluorescence intensity at λ_{em} = 418 nm. FBFdU in duplex **10•11** exhibits significant reduction in fluorescence intensity due to partial stacking interaction followed by more quenching from two adjacent Gs (G_{18} and G_{20}). Although ^{19}F component of nucleoside probes exhibits distinct chemical shifts for different GQ and duplex structures, rationalizing peak positions is not trivial. This is because the environment around the probes and their interaction with neighbouring bases as mentioned above can have varying shielding-deshielding effects on the ^{19}F atom,[66–68] which is difficult to predict and is also evident from an obscure trend in chemical shift of nucleosides in different solvents (Table 5).

Table 5. ^{19}F NMR chemical shift (ppm) of FBFdU (**1**) and FdU (**2**) in different solvents.[48, 58] Although ^{19}F label exhibits distinct chemical shifts in different solvents, the trend based on polarity and viscosity order is complex.

solvent	FBFdU	FdU
water	-121.78	-166.49
methanol	-123.70	-169.39
dioxane	-123.01	-169.06
ethylene glycol	-122.35	-168.05

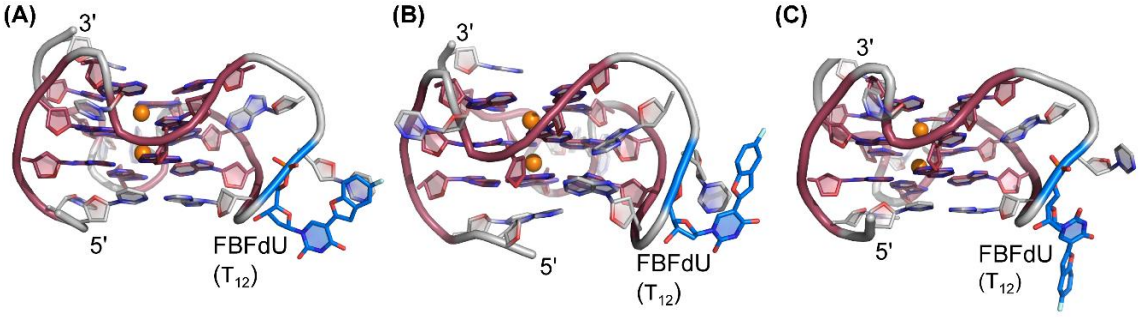


Figure 17. Representative images of three major clusters of LTR-IV ON **9**. (**A**) Cluster 1, (**B**) cluster 2 and (**C**) cluster 3. GQ bases are represented in maroon and FBFdU in blue. K^+ ions are represented as orange spheres. Clusters have been obtained from the 500 ns MD simulation.

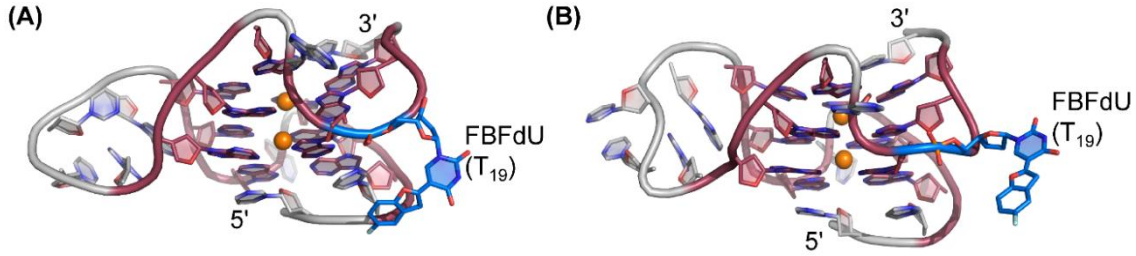


Figure 18. Representative images of two major clusters of LTR-IV ON **10**. (**A**) Cluster 1 and (**B**) cluster 2. GQ bases are represented in maroon and FBFdU in blue. K^+ ions are represented as orange spheres. Clusters have been obtained from the 500 ns MD simulation

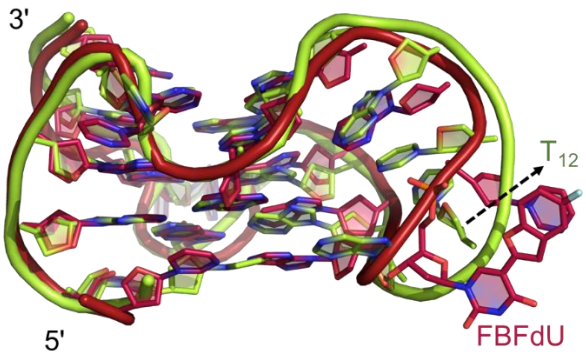


Figure 19. Superimposed images of the major clusters of LTR-IV native ON **8** and modified ON **9**. ON **8** and **9** are represented in green and maroon, respectively. The bases, which show maximum changes in the orientation are labeled in the Figure. The clusters have been obtained from the 500 ns MD simulation.

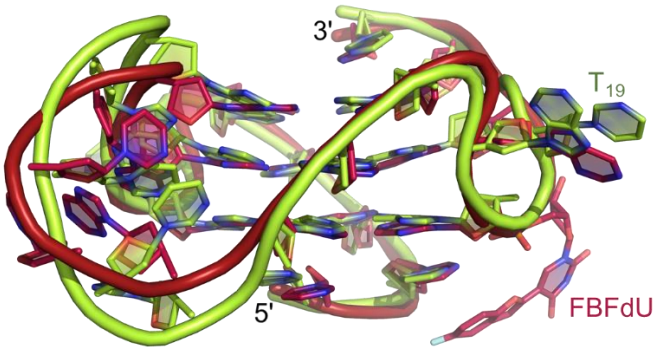


Figure 20. Superimposed images of the major clusters of LTR-IV native ON **8** and modified ON **10**. ON **8** and **10** are represented in green and maroon, respectively. The bases, which show maximum changes in the orientation are labeled in the Figure. The clusters have been obtained from the 500 ns MD simulation.

2.2.5 Probing GQ structure of LTR promoter region in cellular environment by ^{19}F NMR

The LTR G-rich promoter region encompassing both LTR-III and -IV segments were recently characterized by Richter and Phan groups *in vitro* using ^{1}H NMR. Their results suggest that it largely forms a juxtaposed GQ-hairpin motif like LTR-III.[16] As an important step forward, we decided to use the spectral properties of our nucleoside analogs to systematically determine the GQ structure adopted by the LTR region in cellular environment and probe its druggable space.

In-cell ^{19}F NMR has become a powerful tool to study nucleic acid structures in cellular milieu,[48,70,71] as fluorine is 100% abundant, highly sensitive and importantly, absent in cellular systems (no background signal).[72–78] Further, its signal does not undergo significant line broadening in the heterogeneous cellular environment, which is very severe in the case of proton signal.[79] To determine ^{19}F signatures of various domains and survey GQ and hairpin structures, LTR ONs **13**–**15** labeled with FBFdU or FdU at different positions were synthesized (Appendix-I). While ONs **13** and **14** containing FBFdU was synthesized to detect GQs, ON **15** containing FdU was prepared to detect the hairpin domain (Figure 22A). It is to be noted that the position of modification was maintained as in the individual LTR G-rich segments. CD spectra of modified (**13**–**15**) and control unmodified (**12**) ONs exhibited bands corresponding to a hybrid topology similar to the one adopted by the LTR-III region (Figure 21A). While the T_m values of ON **14** (FBFdU at T_{32}) and **15** (FdU at T_{10}) were found to be similar compared to the native ON **12**, ON **13** (FBFdU at T_{25}) displayed a slightly higher T_m

(~4 °C, Figure 21B, Table 6). ON **13** gave a single broad peak at -123.20 ppm and ON **14** exhibited a sharper peak at -122.10 ppm for the GQ domain (Figure 22B). On the other hand, FdU-labeled ON **15** produced a distinct peak (-165.72 ppm) for the hairpin domain. ^{1}H NMR spectra of ONs clearly revealed the presence of imino protons for GQ and hairpin motifs (Figure 22B). Further, absence of multiple ^{19}F peaks suggests that the parallel topology associated to the LTR-IV region is possibly not formed by the longer promoter region (compare with Figure 7C).

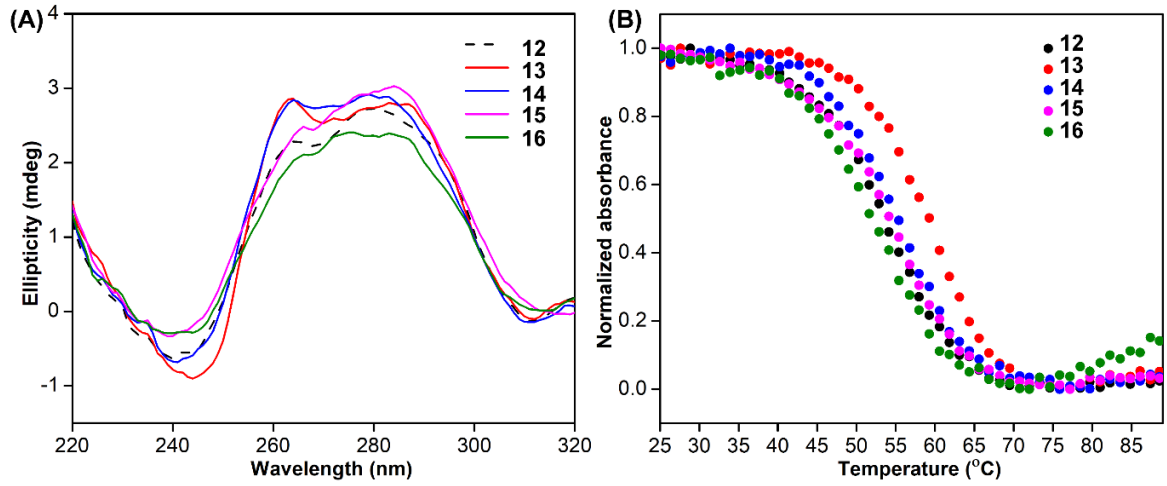


Figure 21. (**A**) CD spectra of control LTR-(III+IV) ON **12** (5 µM), modified LTR-(III+IV) ONs **13**–**16** (5 µM). (**B**) UV-thermal melting profiles for the same at 295 nm (2 µM).

Table 6. T_m °C of LTR-(III+IV) native and modified ONs.

ONs	**12**	**13**	**14**	**15**	**16**
T_m °C	55.6 ± 0.2	60.6 ± 0.3	51.6 ± 1.2	55.6 ± 0.4	55.2 ± 0.4

In order to detect both GQ and hairpin motifs simultaneously, we synthesized ON **16** containing FdU at T_{10} and FBFdU at T_{32} positions. Modification at T_{25} (like in ON **13**) was avoided as it gave a broader peak and poorer base line. Rewardingly, ON **16** displayed two distinct signals, one each for GQ (-122.10 ppm) and hairpin (-165.72 ppm) structures with chemical shifts same as that of ONs **14** and **15** designed to detect the structures independently (Figure 22B). As well, ON **16** depicted a ^{1}H NMR spectrum revealing the presence of GQ and hairpin structures. Collectively, the probe combination provides distinct ^{19}F signatures for GQ and hairpin structures, and our results demonstrate that the LTR G-rich region predominantly folds into a GQ-hairpin motif similar to LTR-III *in vitro*.

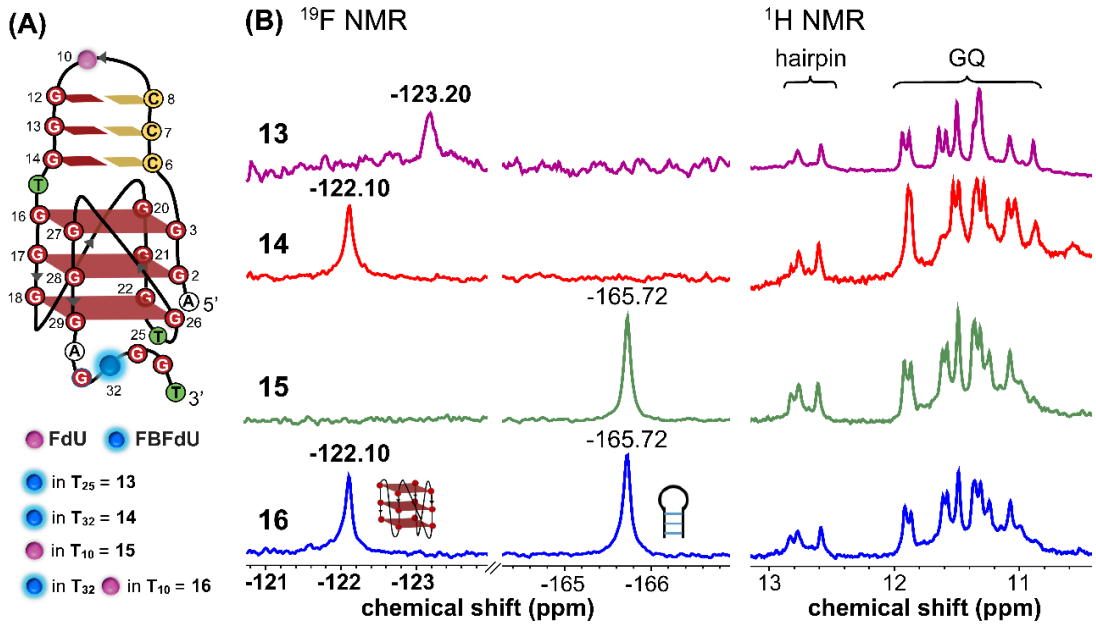


Figure 22. (**A**) Schematic representation of juxtaposed GQ-hairpin structure of LTR-III + IV region. (**B**) ^{19}F NMR and partial ^{1}H NMR spectra of ONs **13**–**16**.

Molecular crowding, complex cellular environment and ionic conditions can have significant effects on the folding process of a G-rich sequence. To obtain a progressive understanding of the LTR GQs in cell-free and cellular environments, we performed a systematic NMR analysis using *Xenopus* oocytes, a commonly used cellular model.[70,72,80,81] While a buffer mimicking intraoocyte ionic conditions serve as a cell-free system, frog egg lysate and extract serve as very good *ex vivo* systems to carryout NMR experiments to determine the structure of nucleic acids. First, we recorded the ^{19}F NMR spectrum of ON **16** 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). The ON acquired a conformation like the LTR-III, reflecting peaks at -122.02 ppm for the GQ and -165.65 ppm for the hairpin motifs (Figure 23, blue line). The formation of the GQ-hairpin structure was confirmed by ^{1}H NMR and CD experiments (Figure 23 and 24). Healthy stage V/VI eggs were selected, mechanically crushed and heat denatured. ON **16** incubated in the lysate obtained by centrifugation supported the formation of a hybrid architecture like in the intraoocyte buffer (Figure 23, red line). The inter-phase egg extract obtained by simply centrifuging crushed eggs maintains metabolite and protein contents mimicking the biological environment.[60] Interestingly, ^{19}F NMR spectrum of ON **16** incubated in the egg extract revealed the presence of GQ and hairpin motifs, albeit a slight broadening and shift in the signal (green line). This is a common effect observed in cell models, wherein slower tumbling rates prolongs the transverse relaxation time (T_2) resulting in peak broadening.[60,70,71,80] In contrast, due to extensive line broadening, the ^{1}H NMR spectrum fails to provide structural information

in egg extract (Figure 23). Henceforth, our fluorine-labeled nucleoside probes outweigh the applicability of proton NMR in cell-based analysis.

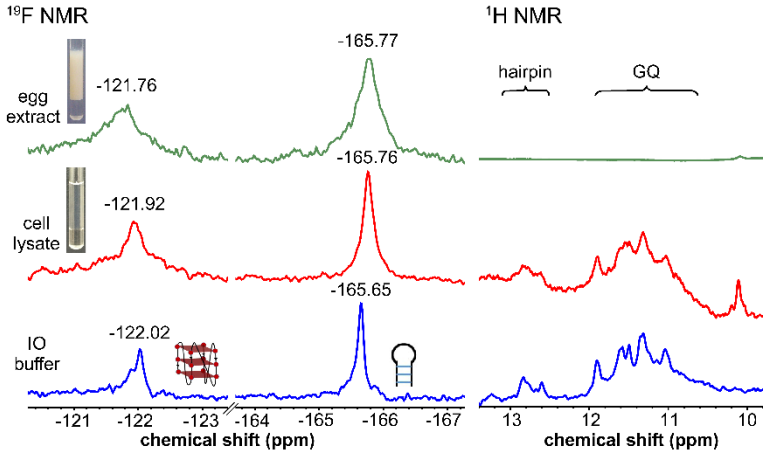


Figure 23. LTR G-rich region forms a GQ-hairpin structure in cellular environment as detected using FBFdU and FdU. ^{19}F and ^{1}H NMR spectra of ON **16** (100 µM) in IO buffer, frog egg lysate and frog egg extract (*ex vivo* cell model)**.**

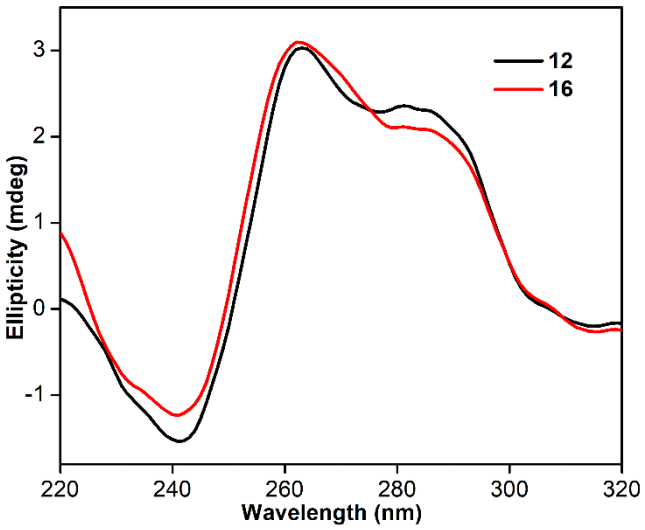


Figure 24. CD spectra of ONs **12** and **16** each at 5 µM in intraoocyte (IO) buffer.

To confirm if the signal is originating from the intact ON in cellular samples, after NMR acquisition, samples were analyzed by HPLC. A comparative analysis of HPLC profiles of the free ON, ON in lysate and free nucleoside **1** clearly pointed out that the ON is not degraded in the cellular environment (Figure 25). Further, ESI-MS of ON **16** recovered from the cell lysate ascertained the intactness of the labeled ON (Figure 26). Taken together, these results provide a clear evidence for the existence of a monomeric architecture preserving juxtaposed GQ and hairpin domains both *in vitro* and in cellular conditions underscoring the potential of LTR GQ as a target of selective therapeutic intervention.

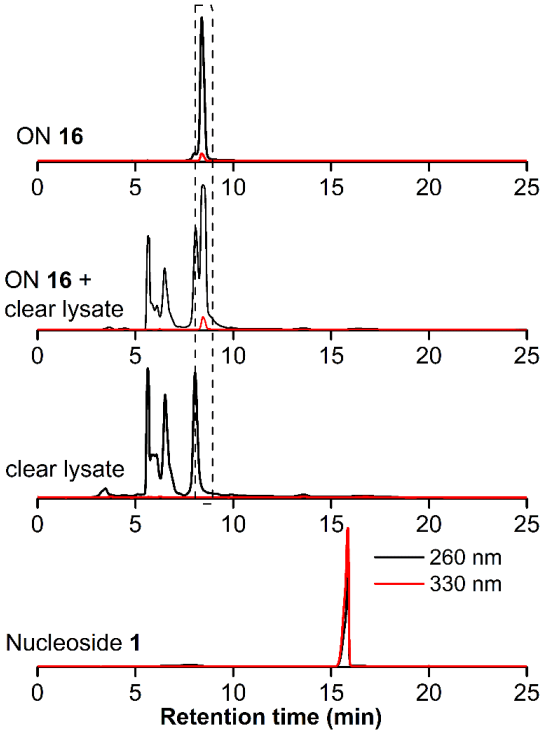


Figure 25. Comparison of RP-HPLC chromatogram of ON **16**, **16** in lysate after the NMR experiments, lysate (control) and nucleoside FBFdU **1** at 260 nm and 330 nm. ON **16** is stable in the lysate and no detectable degradation of ON **16** was observed (see the peak within the dashed line). Peaks between 5–8 min are from metabolites of the clear lysate.

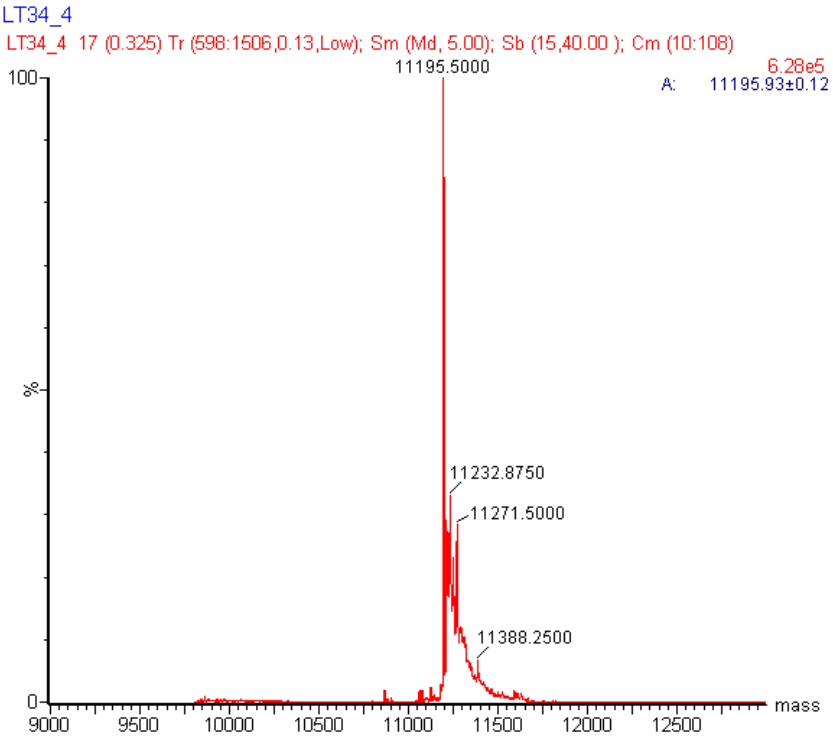


Figure 26. ESI-MS spectra of modified ON **16** extracted from lysate sample after NMR analysis (calculated mass = 11194.75, observed mass =11195.50).

2.3 Conclusions

We have devised a probe platform using two highly environment-sensitive nucleoside analogs (FBFdU and FdU) to study the structural polymorphism of a conserved HIV-1 LTR G-rich region in cell-free and cellular environments. These minimally invasive analogs produced a very simplified spectrum with distinct fluorescence and ^{19}F NMR signatures for different LTR GQ architectures. Importantly, using ^{19}F signatures of FBFdU and FdU we successfully identified that the LTR G-rich region adopts a GQ-hairpin architecture in a cellular environment. MD simulations gave insights on the structural basis by which FBFdU fluorescently senses different GQ topologies and distinguishes them from the duplex form. We envision that the comprehensive study will allow devising new drug molecules targeting the unique G-quadruplexes.

2.4 Experimental Section

2.4.1 Materials: 5-Flurobenzofuran-2′-deoxyuridine (**1**) and phosphoramidite substrates (**1a**) and (**2a**) were synthesised following a reported procedure.[48,57,58] 5-fluoro-2′-deoxyuridine (**2**) was purchased from Carbosynth. Monomers for solid-phase oligonucleotide (ON) synthesis such as *N*-benzoyl-protected dA, *N*-acetyl-protected dC, *N*,*N*-dimethylformamide-protected dG, and dT phosphoramidite substrates were purchased from ChemGenes, Glen Research and Innovassynth. Solid supports for DNA synthesis were procured from Glen Research. All other reagents needed for solid-phase ON synthesis were availed from Sigma-Aldrich. Control DNA ONs **3**, **7**, **8**, **11** and **12** were purchased from Integrated DNA Technology, purified by denaturing polyacrylamide gel electrophoresis (PAGE) and desalted using Sep-Pak Classic C18 cartridges (Waters Corporation). BRACO19 hydrochloride and all reagents (Bio-Ultra grade) used in the preparation of buffers were purchased from Sigma-Aldrich. TMPyP4 and Doxorubicin hydrochloride (DOX) were procured from Merck-Millipore. Millipore water after autoclaving was used for the preparation of all buffer solutions and in all biophysical studies.

2.4.2 Instruments: NMR spectra of small molecules were acquired in Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova software from Mestrelab Research. Mass data was obtained using ESI-MS Waters Synapt G2-Si Mass Spectrometry instrument. Modified DNA ONs were synthesized on a K&A DNA/RNA synthesizer H6. 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). UV-thermal melting analysis of the ONs was carried out on Cary 300 Bio UV-Vis spectrophotometer and Cary 3500 multicell UV-Vis spectrophotometer. CD measurements were 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 Solid-phase DNA synthesis: FBFdU (**1**) and or FdU (**2**) modified DNA ONs **4**–**6**, **9**, **10** and **13**–**16** were synthesized on a 1 μmole scale (1000 Å CPG solid support) with K&A H-6 synthesizer using phosphoramidite substrates. For modified phosphoramidites FBFdU (**1a**) and FdU (**2a**), double coupling of 3 min each was set (total 6 min). After the synthesis, the ONs were cleaved from the solid support using 30% aqueous ammonia and deprotected at 65 °C for 20 h. ONs were purified by denaturing PAGE (18 % or 20% gel) and the product bands were visualized by UV-shadowing. Product bands were excised and ONs were extracted with 4 mL of 0.5 M ammonium acetate buffer in a poly-prep column (Bio-Rad) for 12 h. Desalting was performed using a Waters C-18 cartridge. The purity of ONs was monitored by RP-HPLC (Appendix-I). The integrity was verified by ESI-MS (Appendix-I).

2.4.4 ESI-MS analysis: Negative mode ESI-MS analysis was performed by injecting DNA ONs (~800 pmol) dissolved in 50% acetonitrile in an aqueous solution of 10 mM triethylamine and 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol. See Appendix-I for mass spectra and Table A1 for details.

2.4.5. Circular dichroism (CD) analysis: ONs **3**–**6**, **8**–**10** and **12**–**16** were annealed in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl at 95 °C for 5 min and slowly cooled to RT. CD spectra were recorded from 320–200 nm at 25 °C using 1 nm bandwidth and sample volume of 200 μL using a quartz cuvette (Sterna Scientific, path length 2 mm) on a Jasco J-815 CD spectrometer. Each spectrum was recorded in duplicate with averaging three accumulations at scanning speed of 100 nm/min and baseline corrected for buffer contribution. Each spectrum was smoothened using the software provided by the manufacturer present in the system.

CD in intraoocyte buffer: Control ON **12** (5 μM) and modified ON **16** (5 μM) were annealed in intraoocyte (IO) buffer (25 mM HEPES (pH 7.5), 10.5 mM NaCl, 110 mM KCl, 130 nM

$CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA) at 95 °C for 5 minutes and allowed to cool at RT. CD spectra were recorded as above.

2.4.6. Thermal melting analysis: ONs **3**–**6**, **8**–**10** and **12**–**16** were annealed in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl as mentioned above. The spectra were recorded in Cary 300 Bio UV–Vis spectrophotometer for ONs **3**–**6** and ONs **12**–**16**. Cary 3500 multicell UV-Vis spectrophotometer was used for recording the spectra for ONs **8**–**10** with a temperature interval of 1 °C. Absorbance was recorded at 295 nm with a data interval of 1 °C for ONs **3**–**6**, ONs **12**–**16** and 0.5 °C for ONs **8**–**10**.

2.4.7. Computational analysis: Molecular dynamics (MD) simulations were carried out using the templates HIV LTR-III (PDB ID: 6H1K)[16] LTR-IV (PDB ID:2N4Y)[17]. Force field parameters were generated for FdU and FBFdU to prepare the ON structures with the probes. Both structures were prepared using GaussView 6.0 with phosphate capping at both 3′ and 5′ ends. FBFdU has a rotatable bond between the fluorophore and the base. A dihedral scan was performed for FBFdU with 36 rotations of 10 degrees each in Gaussian 16 version B.01[82] at theory level HF/6-31G* to find the most stable conformer. The stable conformer showing the lowest potential energy was then optimized in Gaussian 16 at the same theory level. FdU was also optimized using a similar strategy. ESP charges were calculated using Gaussian 16, and RESP fitting was done in the antechamber[83] module of AmberTools 19. The parmchk2 program generated an initial set of force field parameters. However, some were missing parameters, and others had high penalty scores. The capping was then removed, and after fixing the overall charges, the GAFF[84] library was used to fill in the missing parameters. Finally, prepin files were generated for complex preparation.

These modifications were incorporated into their respective templates in tleap. Central K^+ of the GQ core was also added using manual coordinate calculations. The complex was solvated with a rectangular water box using TIP3PBOX force field having an edge length of 10 Å. ~27 K^+ were added to neutralize the system. OL15[85] was used to define the DNA. MD simulations were carried out using our previously reported protocol.[86] Briefly, all the complexes were subjected to 10,000 steps of restrained minimization by the steepest descent method with a restraint of 2.0 kcal/mol. $Å^2$ followed by 100 ps of heating and 100 ps of density equilibration. Further, 800 ps of NPT equilibration and 500 ns of production run were carried out in GPU accelerated version of PMEMD[87-89] in AMBER 18.[90] A total of ~2.5 µs (5*500 ns) simulations were carried out. The SHAKE algorithm was applied to subject the hydrogens to

bond length constraints. All the MD analyses were carried out using the CPPTRAJ[91] module of AmberTools 19. The hierarchical agglomerative algorithm was used for clustering the trajectories. The cut-offs for stacking were a COM distance of 5 Å and a vector angle of 45 degrees. The trajectories were visualized using VMD, and images were rendered using PyMOL(Schrodinger LLC.)

2.4.8. Steady-state fluorescence of modified LTR ONs: LTR GQ structures of ONs **4** (0.5 µM), **9** (1 µM) and **10** (1 µM) were formed by heating the samples at 95 °C for 5 min in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl. The corresponding duplexes **4•7**, **9•11** and **10•11** were prepared by heating a 1:1.1 mixture of LTR ONs **4**, **9** and **10** with complementary ONs **7** and **11** at 95 °C for 5 min in the same ionic conditions as mentioned above. All the samples were cooled slowly to RT. Experiments were done in triplicate in a micro-fluorescence cuvette (Hellma, path length 1.0 cm) on a Fluoromax-4 spectrofluorometer (Horiba Scientific) at 25 °C.

2.4.9. NMR of LTR ONs: LTR GQ structures of ONs **3**–**6** (45 µM), **8** (75 µM), **9** (10 µM or 75 µM), **10** (75 µM), **13**–**16** (45 µM) were formed by heating the samples at 95 °C for 5 min in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl in 20% D_2O. The corresponding duplexes **4•7** and **9•11** were prepared by heating a 1:1.1 mixture of LTR ONs **4** and **9** with complementary ONs **7** and **11** respectively at 95 °C for 5 min in the same ionic conditions as mentioned above. ^{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 CryoProbe (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: 90.32 ppm; transmitter frequency offset: -145 ppm; acquisition time: 0.33 s; relaxation delay: 1.0 s; number of scans: 5000–6000. Using these parameters, spectra were obtained in 2–2.5 h. Each spectrum was processed with an exponential window function using lb = 20 Hz. ^{1}H NMR spectra were obtained with water suppression using excitation sculpting with gradients. Number of scans was 1200.

2.5 References

1. Arts, E. J.; Hazuda, D. J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.*, **2012**, *2*, a007161.

2. Peng, Y.; Zong, Y.; Wang, D.; Chen, J.; Chen, Z.-S.; Peng F.; Liu, Z. Current drugs for HIV-1: from challenges to potential in HIV/AIDS. *Front. Pharmacol.,* **2023**, *26*, 1294966.
3. Pomerantz, R. J. Reservoirs of human immunodeficiency virus type 1: the main obstacles to viral eradication. *Clin. Infect. Dis*., **2002**, *34*, 91–97.
4. Gupta, R. K.; Gregson, J.; Parkin, N.; Haile-Selassie, H.; Tanuri, A.; Andrade, F. L.; Kaleebu, P.; Watera, C.; Aghokeng, A.; Mutenda, N. *et al.* HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect. Dis*., **2018**, *18*, 346–355.
5. Archin, A.; Liberty, L.; Kashuba, A. D.; Choudhary, S. K.; Kuruc, J. D.; Crooks, A. M.; Parker, D. C.; Anderson, E. M.; Kearney, M. F.; Strain, M. C. *et al.* Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature*, **2012**, *487*, 482–485.
6. Nguyen, W.; Jacobson, J.; Jarman, K. E.; Jousset Sabroux, H.; Harty, L.; McMahon, J.; Lewin, S. R.; Purcell, D. F.; Sleebs, B. E. Identification of 5-substituted 2-acylaminothiazoles that activate Tat mediated transcription in HIV-1 latency models. *J. Med. Chem.*, **2019,** *62*, 5148–5175.
7. El-Desoky, A. H. H.; Eguchi, K.; Kishimoto, N.; Asano, T.; Kato, H.; Hitora, Y.; Kotani, S.; Nakamura, T.; Tsuchiya, S.; Kawahara, T. *et al*. Isolation, synthesis, and structure-activity relationship study on daphnane and tigliane diterpenes as HIV latency-reversing agents. *J. Med. Chem.*, **2022**, *65*, 3460–3472.
8. Zhang, J.; Crumpacker, C. HIV UTR, LTR, and epigenetic immunity. *Viruses*, **2022**, *14*, 1084.
9. Roebuck, K. A.; Saifuddin, M. Regulation of HIV-1 transcription. *Gene Expr*., **1999**, *8*, 67–84.
10. van Opijnen, T., Jeeninga, R. E., Boerlijst, M. C., Pollakis, G. P., Zetterberg, V., Salminen, M. and Berkhout, B. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. *J. Virol.*, **2004,** *78*, 3675–3683.
11. Pereira, L. A.; Bentley, K.; Peeters, A.; Churchill, M. J.; Deacon, N. J. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res*., **2000**, *28*, 663−668.
12. Jeeninga, R. E.; Hoogenkamp, M.; Armand-Ugon, M.; de Baar, M.; Verhoef, K.; Berkhout, B. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J. Virol.*, **2000**, *74*, 3740−3751.
13. Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palu, G.; Richter, S. N. A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem*., **2013**, *56*, 6521−6530.
14. Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M.-L.; Mergny, J.-L. Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **2014**, *136*, 5249–5252.
15. De Rache, A.; Marquevielle, J.; Bouaziz, S.; Vialet, B.; Andreola, M. L.; Mergny, J.-L.; Amrane, S. Structure of a DNA G-quadruplex that modulates Sp1 binding sites architecture in HIV-1 promoter. *J. Mol. Biol.*, **2023**, *436*, 168359.
16. Butovskaya, E.; Heddi, B.; Bakalar, B.; Richter, S. N.; Phan, A. T. Major G-quadruplex form of HIV-1 LTR reveals a (3 + 1) folding topology containing a stem-loop. *J. Am. Chem. Soc.*, **2018,** *140*, 13654–3662.

17. De Nicola, B.; Lech, C. J., Heddi, B., Regmi, S., Frasson, I., Perrone, R., Richter, S. N.; Phan, A. T. Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.*, **2016**, *44*, 6442–6451.
18. Ruggiero, E.; Frasson, I.; Tosoni, E.; Scalabrin, M.; Perrone, R.; Marušič, M.; Plavec, J.; Richter, S. N. Fused in liposarcoma protein, a new player in the regulation of HIV-1 transcription, binds to known and newly identified LTR G-quadruplexes. *ACS Infect. Dis.*, **2022**, *8*, 958–968.
19. Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palù, G.; Fabris, D.; Richter, S. N. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.*, **2015**, *43*, 8884–8897.
20. Scalabrin, M.; Frasson, I.; Ruggiero, E.; Perrone, R.; Tosoni, E.; Lago, S.; Tassinari, M.; Palù, G.; Richter, S. N. The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.*, **2017**, *7*, 45244.
21. Sheng, Y.; Cao, B.; Ou, M. X.; Wang, Y.; Yuan, S. M.; Zhang, N.; Zou, T. T.; Liu Y. Z. Nucleocapsid protein preferentially binds the stem-loop of duplex/quadruplex hybrid that unfolds the quadruplex structure. *Chem. Commun.*, **2021**, *57*, 5298–5301.
22. Perrone, R.; Lavezzo, E.; Palù, G.; Richter, S. N. Conserved presence of G-quadruplex forming sequences in the long terminal repeat promoter of Lentiviruses. *Sci. Rep.*, **2017**, *7*, 2018.
23. Vorlíčková M.; Kejnovská I.; Sagi J.; Renčiuk D.; Bednářová K.; Motlová J.; Kypr J. Circular dichroism and guanine quadruplexes. *Methods*, **2012**, *57*, 64–75.
24. Darby, R. A.; Sollogoub, M.; McKeen, C.; Brown, L.; Risitano, A.; Brown, N.; Barton, C.; Brown, T.; Fox, K. R. High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.*, **2002**, *30*, e39.
25. Vummidi, B. R.; Alzeer J.; Luedtke, N. W. Fluorescent probes for G-quadruplex structures. *ChemBioChem.*, **2013**, *14*, 540–558.
26. Suseela, Y. V.; Narayanaswamy, N.; Pratihara, S.; Govindaraju, T. Far-red fluorescent probes for canonical and non-canonical nucleic acid structures: current progress and future implications. *Chem. Soc. Rev.*, **2018**, *47*, 1098–1131.
27. Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.*, **2013**, *135*, 367–376.
28. Sun, D.; Hurley, L. H. Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.*, **2010**, *608*, 65–79.
29. Adrian, M.; Heddi, B.; Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods*, **2012**, *57*, 11–24.
30. Haider, S. M.; Parkinson, G. N.; Neidle, S. Structure of a G-quadruplex-ligand complex. *J. Mol. Biol*., **2003**, *326*, 117–125.
31. Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. S. G-quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions. *Pharmaceuticals*, **2021**, *14*, 769.
32. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **2013**, *5*, 182–186.
33. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **2014**, *6*, 75–80.

34. Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Jr, Sen, D.; Lansdrop, P. M. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.,* **2014**, *42*, 860–869.
35. Galli, S.; Melidis, L.; Flynn, S. M.; Varshney, D.; Simeone, A.; Spiegel, J.; Madden, S. K.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex recognition in vitro and in live cells by a structure-specific nanobody. *J. Am. Chem. Soc.,* **2022**, *144*, 23096–23103.
36. Doria, F.; Nadai, M.; Zuffo, M.; Perrone, R.; Freccero, M.; Richter, S. N. A red-NIR fluorescent dye detecting nuclear DNA G-quadruplexes: *in vitro* analysis and cell imaging. *Chem. Commun.,* **2017**, *53*, 2268–2271.
37. Shivalingam, A.; Izquierdo, M. A.; Marois, A. L.; Vyšniauskas, A.; Suhling, K.; Kuimova, M. K.; Vilar, R. The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.,* **2015**, ***6***, 8178–8187.
38. Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M. Y.; Monchaud, D. Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc*., **2015**, ***137***, 8521–8525.
39. Robinson, J.; Stenspil, S. G.; Maleckaite, K.; Bartlett, M.; Antonio, M. D.; Vilar, R.; Kuimova, M. K. Cellular visualization of G-quadruplex RNA via fluorescence-lifetime imaging microscopy. *J. Am. Chem. Soc.*, **2024**, *146*, 1009–1018.
40. Yuan, J.-H.; Shao, W.; Chen, S.-B.; Huang, Z.-S.; Tan, J.-H. Recent advances in fluorescent probes for G-quadruplex nucleic acids. *Biochem. Biophys. Res. Commun.*, **2020**, *531*, 18–24.
41. Pandith, A.; Nagarajachari, U.; Siddappa, R. K. G.; Lee, S.; Park, C.; Sannathammegowda, K.; Seo, Y. J. Loop-mediated fluorescent probes for selective discrimination of parallel and antiparallel G-quadruplexes. *Bioorg. Med. Chem.,* **2021**, *35*, 116077.
42. Vo T.; Oxenford, S.; Angell, R.; Marchetti, C.; Ohnmacht, S. A.; Wilson, W. D.; Neidle, S. Substituted naphthalenediimide compounds bind selectively to two human quadruplex structures with parallel topology. *ACS Med. Chem. Lett.*, **2020**, ***11***, 991–999.
43. Kumar, S.; Pany, S. P. P.; Sudhakar, S.; Singh, S. B.; Todankar, C. S.; Pradeepkumar, P. I. Targeting parallel topology of G-quadruplex structures by indole fused quindoline scaffolds. *Biochemistry*, **2022**, ***61***, 2546–2559.
44. Takahashi, S.; Kotar, A.; Tateishi-Karimata, H.; Bhowmik, S.; Wang, Z. F.; Chang, T. C.; Sato, S.; Takenaka, S.; Plavec, J.; Sugimoto, N. Chemical modulation of DNA replication along G-quadruplex based on topology-dependent ligand binding. *J. Am. Chem. Soc.*, ***2021***, *143*, 16458–16469.
45. Calabrese, D. R.; Chen, X.; Leon, E.; Gaikwad M. S.; Phyo, Z.; Hewitt, M. W.; Alden, S.; Hilimire A. T.; He, F.; Michalowski, M. A. et al. Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat Commun.*, ***2018***, *9*, 4229.
46. Rigo, R.; Groaz, E.; Sissi, C. Polymorphic and higher-order G-quadruplexes as possible transcription regulators: novel perspectives for future anticancer therapeutic applications. *Pharmaceuticals*, **2022**, *15*, 373.
47. Grün, J. T.; Schwalbe, H. Folding dynamics of polymorphic G-quadruplex structures. *Biopolymers*, **2022,** *113*, e23477.
48. Manna, S.; Sarkar, D.; Srivatsan, S. G. A dual-app nucleoside probe provides structural insights into the human telomeric overhang in live cells. *J. Am. Chem. Soc.,* **2018**, *140*, 12622–12633.

49. 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.
50. Tanpure A. A.; Srivatsan S. G. Conformation-sensitive nucleoside analogues as topology-specific fluorescence turn-on probes for DNA and RNA G-quadruplexes. *Nucleic Acids Res.,* 2015, **43**, e149.
51. Cheng, M.; Cheng, Yu.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.,* **2018**, *46*, 9264–9275.
52. Jana, J.; Vianney, Y. M.; Schröder, N.; Weisz, K. Guiding the folding of G-quadruplexes through loop residue interactions. *Nucleic Acids Res.,* **2022**, *50*, 7161–7175.
53. Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A. Positional impact of fluorescently modified G-tetrads within polymorphic human telomeric G-quadruplex structures. *ACS Chem. Biol.*, **2015**, *10*, 1311–1318.
54. Manna, S.; Srivatsan, S. G. Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.,* **2018**, *8*, 25673–25694.
55. Olejniczak, M.; Gdaniec, Z.; Fischer, A.; Grabarkiewicz, T.; Bielecki, L.; Adamiak, R. W. The bulge region of HIV-1 TAR RNA binds metal ions in solution. *Nucleic Acids Res.,* **2002**, *30*, 4241–4249.
56. Tanabe, K.; Sugiura, M.; Nishimoto, S. Monitoring duplex and triplex formation by ^{19}F NMR using oligodeoxynucleotides possessing 5-fluorodeoxyuridine unit as ^{19}F signal transmitter. *Bioorg. Med. Chem.*, **2010**, *18*, 6690–6694.
57. Sacre, L.; Wilds, C. J. *Eur. J. Org. Chem.* **2017**, 3003–3008.
58. Pandey, A.; Roy, S.; Srivatsan, S. G. *Chem. Asian J.* **2023**, *18*, e20230051.
59. Do, N. Q.; Phan, A. T. Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry,* **2012**, *18*, 14752–14874.
60. Hänsel, R.; Löhr, F.; Foldynová-Trantírková, S.; Bamberg, E.; Trantírek, L.; Dötsch, V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.,* **2011**, *39*, 5768–5775.
61. Gao, C.; Mohamed, H. I.; Deng, J.; Umer, M.; Anwar, N.; Chen, J.; Wu, Q.; Wang, Z.; He, Y. Effects of molecular crowding on the structure, stability, and interaction with ligands of G-quadruplexes. *ACS Omega*, **2023**, *8*, 14342–14348.
62. Doose, S.; Neuweiler, H.; Sauer, M. Fluorescence quenching by photoinduced electron transfer: A reporter for conformational dynamics of macromolecules. *ChemPhysChem*, **2009**, *10*, 1389–1398.
63. Hayatshahi, H. S.; Henriksen, N. M.; Cheatham, T. E. Consensus conformations of dinucleoside monophosphates described with well-converged molecular dynamics simulations. *J. Chem. Theory Comput.*, **2018**, *14*, 1456–1470.
64. Greco, N. J.; Tor, Y. Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. *J. Am. Chem. Soc.,* **2005**, *127*, 10784–10785.
65. Riedl, J.; Pohl, R.; Rulíšek, L.; Hocek, M. Synthesis and photophysical properties of biaryl-substituted nucleos(t)ides. Polymerase synthesis of DNA probes bearing solvatochromic and pH-sensitive dual fluorescent and ^{19}F NMR labels. *J. Org. Chem.*, **2012**, *77*, 1026–1044.
66. Sapper, H.; Lohmann, W. Stacking interaction of nucleobases: NMR investigations III. Molecular aspects of the solvent dependence. *Biophys. Struct. Mechanism,* **1978**, *4*, 327–335.

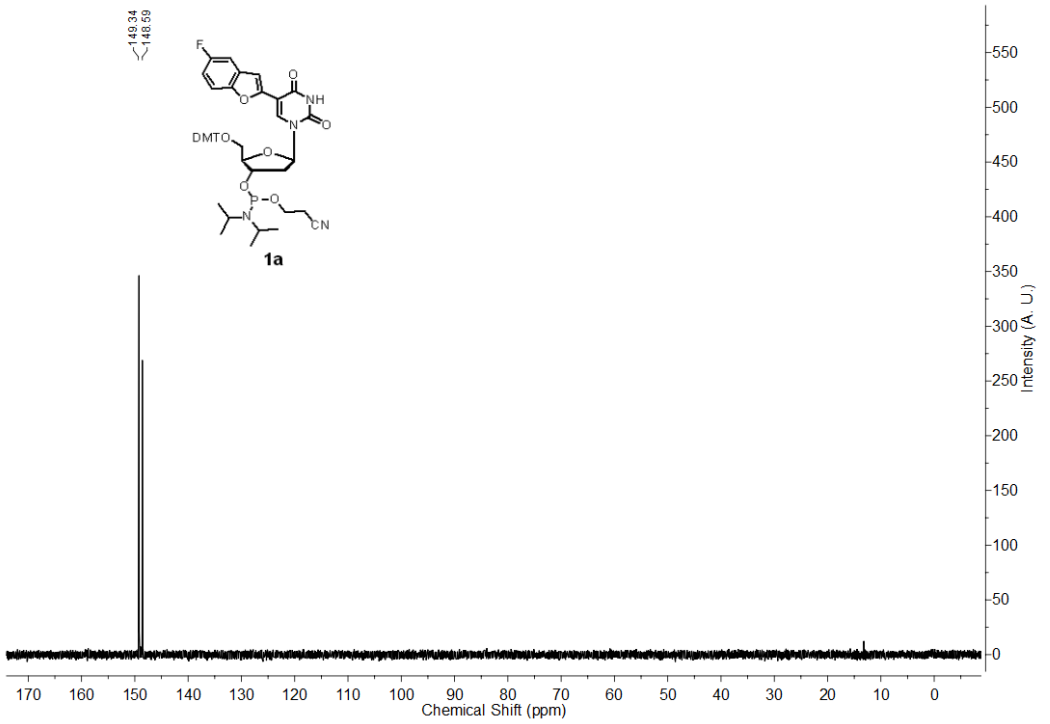
67. Giam, C. S.; Lyle, J. L. Medium effects on the fluorine-19 magnetic resonance spectra of fluoropyridines. *J. Am. Chem. Soc.*, **1973**, *95*, 3235–3239.
68. Dahanayake, J. N.; Kasireddy, C.; Karnes, J. P.; Verma, R.; Steinert, R. M.; Hildebrandt, D.; Hull, O. A.; Ellis, J. M.; Mitchell-Koch, K. R. Progress in our understanding of ^{19}F chemical shifts. *Annu. Rep. NMR Spectrosc.*, **2018**, *93*, 281–365.
69. Le, D. D.; Antonio, M. D.; Chan, L. K. M.; Balasubramanian, S. G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun.*, **2015**, ***51***, 8048–8050.
70. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of human telomere RNA G-quadruplex structures in vitro and in living cells using ^{19}F NMR spectroscopy. *Nucleic Acids Res*., **2017**, *45*, 5501–5511.
71. Bao, H.-L.; Masuzawa, T.; Oyoshi, T.; Xu, Y. Oligonucleotides DNA containing 8-trifluoromethyl-2′-deoxyguanosine for observing Z-DNA structure. *Nucleic Acids Res*., **2020**, *48*, 7041–7051.
72. 70. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. ^{19}F NMR: a valuable tool for studying biological events. *Chem. Soc. Rev.*, **2013**, *42*, 7971–7982.
73. Gimenez, D.; Phelan, A.; Murphy, C. D.; Cobb, S. L. ^{19}F NMR as a tool in chemical biology. *Beilst. J. Org. Chem.*, **2021**, *17*, 293–318.
74. Xu, Y. In Cell ^{19}F NMR for G-quadruplex. Handbook of Chemical Biology of Nucleic Acids Springer Nature. (2023)
75. Barhate, N. B.; Barhate, R. N.; Cekan, P.; Drobny, G.; Sigurdsson, S. Th. A Nonafluoro nucleoside as a sensitive ^{19}F NMR probe of nucleic acid conformation. *Org. Lett.*, **2008**, ***10***, 2745–2747.
76. Puffer, B.; Kreutz, C.; Rieder, U.; Ebert, M.-O.; Konrat, R.; Micura, R. 5-Fluoro pyrimidines: labels to probe DNA and RNA secondary structures by 1D ^{19}F NMR spectroscopy. *Nucleic Acids Res.,* **2009**, *37*, 7728–7740.
77. Li,Q.; Chen, J.; Trajkovski, M.; Zhou, Y.; Fan, C.; Lu, K.; Tang, P.; Su, X.; Plavec, J.; Xi, Z.; Zhou, C. 4′-Fluorinated RNA: synthesis, structure, and applications as a sensitive 19F NMR probe of RNA structure and function. J. Am. Chem. Soc., 2020, 142, 4739–4748.
78. Baranowski, M. R.; Warminski, M.; Jemielity, J.; Kowalska, J. 5′-fluoro(di)phosphate-labeled oligonucleotides are versatile molecular probes for studying nucleic acid secondary structure and interactions by ^{19}F NMR. *Nucleic Acids Res.*, **2020**, *48*, 8209–8224.
79. Pham, L. B. T.; Costantino, A.; Barbieri, L.; Calderone, V.; Luchinat, E.; Banci, L. Direct expression of fluorinated proteins in human cells for ^{19}F in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **2023**, *145*, 1389–1399.
80. Hänsel, R.; Foldynová-Trantírková, S.; Löhr, F.; Buck, J.; Bongartz, E.; Bamberg, E.; Schwalbe, H.; Dötsch, V.; Trantírek, L. Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. *J. Am. Chem. Soc.*, **2009**, *131*, 15761–15768.
81. Krafcikova, M.; Hänsel-Hertsch, R.; Trantirek, L.; Foldynova-Trantirkova, S. In cell NMR spectroscopy: investigation of G-quadruplex structures inside living *Xenopus laevis* oocytes. In *Methods in Molecular Biology*; Springer: Berlin, Germany, **2019**, Vol. 2035, pp. 397–405.
82. Gaussian 16, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M.

Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

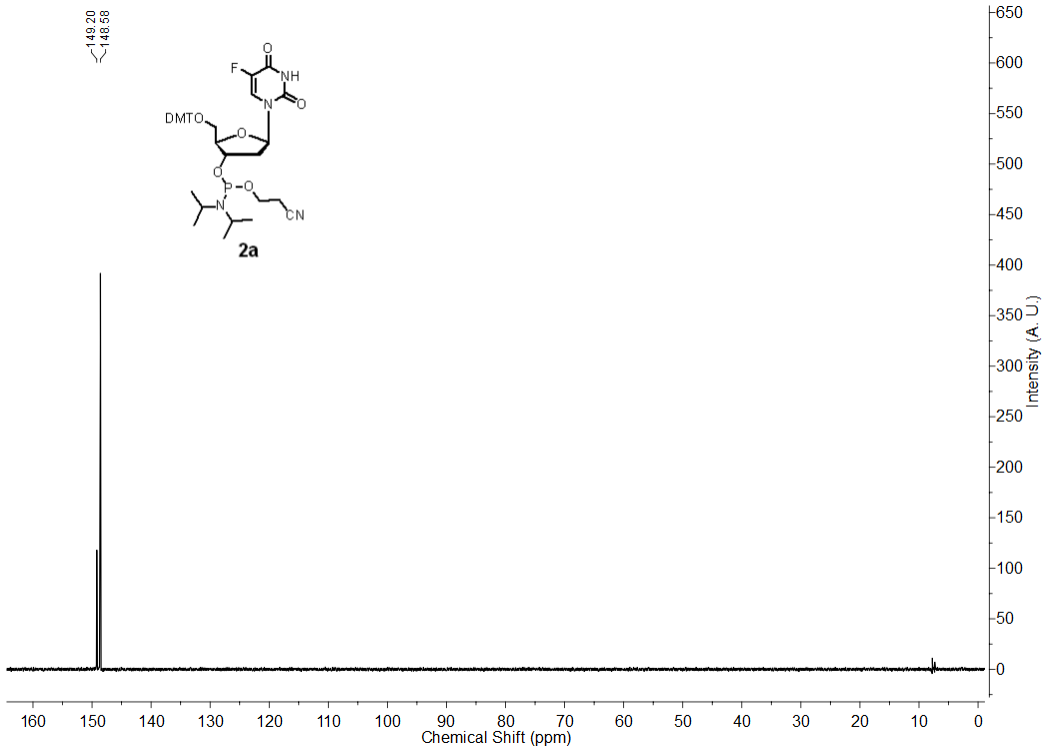
83. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
84. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
85. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer,J.; Otyepka, M.; Jurecka,ˇ P.; Cheatham, T. E. Assessing the current state of Amber Force Field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114−4127.
86. Kumar, S.; Pany, S.P.; Todankar, C. S.; Sudhakar, S.; Singh, S. B.; Todankar, C. S.; Pradeepkumar, P.I. Targeting parallel topology of G-quadruplex structures by indole-fused quindoline scaffolds. *Biochemistry* **2022**, *61*, 2546−2549.
87. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput*. **2013**, *9*, 3878−3888.
88. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
89. Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
90. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K. AMBER 2018; Univ. California: San Fr, 2018; Vol. 2018.
91. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
92. Humphrey, W.; Dalke, A.; Schulten, K. "VMD - Visual Molecular Dynamics". *J. Molec. Graphics* **1996**, *14*, 33−38.

2.6 Appendix-I (NMR, HPLC, ESI-MS spectra)

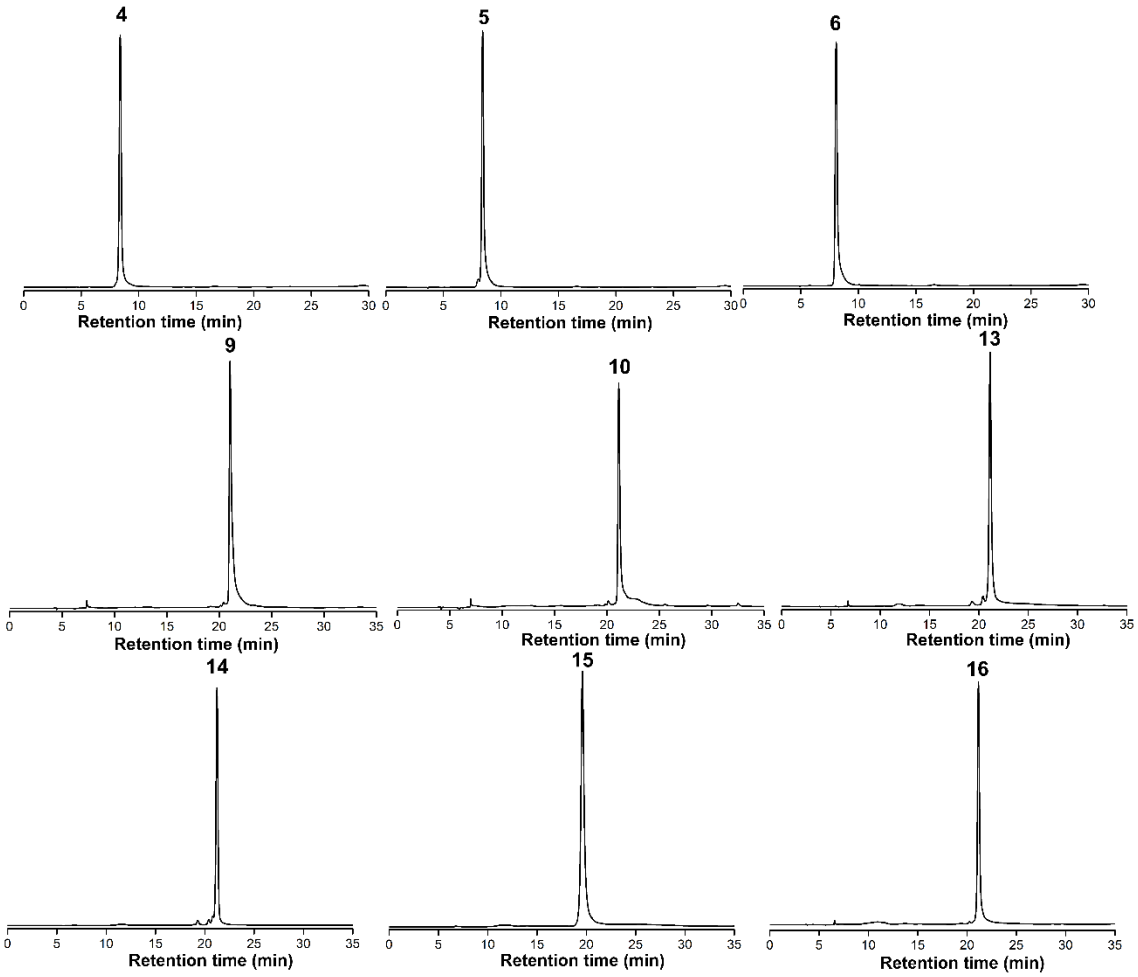
^{31}P NMR of **1a** (162 MHz, $CDCl_3$)



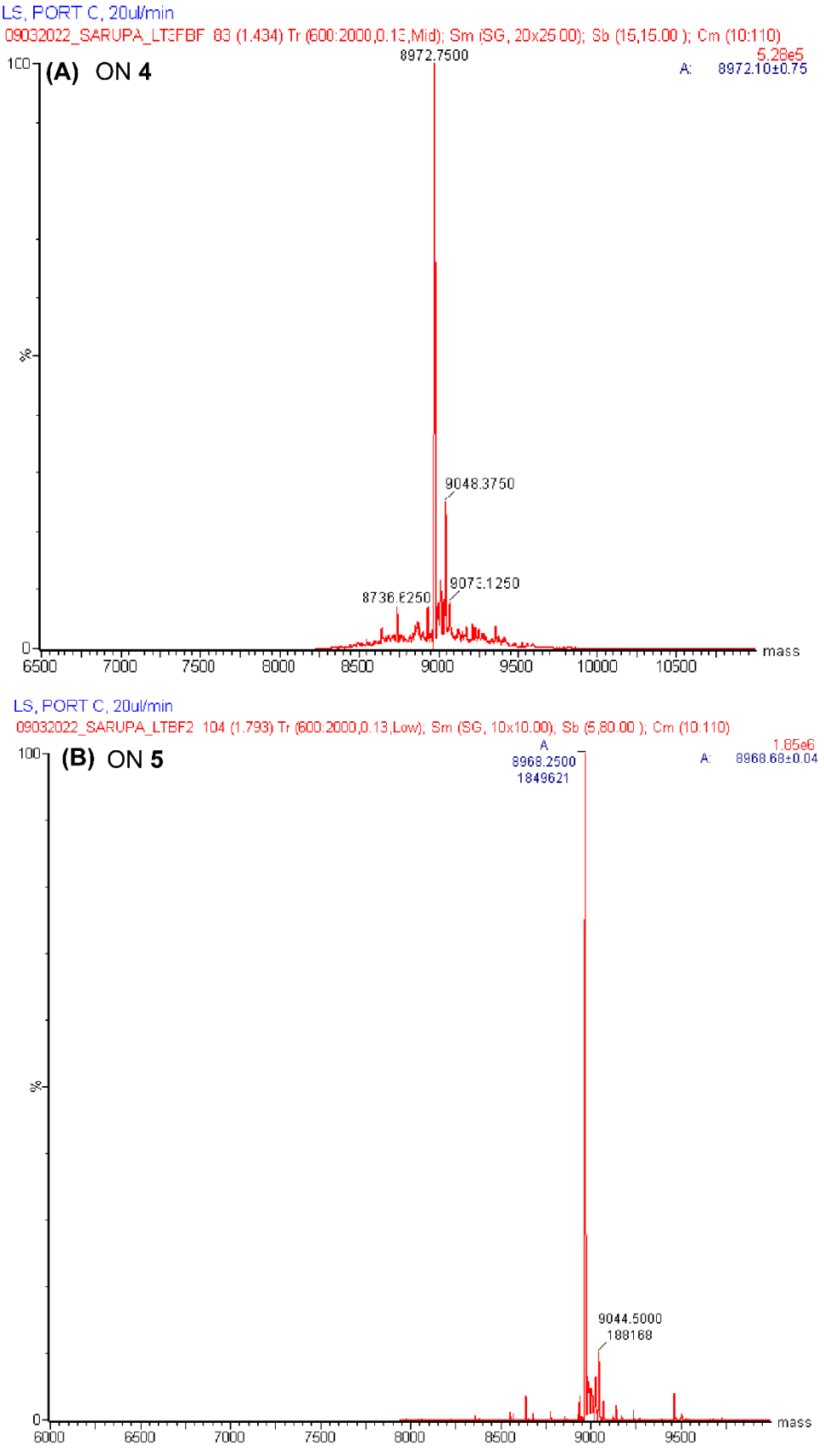
^{31}P NMR of **2a** (162 MHz, $CDCl_3$)

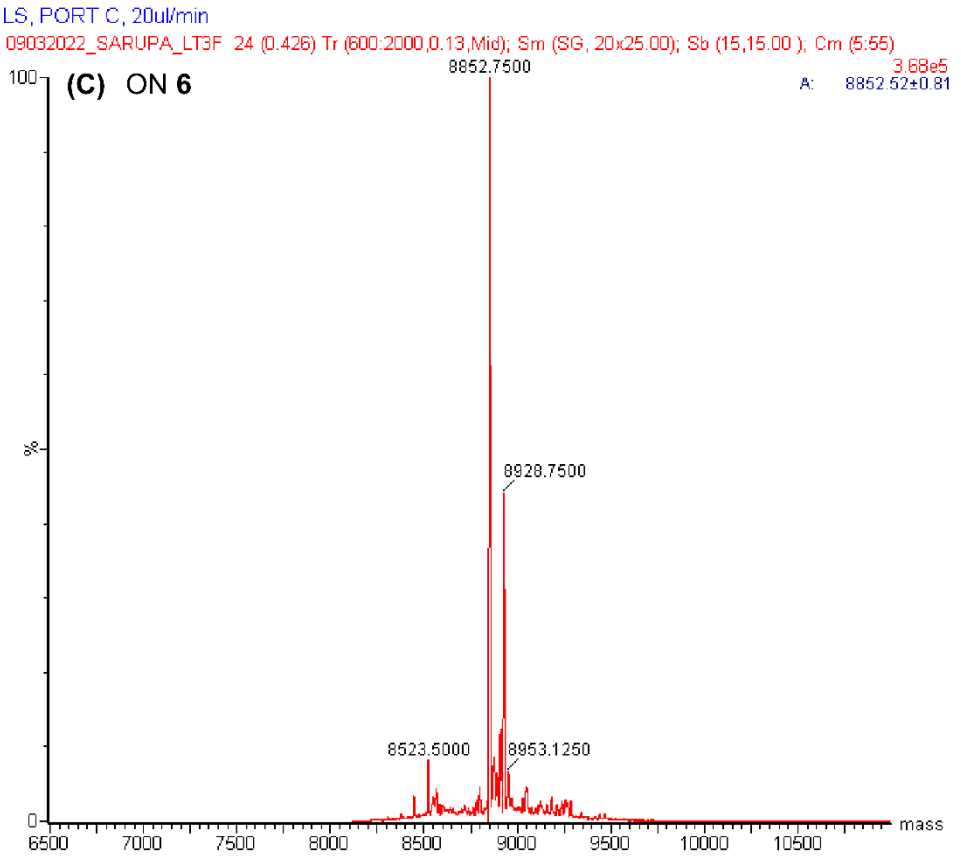


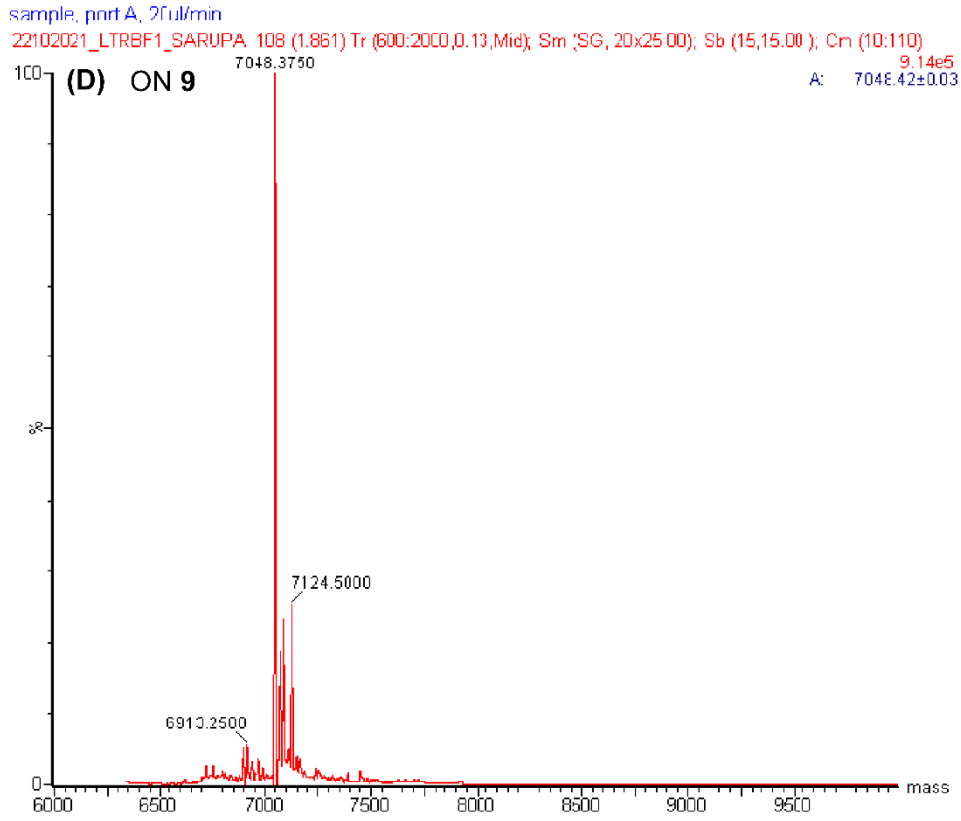
The purity of PAGE purified ONs containing the modification was analysed by RP-HPLC at 260 nm using Luna C18 column (250 x 4.6 mm, 5 micron). Mobile phase A= 0.5 mM triethylammonium acetate (pH=7.3) and B= acetonitrile. Gradient: 0–100 % B in 30 min with a flow rate of 1 mL/min was used for ONs **4**–**6**. Gradient: 0–30 % B in 40 min and 30–100% in 10 min with a flow rate of 1 mL/min was used for ONs **9**, **10**, **13**–**16**

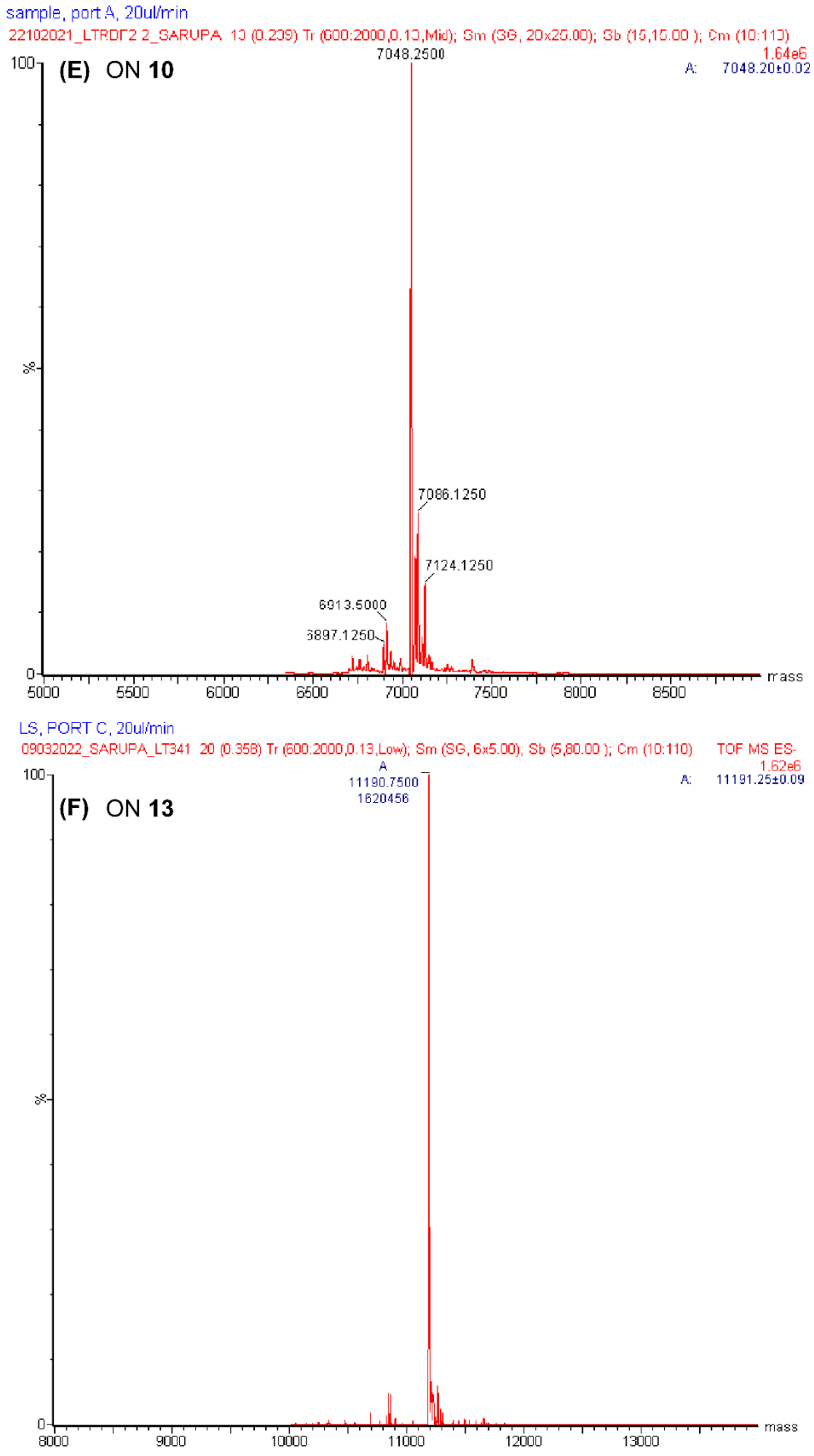


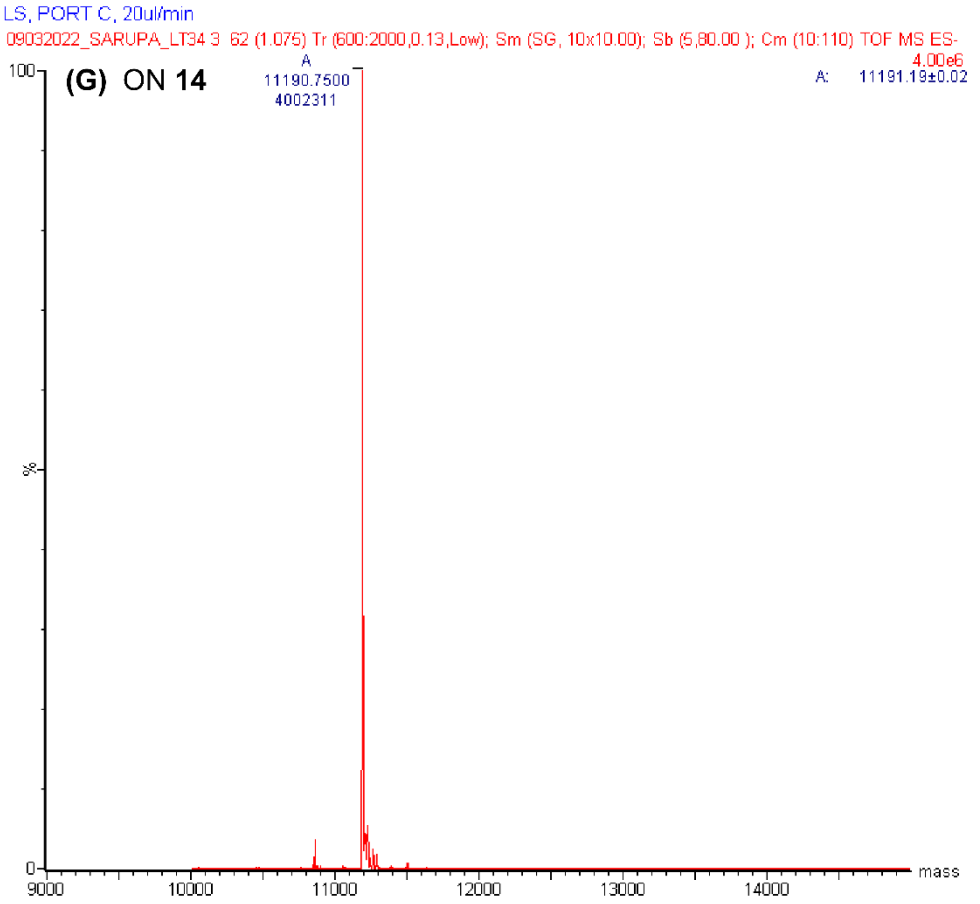
ESI-MS of purified LTR ONs containing FBFdU and or FdU. See Table A1 for details.

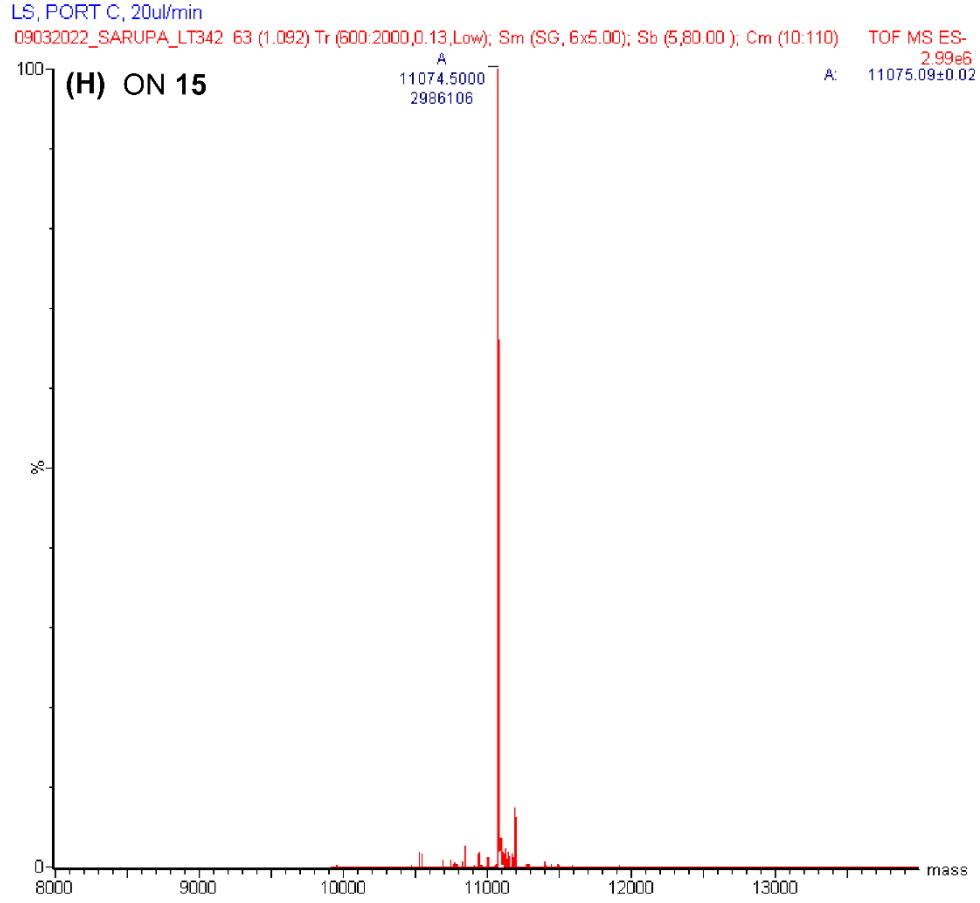



LS, PORT C, 20ul/min
09032022_SARUPA_LT3F 24 (0.426) Tr (600:2000,0.13,Mid); Sm (SG, 20x25.00); Sb (15,15.00); Cm (5:55)
(C) ON 6
3.68e5
A: 8852.52±0.81
8852.7500
8928.7500
8523.5000
8953.1250
mass


sample, port A, 20ul/min
22102021_LTRBF1_SARUPA 108 (1.861) Tr (600:2000,0.13,Mid); Sm (SG, 20x25.00); Sb (15,15.00); Cm (10:110)
(D) ON 9
9.14e5
A: 7048.42±0.03
7048.3750
7124.5000
6910.2500
mass


sample, port A, 20ul/min
22102021_LTRDF2 2_SARUPA 10 (0.239) Tr (600:2000,0.10,Mid); Sm (SG, 20x25.00); Sb (15,15.00); Cm (10:110)
(E) ON 10
7048.2500
1.64e6
A: 7048.20±0.02
7086.1250
7124.1250
6913.5000
3897.1250
100
%
0
5000
5500
6000
6500
7000
7500
8000
8500
mass
LS, PORT C, 20ul/min
09032022_SARUPA_LT341 20 (0.358) Tr (600:2000,0.13,Low); Sm (SG, 6x5.00); Sb (5,80.00); Cm (10:110)
TOF MS ES-
A
11190.7500
1620456
1.62e6
A: 11191.25±0.09
(F) ON 13
100
%
0
8000
9000
10000
11000
12000
13000
mass


LS, PORT C, 20ul/min
09032022_SARUPA_LT34 3 62 (1.075) Tr (600:2000,0.13,Low); Sm (SG, 10x10.00); Sb (5,80.00); Cm (10:110) TOF MS ES-
4.00e6
(G) ON 14
A
11190.7500
4002311
A: 11191.19±0.02
100
%
0
9000
10000
11000
12000
13000
14000
mass


LS, PORT C, 20ul/min
09032022_SARUPA_LT342 63 (1.092) Tr (600:2000,0.13,Low); Sm (SG, 6x5.00); Sb (5,80.00); Cm (10:110)
TOF MS ES-
2.99e6
(H) ON 15
A
11074.5000
2986106
A: 11075.09±0.02
100
%
0
8000
9000
10000
11000
12000
13000
mass

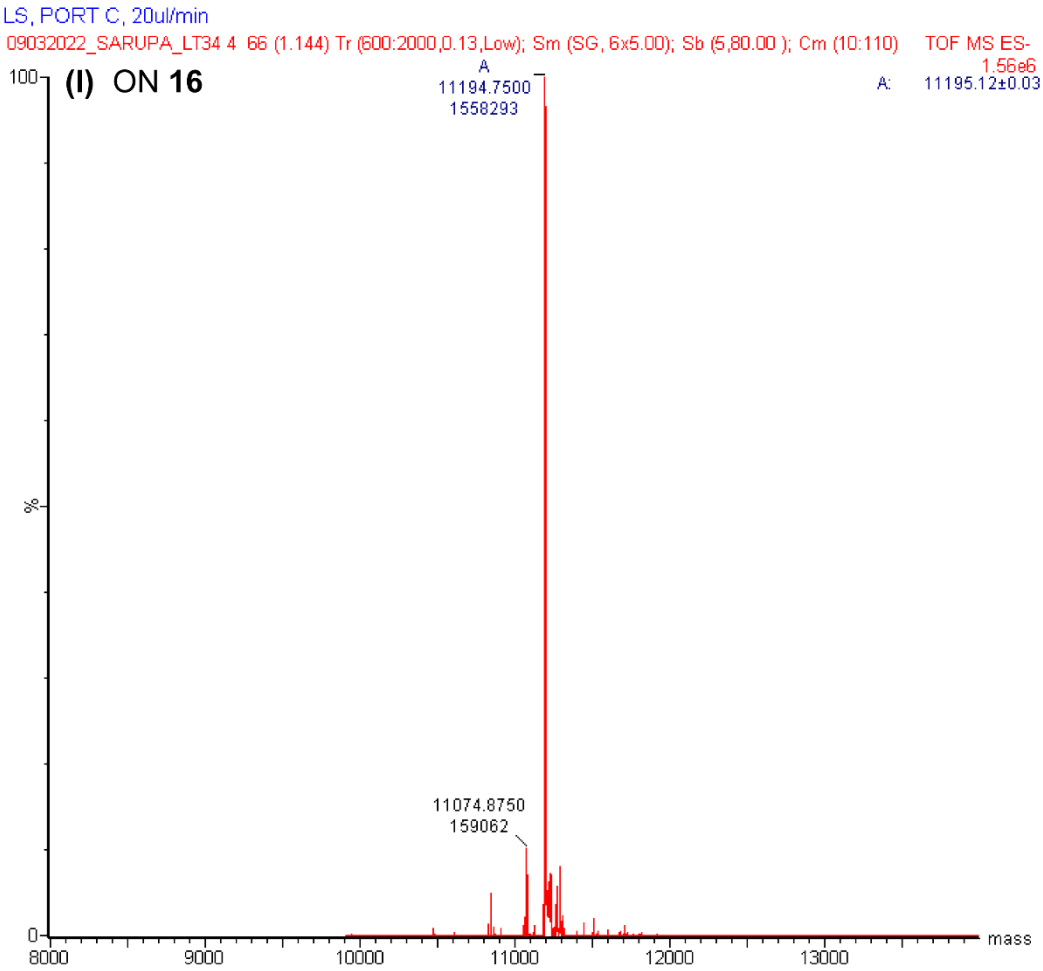


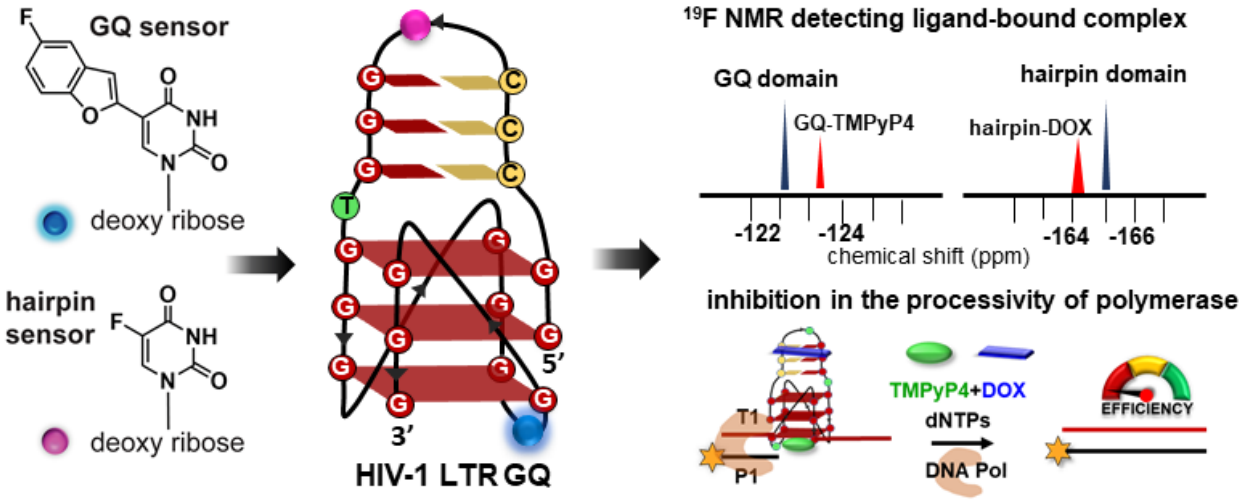
Table A1. ε_{260} and mass data of modified LTR ONs

ON sequence	ε_{260} [M^{-1} cm^{-1}][a]	calculated (g/mol)	found (g/mol)
4	258418	8972.76	8972.75
5	268710	8968.79	8968.25
6	266087	8852.68	8852.75
9	212810	7048.58	7048.37
10	212810	7048.58	7048.25
13	342910	11191.23	11190.75
14	342910	11191.23	11190.75
15	348117	11075.11	11074.50
16	341896	11195.19	11194.75

[a]Molar absorption coefficient (ε) of modified ONs was determined by using OligoAnalyzer 3.1. The molar absorptivity of modified nucleosides FBFdU (ε_{260}= 10310 M^{-1} cm^{-1}) and FdU (ε_{260}= 7687 M^{-1} cm^{-1}) was used in place of dT.

Chapter 3

Probing druggable space of HIV-1 LTR G-quadruplexes using responsive ^{19}F-labeled nucleoside probes



***Chapter 3* is a reprint of part of the data published in the journal:**

Roy, S.; Majee, P.; Sudhkar, S.; Misra, S.; Kalia, J; Pradeepkumar, P. I.; Srivatsan, S. G. *Chem Sci.,* **2024**, *15,* 7982–7991.

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

3.1 Introduction

Given the functional implication of G-quadruplexes (GQs) in regulating key biological processes such as replication, transcription, and translation, targeting these structural motifs that are associated with disease states (e.g., cancer progression) presents a promising therapeutic strategy.[1] However, these motifs are naturally dynamic and prone to unwinding by cellular helicases, which can rapidly resolve them after formation.[2] In this context, ligands—chemical compounds that selectively bind to GQs—are crucial because they stabilize these structures. These secondary structures serve as recognition sites for ligands due to the distinct topologies adopted by different GQ-forming sequences in the human genome.[3] Apart from stabilization, these ligands must demonstrate a high degree of selectivity for quadruplexes over duplex structures possessing several key features a) a π-delocalized electronic system that can stack with the planar G-quartet and b) positively charged groups capable of interacting with the grooves, loops, and negatively charged phosphate backbone of quadruplex DNA (Figure 1). Additionally, these charged groups often improve the solubility of the quadruplex stabilizers[3,4] so that these ligands remain stable under physiological conditions.[5,6] These ligands can bind to GQs through various patterns, including stacking with the outer G-quartets, groove binding, loop binding, and combined binding modes (Figure 2).[4,7]

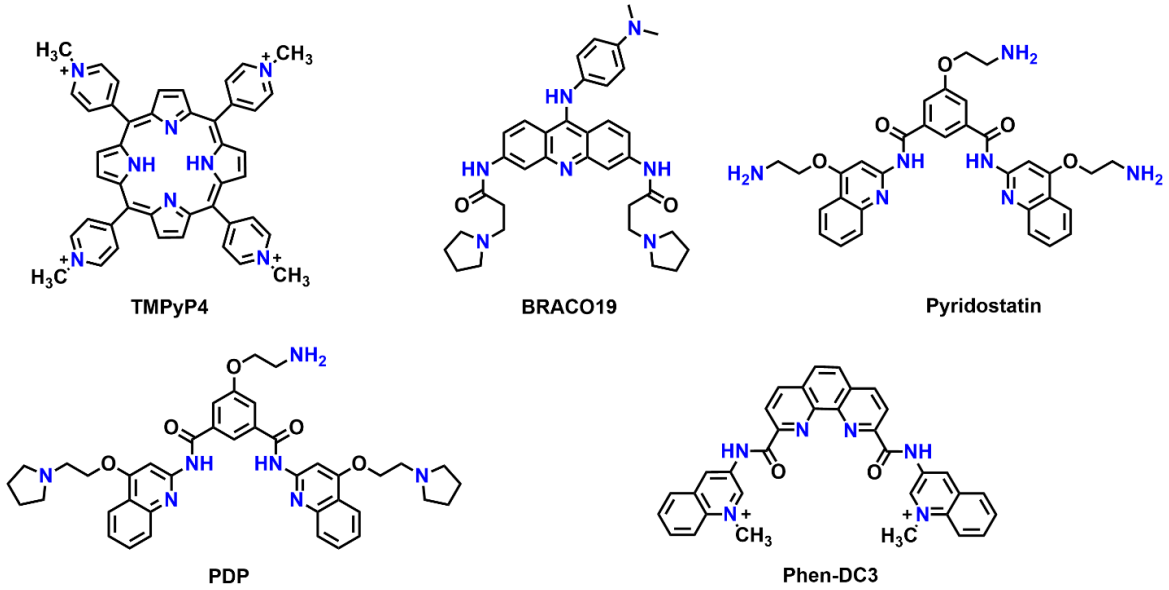


Figure 1. Structures of GQ ligands.

With the growing recognition of GQs and their biological significance, the development of compounds that target these structures has rapidly progressed. Some of these compounds have shown promise as potential anticancer agents with relatively low toxicity.[8] Among them, Quarfloxin, the first GQ-interactive drug, advanced to phase 2 clinical trials,[9] though it failed to demonstrate sufficient efficacy for further clinical development.[10,11] Alongside these advancements, various GQ ligands—such as BRACO19, TMPyP4, Pyridostatin, PhenDC, and their derivatives, as well as quindoline derivatives (Figure 1) have been tested for antiviral efficacy. While these ligands interact with both human and viral GQs, they have shown promising antiviral activity and greater selectivity.[12,13] Examples include ThT-NE, an indicator of HCV GQs[14] and more specialized antiviral agents such as the naphthalene diimide derivative c-exNDI,[15] gamma-PNAs[16], and benzoselenoxanthenes.[17]

The experimental techniques used to study GQ/ligand interactions can be classified into three main categories: structure-based methods, affinity-based approaches, and high-throughput techniques. Structure-based methods, such as circular dichroism (CD),[18] nuclear magnetic resonance (NMR) spectroscopy,[19,20] and X-ray crystallography,[21] provide detailed structural information. Affinity methods including surface plasmon resonance (SPR),[22] isothermal titration calorimetry (ITC),[23] and mass spectrometry (MS),[24] focus on evaluating the binding strength and interaction dynamics. High-throughput methods like fluorescence resonance energy transfer (FRET) melting,[25] GQ-fluorescent intercalator displacement assay (GQ-FID),[26] affinity chromatography,[27] and microarrays,[28] are used for large-scale assessments. Each of these techniques has its own advantages and limitations, making it essential to choose the appropriate method based on the specific biological question being addressed.

To complement these methods and provide a more comprehensive understanding, we employed the microenvironment-sensitive nucleoside probes to determine the GQ/ligand interactions. The probe FBFdU not only facilitates the estimation of binding affinity using fluorescence assays but also, in conjunction with FdU, enables the examination of binding modes with the aid of ^{19}F NMR. This integrated approach helps overcome the limitations of individual techniques by providing additional insights into both the affinity and the binding mode of GQ/ligand interactions.

In this study, we investigated ligand recognition of small molecule ligands with the HIV-1 LTR region, leveraging the responsiveness of the FBFdU/FdU probe system. The probe system

not only helped in mapping the binding interactions of individual LTR-segments, but assisted us to identify the druggable space of the longer LTR region. Since the LTR-III structure contains a hybrid-GQ/hairpin motif, the probes enabled tracking of structure-selective binding of the ligands. Additionally, a polymerase stop assay was utilized to characterize ligand interactions and assess their impact on the LTR replication process.

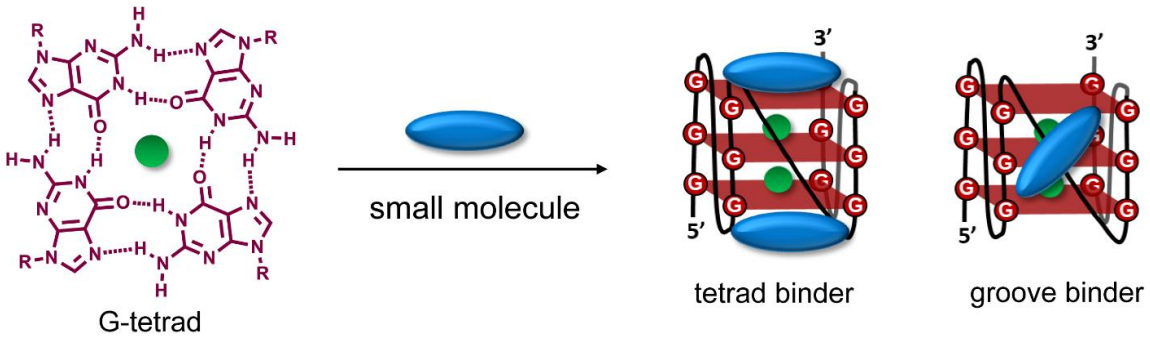


Figure 2. Different binding modes of ligands with GQ.

3.2 Results and Discussion

3.2.1 Probing ligand binding to LTR-III GQ structures by fluorescence and ^{19}F NMR

The modified LTR ONs were synthesized using solid-phase synthesis as discussed in the previous chapter 2 (Table 1).

Table 1: Native and modified LTR ONs used for ligand binding.

G-rich domain	ON[a]	5′----------3′
LTR-III	**4**	GGG AGG CG**2** GGC CTG GGC GGG AC**1** GGG G
LTR-IV	**9**	CTG GGC GGG AC**1** GGG GAG TGG T
LTR-III + IV	**12**	AGG GAG GCG TGG CCT GGG CGG GAC TGG GGA GTG GT
	16	AGG GAG GCG **2**GG CCT GGG CGG GAC TGG GGA G**1**G GT

[a]ONs **4** is native LTR-III modified with FBFdU (**1**) and or FdU (**2**) at T_{24} and T_9, respectively. ONs **9** is native LTR-IV modified with FBFdU (**1**) at T_{12} and T_{19}, respectively. ON **12** is native unmodified ON. ON **16** are native LTR-(III+IV) modified with FBFdU (**1**) and or FdU (**2**) at T_{25} or T_{32} and T_{10}, respectively.

Using the spectral properties of the probes, we evaluated ligand recognition of LTR ONs using two structurally different GQ binders namely, TMPyP4 and BRACO19 (Figure 1). Upon ligand binding to the GQ structure, the fluorescence of FBFdU placed in the GQ domain is known

to diminish significantly because of its proximity to the polyaromatic ligands.[29,30] This effect was used in setting up a fluorescence-binding assay to quantify the affinity of ligands to LTR GQs. LTR-III ON **4** was titrated with increasing concentrations of the ligands and changes in fluorescence were recorded. Titration with TMPyP4 and BRACO19 resulted in a dose-dependent quenching in fluorescence intensity with minimum changes in the emission maximum (Figure 3A and 3B). A plot of normalized fluorescence intensity *versus* ligand concentration fitted to the Hill equation gave an apparent K_d value of 0.28 ± 0.05 µM and 0.56 ± 0.09 µM, respectively (Figure 3C). Whereas, titration with pyridostatin (PDS) did not result in any stepwise quenching and did not result in a saturation-binding isotherm, and hence, reliable K_d values could not be obtained. This is because addition of higher amounts of the ligands interferes with the fluorescence measurement due to background signal.

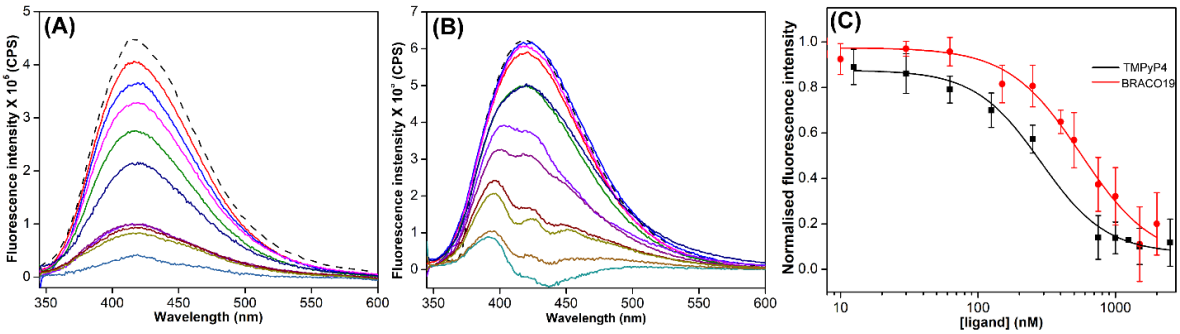


Figure 3. Emission spectra for the titration of labeled LTR-III ON **4** (0.5 µM) with increasing concentration of (**A**) TMPyP4 (12.5 nM–2.5 µM) and (**B**) BRACO19 (10 nM–2.5 µM). Samples were excited at 330 nm with an excitation and emission slit widths of 7 nm and 9 nm, respectively. The dashed line represents the spectrum of ON **4** without any ligand. (**C**). Curve fits for the binding of TMPyP4 and BRACO19 to LTR-III ON **4**. Normalized fluorescence intensity at the emission maximum is plotted against ligand concentration. Values are denoted as mean ± s.d for 3 independent experiments.

^{19}F label of FBFdU efficiently reported the formation of different GQ-ligand complexes with distinct chemical shifts. TMPyP4 and BRACO19 binding to the GQ domain of LTR-III ON **4** produced a new peak at -120.75 ppm and -120.78 ppm, respectively, with a concomitant decrease in the GQ signal (Figure 4A and 4B). Notably, PDS titration with ON **4**, resulted in the emergence of a new peak at -120.78 ppm in the GQ domain (Figure 4C). Thus, ^{19}F NMR facilitated precise monitoring of the interaction dynamics between PDS and ON **4**, which was not achievable through fluorescence spectroscopy alone. Therefore, these techniques serve to complement and reinforce the findings, particularly when one method proves inadequate in detecting the interactions.

Gratifyingly, as these ligands bind preferentially to the G-tetrad, they did not exhibit detectable interaction with the hairpin structure. This is evident from the chemical shift of FdU (-165.73 ppm), placed in the hairpin domain, which remains mostly unchanged throughout the titration experiment.

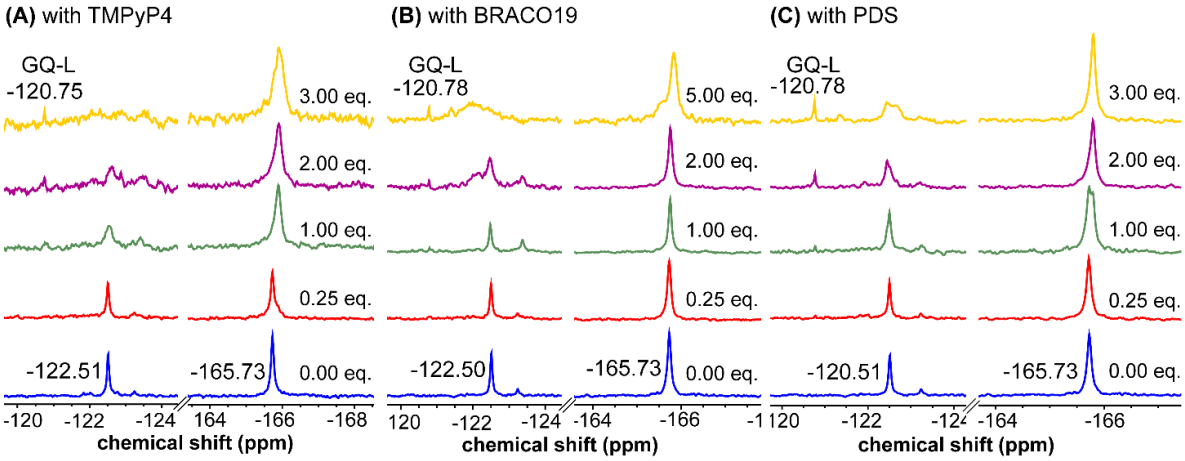


Figure 4. ^{19}F NMR spectra of ON **4** as a function of increasing concentration of ligands. GQ-L represents the peak corresponding to the GQ-ligand complex. (**A**) TMPyP4 (**B**) BRACO19 (**C**) PDS.

3.2.2 Probing ligand binding to LTR-IV GQ structures by fluorescence and ^{19}F NMR

Similarly, the parallel GQ structure of LTR-IV ON **9** titrated with TMPyP4 and BRACO19 resulted in a dose-dependent quenching in fluorescence intensity with no apparent effect on the emission maximum (Figure 5A and 5B). K_d values were found to be 0.33 ± 0.02 µM and 0.39 ± 0.05 µM, respectively (Figure 5C). In this case too, we could not find the isotherm for PDS titration to ON **9**.

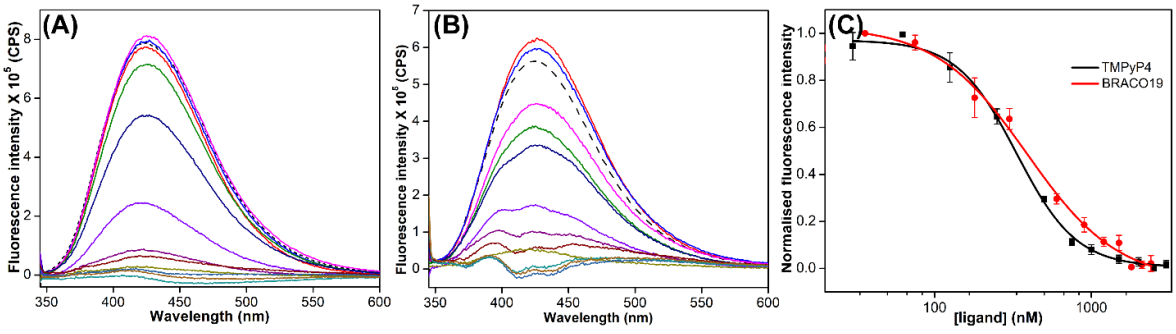


Figure 5. Emission spectra for the titration of labeled LTR-IV ON **9** (0.5 µM) with increasing concentration of (**A**) TMPyP4 (12.5 nM–2.5 µM) and (**B**) BRACO19 (36 nM–2.8 µM)). Samples were excited at 330 nm with an excitation and emission slit widths of 7 nm and 9 nm, respectively.

The dashed line represents the spectrum of ON **9** without any ligand. (**C**). Curve fits for the binding of TMPyP4 and BRACO19 to LTR-IV ON **9.** Normalized fluorescence intensity at the emission maximum is plotted against ligand concentration. Values are denoted as mean ± s.d for 3 independent experiments.

In the same way, ligands binding to the parallel topology of LTR-IV ON **9** exhibited a distinct peak for each complex (TMPyP4: -120.77 ppm, BRACO19: -120.60 ppm and PDS: -120.77, Figure 6A, 6B and 6C). Interestingly, upon ligand binding, multiple GQs formed by **9** coalesce into one ligand-bound form.

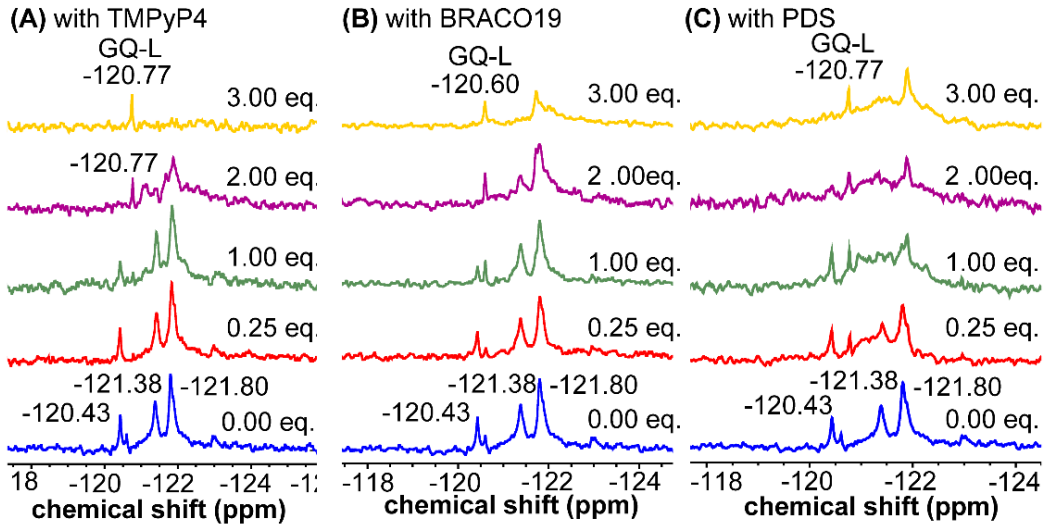


Figure 6. ^{19}F NMR spectra of ON **9** as a function of increasing concentration of ligands. GQ-L represents peak corresponding to GQ-ligand complex. (**A**) TMPyP4 (**B**) BRACO19 (**C**) PDS

3.2.3 Probing structure-specific ligand interaction of juxtaposed GQ-hairpin

As structural evaluation of LTR region in cellular environment reveals the retention in the folding of GQ structure akin to LTR-III, we next tried to stabilize both the motifs GQ and hairpin. Bioinformatics and biophysical experiments suggest the prevalence of quadruplex- hairpin/duplex junctions in different genomes.[31,32] These hybrid structures consist of external double-helical overhangs or internal duplex stem loops, which can be either co-axial or orthogonal with respect to the GQ.[33–36] Further, they feature variations in the size, and conformation of loops that connect GQ and hairpin/duplex domains. Hence, adjacently placed GQ-hairpin/duplex architectures are considered as unique scaffolds that can provide chemical space to target the junction or simultaneously both GQ and duplex elements.[37–41] Bimodal ligands capable of doing the same can potentially have a synergistic functional effect and can significantly enhance the specific

targeting of hybrid GQs as opposed to autonomous GQ structures. In this direction, we evaluated the recognition properties of GQ-hairpin motif of LTR promoter using TMPyP4 (GQ binder) and doxorubicin (DOX, duplex binder) by fluorescence and ^{19}F NMR (Figure 7A). Addition of increasing concentrations of TMPyP4 (30 nM–2.5 μM) to ON **16** (0.5 μM) resulted in a progressive quenching in fluorescence intensity as before and gave an apparent K_d value of 0.52 ± 0.03 μM for the formation of ligand-GQ complex (Figure 7B and 7C). Preferential binding of the ligand to the GQ region was ascertained by ^{19}F NMR. FBFdU placed at the GQ domain responded to increasing concentrations of the ligand, giving rise to a new peak at -120.69 ppm for the complex (Figure 8A). Notably, the chemical shift of FdU (-165.72 ppm) placed in the hairpin domain remained practically unaltered, indicating that TMPyP4 interacts specifically with the GQ structure.

DOX, a well-known DNA duplex intercalator also binds to GQ structures of telomeric repeat, *VEGF*, *Pu22,* and *c-MYC* oncogenes with varying affinities.[42–44] DOX is intrinsically fluorescent and it shows changes in fluorescence upon binding to DNA. To avoid interference from FBFdU, DOX (2 μM) was titrated with a control unmodified ON **12** (2.5 nM–2 μM). We observed a sigmoidal quenching behavior, which gave a K_d value of 0.10 ± 0.02 μM (Figure 7B and 7D). ^{19}F NMR using doubly labeled ON **16** gave better insights into the recognition process. Addition of DOX (1 equiv.) to the ON resulted in the emergence of two new peaks–(i) -120.79 ppm associated to GQ-DOX and (ii) -165.98 ppm associated to hairpin-DOX (Figure 8B). Notably, at a higher equivalent of DOX, the ligand largely occupies the hairpin domain and to some extent the GQ domain. These results highlight the advantage of ^{19}F-labeled nucleoside analogs in probing structure-specific interactions. Based on these observations, we designed a polymerase stop assay to study the inhibitory effect of GQ structure and the ligands independently and in a combination.

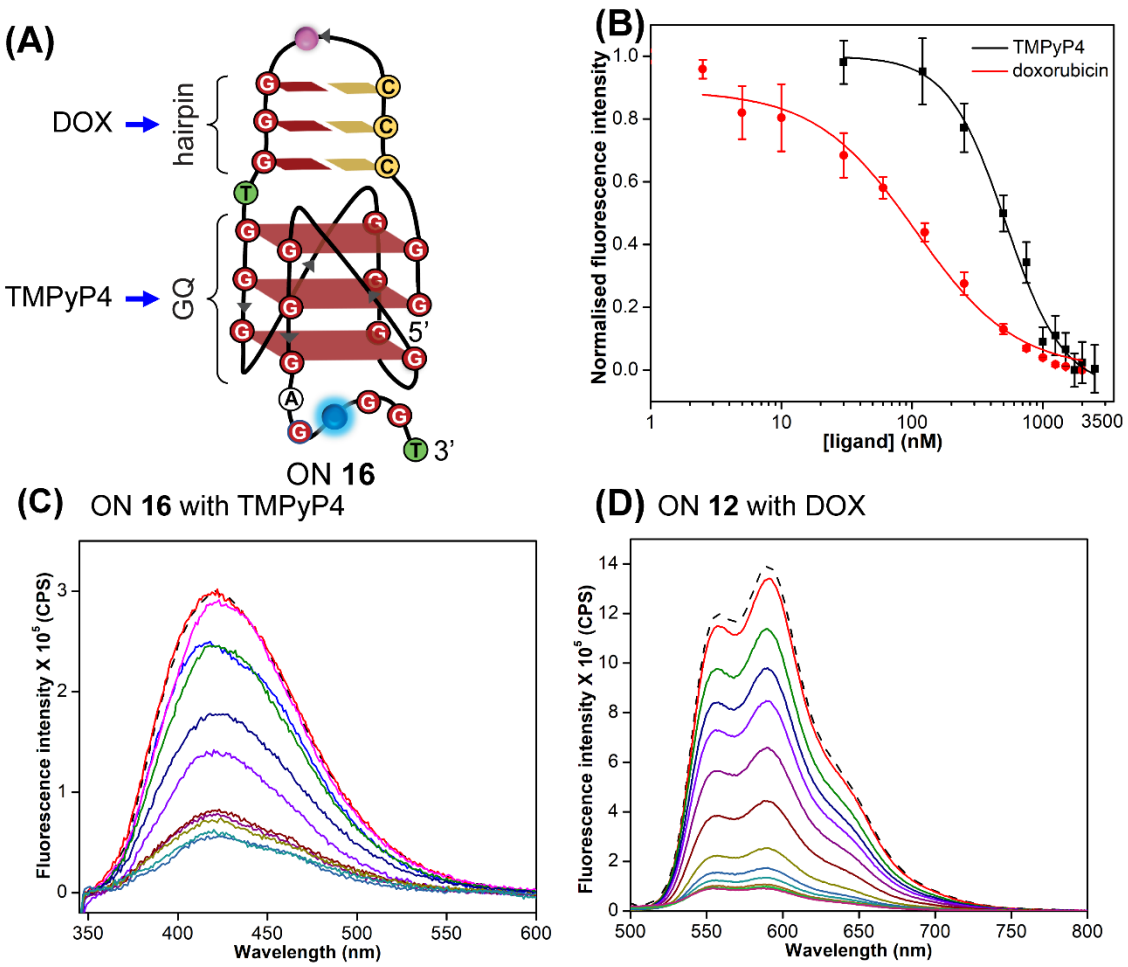


Figure 7. FBFdU and FdU report structure-specific ligand binding to LTR GQ-hairpin structure. (**A**) Schematic representation of the doubly-labeled LTR ON **16** showing the preferred site of ligand interaction (TMPyP4 to GQ and DOX to hairpin). (**B**) Curve fits for the binding of TMPyP4 and DOX to ON **16** and ON **12**, respectively. Normalized fluorescence intensity at the emission maximum of ON **16** (λ_{em}= 421 nm) and DOX (λ_{em}= 590 nm) is plotted against TMPyP4 /ON **12** concentration. Values are denoted as mean ± s.d for 3 independent experiments. (**C**) Emission spectra for the ligand titration of labeled LTR-(III+IV) ON **16** (0.5 µM) with increasing concentration of TMPyP4 (30 nM–2.5 µM). Samples were excited at 330 nm with an excitation and emission slit widths of 7 nm and 9 nm, respectively. The dashed line represents the spectrum of ON **16** without any ligand. (**D**) Emission spectra for the titration of DOX (2 µM) with increasing concentration of control ON **12** (2.5 nM–2.0 µM). Samples were excited at 480 nm with an excitation and emission slit widths of 7 nm and 9 nm, respectively. The dashed line represents the spectrum of DOX without any ON **12**.

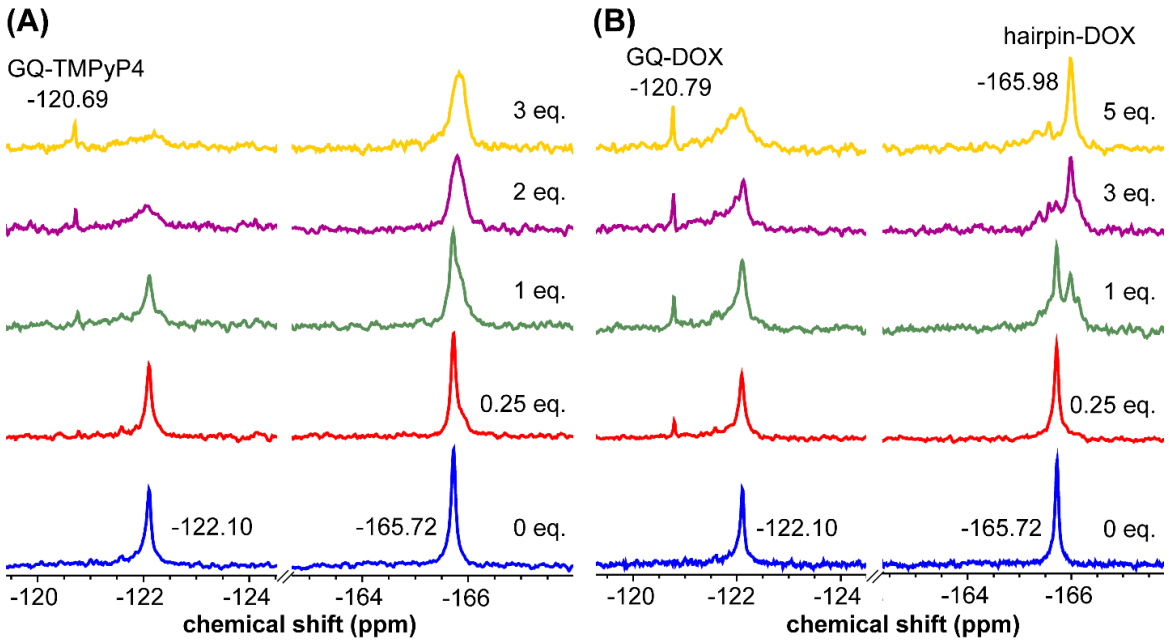


Figure 8. (**A**) and (**B**) ^{19}F NMR spectra of ON **16** as a function of increasing TMPyP4 and DOX concentration, respectively.

3.2.4 TMPyP4-DOX combination exhibits a synergistic inhibitory effect on the viral DNA replication

The effect of GQ structures on DNA polymerase activity was first evaluated by using a native LTR template **T1** encompassing III and IV regions and a mutated template **T2** (does not fold into a GQ, Figure 9A, Table 2). A 5′-FAM labeled primer **P1** was annealed with templates **T1** and **T2** and primer extension reactions were initiated by adding dNTPs and *Taq* DNA polymerase. Reactions were quenched at different time intervals and replication products were resolved by PAGE under denaturing conditions and visualized using a fluorescence scanner. GQ forming template **T1** significantly halted the polymerization process yielding largely stalled products near the GQ site (Figure 10, lanes 2–6, Figure 9B). Longer reaction times (30 min) produced only ~33% of the full-length product. In contrast, reactions in the presence of a non-GQ forming template **T2** produced significant amounts of the full-length product in only 2 min (~45%), which progressively increased to ~80% at 30 min (Figure 10, lanes 7–11, Figure 9B). These observations indicate that the stalling of the primer extension reaction is due to the formation of a stable LTR GQ structure by **T1**.

Table 2. Sequence of templates and primer used in *Taq* DNA polymerase stop assay.

ON	5′----------3′
P1	FAM-GGCAAAAAGCAGCTGCTTATATGCAG
T1	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTGC ATATAAGCAGCTGCTTTTTGCC
T2[a]	TTTTTGGGAGGCGTGGCCTGTGCGTGACTGGGGAGTGGTTTTTCTGCA TATAAGCAGCTGCTTTTTGCC

[a]**T** represents G-T mutation. This mutation does not support GQ formation.[45]

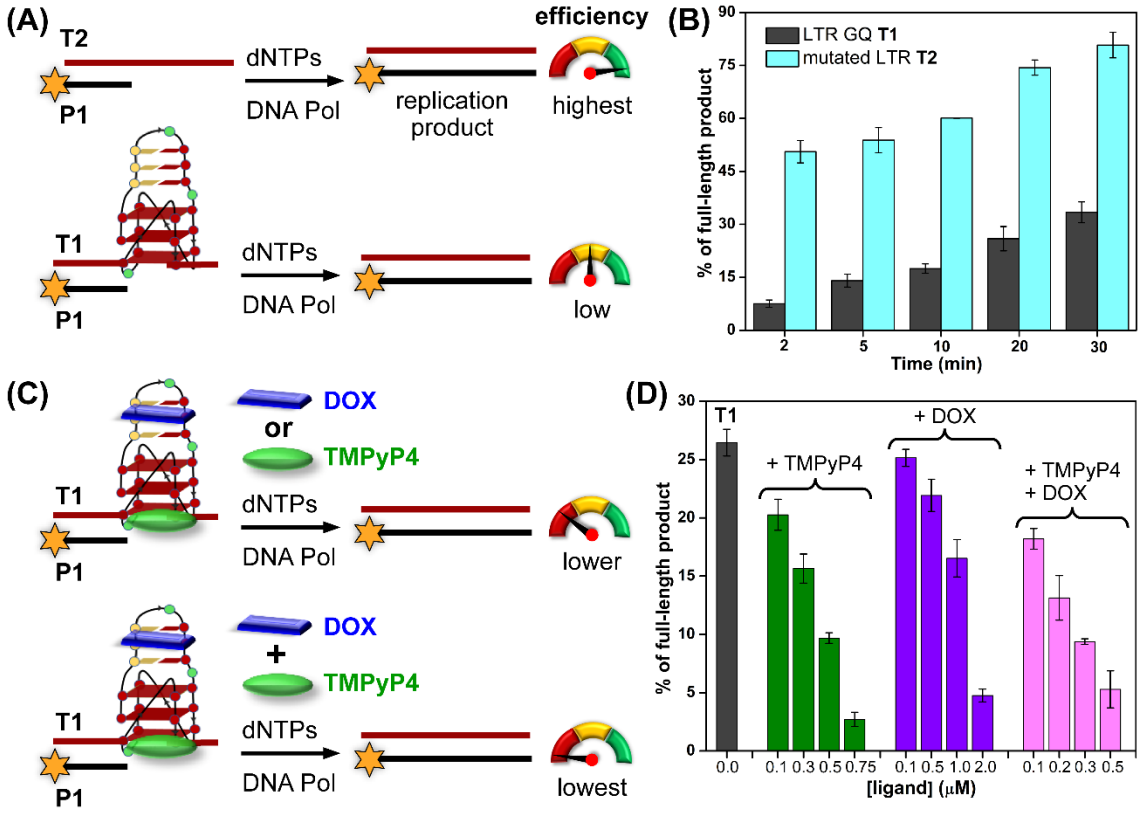


Figure 9. Schematic cartoon representation of primer extension reactions (**A**) with non-GQ forming template **T2** and GQ forming template **T1**, (**C**) in the presence of ligands TMPyP4 or DOX and TMPyP4+DOX. Percentage of the full-length product obtained from *Taq* polymerase reactions. (**B**) Reactions performed using templates **T1** and **T2** at different time intervals. (**D**) Reactions performed using **T1** with increasing concentrations of ligands TMPyP4, DOX and TMPyP4+DOX at 20 min. For gel images, see Figure 10 and 11. Values are denoted as mean ± s.d for 2 independent experiments.

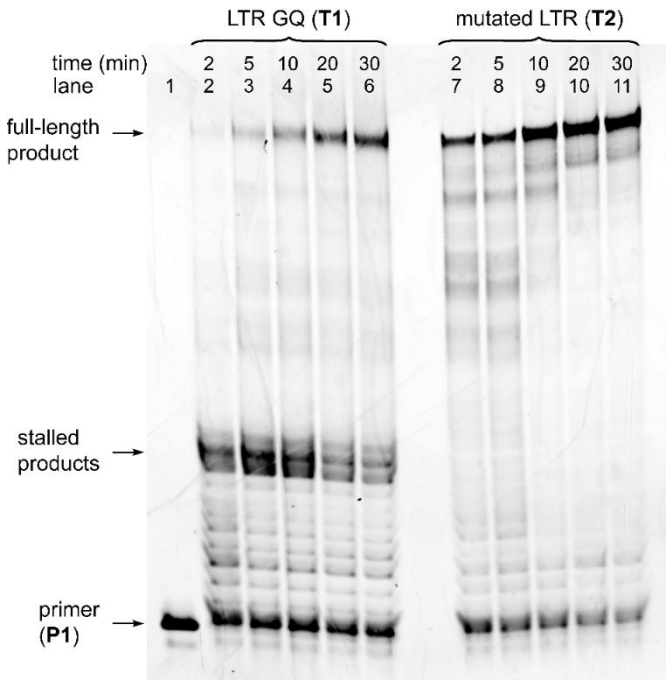


Figure 10. Gel image of primer extension reactions using native LTR G-rich ON template **T1** and mutated LTR template **T2**. See experimental section for details.

Next, we studied the effect of ligand binding to the GQ structure on the polymerase activity using **T1** (Figure 9C). For this purpose, a reaction time of 20 min was chosen as it gave reasonable amounts of the full-length product (~25%, Figure 10, lane 5). Varying concentrations of TMPyP4 and DOX were independently added, allowed to bind with LTR GQ, and then primer extension reactions were carried out as before. Upon addition of TMPyP4 there was a noticeable decrease in the formation of the full-length product, accompanied by a simultaneous rise in stalled products (Figure 11, lanes 2–5, Figure 9D). At 0.75 μM of the ligand (7.5 equiv. w.r.t **T1**) no detectable full-length product was observed (Figure 11, lane 6). Similarly, increasing amounts of DOX resulted in a progressive reduction in the formation of the full-length product (Figure 11, lane 7–10, Figure 9D). Although DOX inhibited the polymerase activity, it required a higher amount (20 equiv.) to produce an effect comparable to TMPyP4 (Figure 11, compare lane 6 and 10). This may be due to DOX competing for the primer-template duplex, hairpin and GQ regions. These results prompted us to study the combined effect of ligands, wherein different concentrations of TMPyP4 and DOX at 1:1 ratio were added to the reaction mixture. The gel image revealed the synergistic effect of ligands as the formation of the full-length product considerably decreased with concomitant increase in truncated products in comparison to reactions in which only one ligand

was added at an equivalent concentration (Figure 9D, Figure S26, lanes 11–14). The effect is noticeable when we compare TMPyP4 (0.3 μM, ~16%), DOX (1 μM, ~17%), and TMPyP4+DOX (0.3 μM, ~9%, Figure 9D). These results suggest that GQ-hairpin motif of LTR G-rich region serves as regulatory point and it can be used as a conserved target to block the viral replication process by targeting simultaneously both GQ and hairpin domains. Based on our results and the formation of a unique GQ-hairpin architecture in cellular environment, we propose that bimodal ligand scaffolds composed of GQ and duplex binders, clamped using an appropriate linker, could selectively target the HIV-1 LTR and profoundly attenuate its pathogenesis (Figure 12). Needless to say, careful consideration should be exercised when optimizing the ligand design. Choice of GQ and hairpin binders from available examples and linker length, flexibility and point of attachment to the ligands will be very crucial.

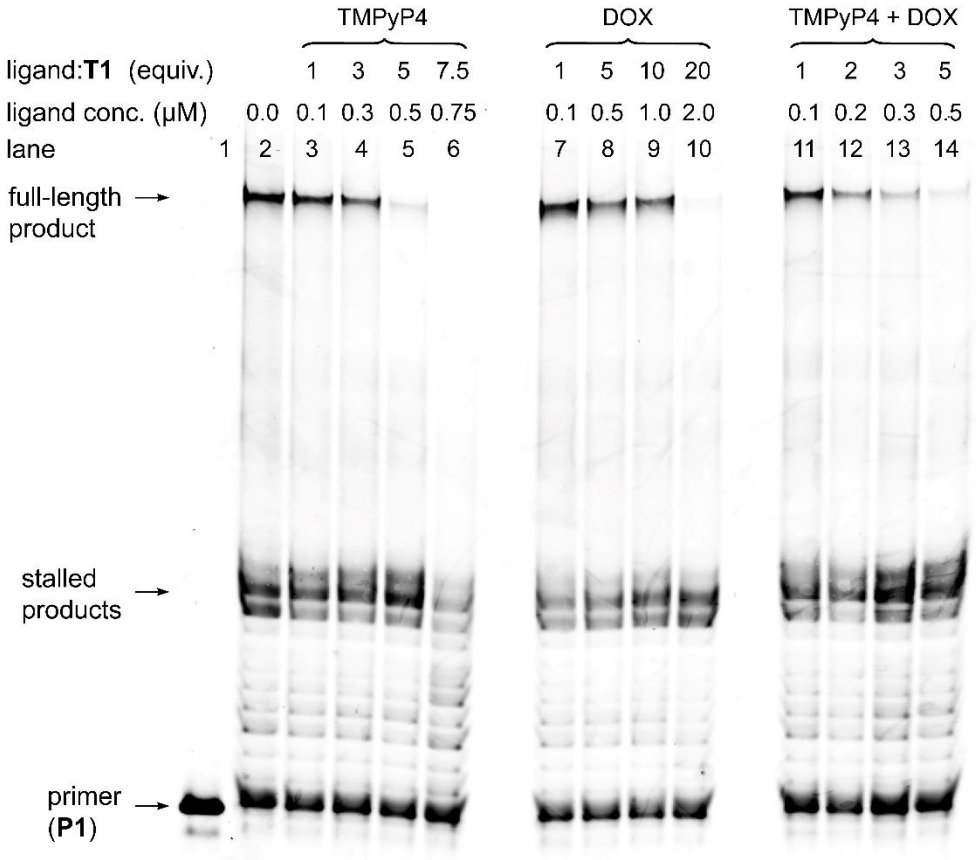


Figure 11. Gel image of primer extension reactions using native LTR G-rich ON template **T1** in the presence of different ligands namely, TMPyP4/DOX and TMPyP4+DOX. See experimental section for details.

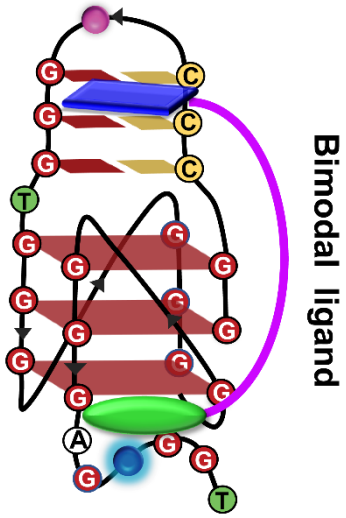


Figure 12. Tailor-made bimodal ligand scaffolds composed of GQ and duplex binders could selectively interact with respective domains of LTR G-rich motif. The design of clamping linker (length and interaction partners) will be crucial in adding to the selectivity and location of binding within the GQ, hairpin and GQ-hairpin junction domains. Here, a representative mode of binding of the ligand scaffold is shown.

3.3 Conclusions

We utilized the probe platform composed two highly environment-sensitive nucleoside analogs, FBFdU and FdU, to investigate the druggable space conserved G-rich region in the HIV-1 LTR. the nucleoside probes facilitated the detection and estimation of structure-specific ligand interactions by fluorescence and ^{19}F NMR techniques. Polymerase stop assay confirmed the regulatory function of LTR GQ structures. While TMPyP4 (GQ binder) and DOX (duplex binder) individually decreased polymerase activity, an equimolar mixture exhibited a synergistic inhibitory effect on LTR replication. Taken together, these findings suggest that simultaneous targeting of the juxtaposed GQ-hairpin motif using bimodal ligands could be a rational plan to selectively and efficiently inhibit the pathogenesis of the virus. In this direction, the FBFdU and FdU combination offers a versatile platform to study the structure as well as devise screening assays to identify hybrid ligands targeting GQ-hairpin/duplex motifs.

3.4 Experimental Section

3.4.1 Materials. BRACO19 hydrochloride and all reagents (Bio-Ultra grade) used in the preparation of buffers were purchased from Sigma-Aldrich. TMPyP4 and Doxorubicin

hydrochloride (DOX) were procured from Merck-Millipore. Millipore water after autoclaving was used for the preparation of all buffer solutions and in all biophysical studies.

3.4.2 Instruments. Fluorescence of the ONs samples were recorded using a Fluoromax-4 spectrophotometer (Horiba Scientific). 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.

3.4.3 Ligand binding using fluorescence. LTR GQ structures (0.5 μM) of ONs **4**, **9** and **16** were formed by heating the samples at 95 °C for 5 min in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl. The samples were incubated with increasing concentrations of the ligands (TMPyP4 and BRACO19) at 25 °C for 1 h. Samples were excited at 330 nm with an excitation and emission wavelength slit widths of 7 nm and 9 nm, respectively. For DOX titration, increasing concentrations of annealed GQ formed by ON **12** in the same ionic conditions as mentioned above was incubated with DOX (2 μM) at 25 °C for 1 h. The final volume of each sample solution was kept at 200 μL. Samples were excited at 480 nm, with an excitation and emission slit widths of 7 nm and 9 nm respectively. Fluorescence experiments were performed in triplicate in a micro-fluorescence cuvette at 25 °C. For titration with TMPyP4 and BRACO19, an appropriate blank containing ligand in a buffer was subtracted from each reading for the corresponding ligand concentration. Normalized fluorescence intensity (F_N) against ligand concentration was plotted and fitted to a Hill equation (see below) to determine the apparent K_d values. Fitted graphs were prepared using OriginPro 8.5 software.

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

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

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

3.4.4 Ligand binding using ^{19}F NMR. LTR GQ structures of ONs **4** (45 μM), **9** (75 μM), **16** (45 μM) were formed by heating the samples at 95 °C for 5 min in 20 mM potassium phosphate buffer (pH 7) containing 70 mM KCl in 20% D_2O. The samples were allowed to cool at RT and then they were transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F NMR spectra were recorded at a frequency 564.9 MHz 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). After each experiment, increasing concentrations of ligands were added and incubated at RT for 1 h prior to the experiment. 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 8. ^{19}F NMR spectra were obtained in 2–2.5 h with 5000–6000 scans. Spectra were processed with an exponential window function using lb = 20 Hz.

3.4.5 *Taq* polymerase assay: 5′-FAM labeled primer **P1** (5 μM) and template DNA ONs **T1** and **T2** (5 μM) were annealed in 10 mM Tris-HCl (pH 7.8) containing 100 mM KCl at 95 °C for 5 minutes and slowly cooled to RT (Table S5). The primer-template duplexes were further diluted to 1 μM in 10 mM Tris-HCl buffer containing 100 mM KCl. Primer extension reactions were performed with primer-template duplex (100 nM), KCl (100 mM), 1× DNA polymerase buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) at 37 °C. Reactions were initiated by adding dNTPs (500 μM) and 0.5 μL of *Taq* DNA polymerase (5 U/μL, New England Biolabs, Catlog. M0273S) in a total reaction volume of 20 μL. Reactions were stopped at different time intervals by adding 10 μL of denaturing loading buffer (8.3 M urea in 10 mM Tris-HCl, 100 mM EDTA, 0.05% bromophenol blue, pH 8) further flash cooled on a dry-ice bath. The reaction mixture was then concentrated in a speedvac concentrator. The extension products were resolved by 15% denaturing PAGE containing 8.3 M urea and was electrophoresed at a constant power of 35 W for 2.5–3 h. The gel was scanned using an Amersham Typhon 600 (GE Healthcare) at the FAM wavelength and quantified with the help of the ImageJ software. Impact of ligands on the polymerase activity was studied by adding different concentrations of TMPyP4 and or DOX. The ligands were first added to the reaction mixture and incubated for 1 h at RT and then initiated by adding dNTPs and enzyme as above. The reaction products were analyzed as described above.

3.5 References

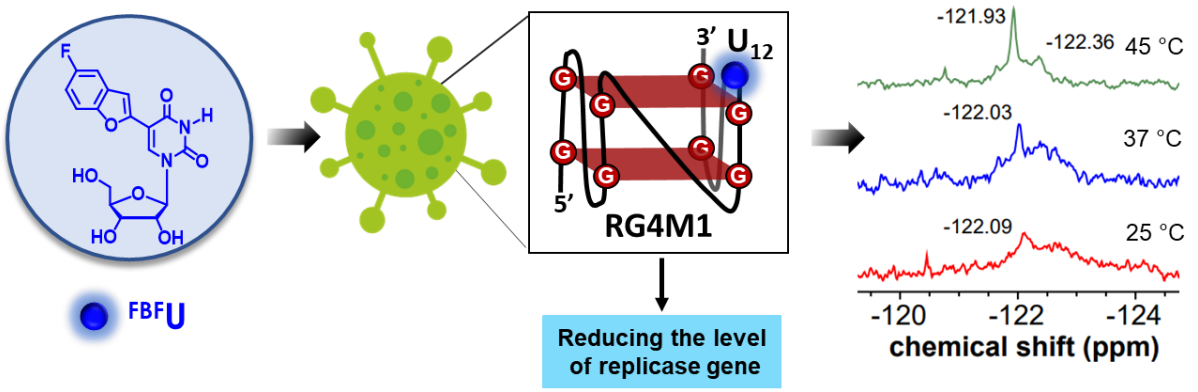
1. Kosiol, N.; Juranek, S.; Brossart, P.; Annkristin, H.; Katrin P. G-quadruplexes: a promising target for cancer therapy. *Mol. Cancer* **2021**, *20*, 40.
2. Mendoza, O.; Bourdoncle, A.; Boulé, J.B.; Brosh, R. M.; Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res*. **2016**, *44*, 1989–2006.
3. Murat, P.; Singh, Y.; Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev.*, **2011**, *40*, 5293–5307.
4. Current Topics in Medicinal Chemistry, 2008, Vol. 8, No. 15
5. Yan, Y.; Tan, J.; Ou, T.; Huang, Z.; Gu, L. DNA G-quadruplex binders: a patent review. *Expert Opin. Ther. Pat.* **2013**, *23*, 1495–1509.
6. Sanchez-Martin, V.; Soriano, M.; Garcia-Salcedo, J. A. Quadruplex Ligands in Cancer Therapy. *Cancers (Basel)*. **2021**, *13*, 3156.
7. Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
8. Yang, D.; Okamoto, K. Structural insights into G-quadruplexes: Towards new anticancer drugs. *Future Med. Chem.* **2010,** *2*, 619–646.
9. Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaelbe, M.; Lin, A.; Bliesath, J.; Ho, C.B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K.; Rice, W. G. Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. Cancer Res. **2009**, *69*, 7653–7661.
10. Ruggiero, E.; Richter, S. N. Targeting G-quadruplexes to achieve antiviral activity. *Bioorg. Med. Chem. Lett.*, **2023**, *79*, 129085.
11. Ruggiero, E.; Zanin, I.; Terreri, M.; Richter, S. N. G-Quadruplex Targeting in the Fight against Viruses: An Update. *Int. J. Mol. Sci.*, **2021**, *22*, 10984.
12. Abiri, A.; Lavigne, M.; Rezaei, M. et al. Unlocking G-quadruplexes as antiviral targets. *Pharmacol Rev*. **2021**, *73*, 897–923.
13. Pathak, R. G-Quadruplexes in the Viral Genome: Unlocking Targets for Therapeutic Interventions and Antiviral Strategies. *Viruses* **2023**, *15*, 2216.
14. Luo, X.; Xue, B.; Feng, G. *et al.* Lighting up the native viral RNA genome with a fluorogenic probe for the live-cell visualization of virus infection. *J. Am. Chem. Soc.*, **2019**, *141*, 5182–5191.
15. Perrone, R.; Doria, F.; Butovskaya, E. *et al.* Synthesis, binding and antiviral properties of potent core-extended naphthalene diimides targeting the HIV-1 long terminal repeat promoter G-quadruplexes. *J. Med. Chem.*, **2015**, *58*, 9639–9652.
16. Sarkar, S.; Armitage, B. A. Targeting a potential G-quadruplex forming sequence found in the west nile virus genome by complementary gamma-peptide nucleic acid oligomers. *ACS Infect. Dis.*, **2021**, *7*, 1445–1456.
17. Shen, L.W.; Qian, M. Q.; Yu, K. *et al.* Inhibition of Influenza A virus propagation by benzoselenoxanthenes stabilizing TMPRSS2 Gene G-quadruplex and hence down-regulating TMPRSS2 expression. *Sci. Rep.*, **2020**, *10*, 7635.
18. Carvalho, J.; Queiroz, J.A.; Cruz, C. Circular dichroism of G-Quadruplex: A laboratory experiment for the study of topology and ligand binding. *J. Chem. Educ.* **2017**, *94*, 1547–1551.
19. Krafcikova, M.; Dzatko, S.; Caron, C.; Granzhan, A.; Fiala, R.; Loja, T.; Teulade-Fichou, M.P.; Fessl, T.; Hänsel-Hertsch, R.; Mergny, J.L.; et al. Monitoring DNA-Ligand

Interactions in Living Human Cells Using NMR Spectroscopy. *J. Am. Chem. Soc.*, **2019**, *141*, 13281–13285.
20. Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J. L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **2015**, *6*, 3314–3320.
21. Parkinson, G. N.; Collie, G. W. X-ray crystallographic studies of G-quadruplex structures. *In Methods in Molecular Biology; Springer: Berlin, Germany*, **2019**, *2035*, 131–155.
22. Prado, E.; Bonnat, L.; Bonnet, H.; Lavergne, T.; Van Der Heyden, A.; Pratviel, G.; Dejeu, J.; Defrancq, E. Influence of the SPR Experimental Conditions on the G-Quadruplex DNA Recognition by Porphyrin Derivatives. *Langmuir* **2018**, *34*, 13057–13064.
23. Pagano, B.; Mattia, C.A.; Giancola, C. Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. *Int. J. Mol. Sci.* **2009**, *10*, 2935–2957.
24. Lecours, M. J.; Marchand, A.; Anwar, A.; Guetta, C.; Hopkins, W.S.; Gabelica, V. What stoichiometries determined by mass spectrometry reveal about the ligand binding mode to G-quadruplex nucleic acids. *Biochim. Biophys. Acta Gen. Subj.* **2017**, *1861*, 1353–1361.
25. De Cian, A.; Guittat, L.; Shin-ya, K.; Riou, J. F.; Mergny, J. L. Affinity and selectivity of G4 ligands measured by FRET. *Nucleic Acids Symp. Ser.* **2005**, *49*, 235–236.
26. Monchaud, D.; Teulade-Fichou, M. P. G4-FID: A fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands. *Methods Mol. Biol.* **2010**, *608*, 257–271.
27. Chang, T.; Liu, X.; Cheng, X.; Qi, C.; Mei, H.; Shangguan, D. Selective isolation of G-quadruplexes by affinity chromatography. *J. Chromatogr.* **2012**, *1246*, 62–68.
28. Wu, G.; Tillo, D.; Ray, S.; Chang, T. C.; Schneekloth, J. S.; Vinson, C.; Yang, D. Custom G4 microarrays reveal selective G-quadruplex recognition of small molecule BMVC: A large-scale assessment of ligand binding selectivity. *Molecules* **2020**, *25*, 3465.
29. Manna, S.; Sarkar, D; Srivatsan, S. G. A dual-app nucleoside probe provides structural insights into the human telomeric overhang in live cells. *J. Am. Chem. Soc.*, **2018,** *140*, 12622–12633.
30. Le, D. D.; Antonio, M. D.; Chan, L. K. M; Balasubramanian, S. G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun.*, **2015**, *51*, 8048–8050.
31. Lim, K. W.; Jenjaroenpun, P.; Low, Z. J.; Khong, Z. J.; Ng, Y. S.; Kuznetsov, V. A.; Phan, A. T. Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.*, **2015**, *43*, 5630–5646.
32. Vianney, Y. M.; Weisz, K. High-affinity binding at quadruplex–duplex junctions: rather the rule than the exception. *Nucleic Acids Res.*, **2022**, *50*, 11948–11964.
33. Greco, M. L.; Kotar, A.; Rigo, R.; Cristofari, C.; Plavec, J.; Sissi, C. Coexistence of two main folded G-quadruplexes within a single G-rich domain in the EGFR promoter. *Nucleic Acids Res.*, **2017** *45*, 10132–10142.
34. Yang, M.; Carter, S.; Parmar, S.; Bume, D. D.; Calabrese, D. R.; Liang, X.; Yazdani, K.; Xu, M.; Liu, Z.; Thiele, C. J.; Schneekloth, J. S. Targeting a noncanonical, hairpin-containing G-quadruplex structure from the MYCN gene. *Nucleic Acids Res.*, **2021**, *49*, 7856–7869.
35. 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**, *14*, 5627–5637.
36. Sannapureddi, R. K. R.; Mohanty, M. K.; Salmon, L.; Sathyamoorthy, B. Conformational plasticity of parallel G-quadruplex–implications on duplex-quadruplex motifs. *J. Am. Chem. Soc.*, **2023** *145*, 15370–15380.

37. Díaz-Casado, L.; Serrano-Chacón, I.; Montalvillo-Jiménez, L.; Corzana, F.; Bastida, A.; Santana, A. G.; González, C.; Asensio, J. L. De novo design of selective quadruplex–duplex junction ligands and structural characterisation of their binding mode: targeting the G4 hot-spot. *Chem. -Eur. J.*, **2021**, *27*, 6204–6212.
38. Vianney, Y. M.; Preckwinkel, P.; Mohr, S.; Weisz, K. Quadruplex-duplex junction: a high-affinity binding site for indoloquinoline ligands. *Chemistry*, **2020**, *26*, 16910–16922.
39. Nguyen, T. Q. N.; Lim, K. W.; Phan, A.T. A dual-specific targeting approach based on the simultaneous recognition of duplex and quadruplex motifs. *Sci. Rep.*, **2017**, *7*, 11969.
40. Asamitsu, S.; Obata, S.; Phan, A. T.; Hashiya, K.; Bando, T.; Sugiyama, H. Simultaneous binding of hybrid molecules constructed with dual DNA-binding components to a G-quadruplex and its proximal duplex. *Chemistry*, **2018**, *24*, 4428–4435.
41. Mandal, S.; Kawamoto, Y.; Yue, Z.; Hashiya, K.; Cui, Y.; Bando, T.; Pandey, S.; Hoque, M. E.; Hossain, M. A.; Sugiyama, H.; Mao, H. Submolecular dissection reveals strong and specific binding of polyamide-pyridostatin conjugates to human telomere interface. *Nucleic Acids Res.*, **2019**, *47*, 3295–3305.
42. Manet, I.; Manoli, F.; Zambelli, B.; Andreano, G.; Masi, A.; Cellai, L.; Monti, S. Affinity of the anthracycline antitumor drugs doxorubicin and sabarubicin for human telomeric G-quadruplex structures. *Phys. Chem. Chem. Phys.*, **2010**, *13*, 540–551.
43. Bilgen, E.; Persil Çetinkol, Ö. Doxorubicin exhibits strong and selective association with VEGF Pu22 G-quadruplex. *Biochim. Biophys. Acta Gen. Subj.*, **2020**, *1864*, 129720.
44. Scaglioni, L.; Mondelli, R.; Artali, R.; Sirtori, F. R.; Mazzini, S. Nemorubicin and doxorubicin bind the G-quadruplex sequences of the human telomeres and of the c-MYC promoter element Pu22. *Biochim. Biophys. Acta*, **2016**, *1860*, 1129–1138.
45. Scalabrin, M.; Frasson, I.; Ruggiero, E.; Perrone, R.; Tosoni, E.; Lago, S.; Tassinari, M.; Palù, G.; Richter, S. N. The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes. *Sci. Rep.* **2017**, *7*, 45244.

Chapter 4

Probing SARS-CoV-2 Nsp3 G-quadruplexes by fluorescence and ^{19}F NMR using a functionalized nucleoside analog



4.1 Introduction

SARS-CoV-2, a betacoronavirus belonging to the Coronaviridae family, is the main etiological agent causing the COVID-19 pandemic.[1,2] It is a positive-sense single-stranded RNA (+)ssRNA virus that functions as a mRNA encoding viral proteins necessary for replication, and structure composition.[3,4] The emergence of different SARS-CoV-2 variants has created significant challenges in controlling the COVID-19 pandemic.[5] As a result, there remains a critical need for developing effective antiviral drugs. So far, the traditional strategy for developing small-molecule antiviral drugs has focused on viral proteins like polymerases, proteases, structural and non-structural proteins.[6–10] However, a significant challenge in targeting these viral proteins is the inevitable emergence of drug-resistant strains, driven by the high mutation rates of viruses.[11,12] Owing to these challenges, targeting structural motifs of highly conserved regions of the viral genome may offer a novel approach to expand the range of available antiviral treatments and block viral proliferation.

Conserved regions in the viral genome can adopt various secondary structures including G-quadruplexes (GQs).[13] Despite the diversity of viral genomes, these GQs exhibit remarkable conservation in their sequences, location, and occurrence.[14–17] The viral GQs play critical roles in regulating several stages of the viral life cycle including replication, genome stability, evolution, transcription, and translation.[18,19] Consequently, they present promising targets for the development of antiviral therapies.

The SARS-CoV-2 genomic RNA is ~30 kb long encoding about 25 non-structural proteins (Nsps) and accessory proteins, four structural proteins including spike (S), envelope (E), membrane (M), and nucleocapsid (N) (Figure 1A).[20,21] Bioinformatics and experimental studies reveal the widespread presence of putative GQ-forming sequences in various regions of the SARS-CoV-2 genome associated with structural and functional viral proteins. These regions majorly include Nsp3,[22,23] Nsp10,[24] S,[25], and N[26] genes. Importantly, the Nsp3 fragment within the open reading frame (ORF1ab) region is critical for producing the replicase gene[22] and assembling the viral replication and transcription complex (RTC).[27] Given its influence on translation-related processes such as elongation, ribosomal frameshifting, and mRNA surveillance (no-go mRNA decay), the Nsp3 GQ in the ORFs of the replicase gene is hypothesized to inhibit its translation. This, in turn, may reduce polyprotein expression.[28] We therefore envision that probing the structural features of the Nsp3 G-rich sequence and its recognition properties could provide valuable insights to support the development of a feasible therapeutic strategy.

The formation, stability, and interactions of GQ structures with small molecules are typically investigated *in vitro* using techniques such as circular dichroism (CD), UV thermal melting, fluorescence, NMR, and X-ray crystallography.[29–32] Notably, antibodies and probes for in-cell detection of RNA GQs have been recently developed.[33–36] Despite that, when multiple GQ species coexist, most techniques and probes, aside from a few exceptions,[37–41] struggle to distinguish distinct topologies or structures. This limitation makes it difficult to obtain precise information about the individual structures and their equilibrium states. Further, most of the tools pose challenges in assessing the structures in the cellular environment.

In this context, to study nucleic acid structures and interactions in greater detail, we designed microenvironment-sensitive nucleoside probes with dual functionalities.[42,43] These probes are designed by attaching a heterocyclic ring to the C5 position of the uracil, extending the π conjugation while maintaining the integrity of the native nucleic acid structures.[44] Following this approach, we designed selenophene, benzofuran, and benzothiophene conjugated probes. These carefully designed probes are ideal for studying complex structures like GQs and i-motifs, as well as nucleic acid–drug interactions.[45] Additionally, we broadened the probe design by incorporating a ^{19}F atom into the benzofuran-modified nucleoside, thereby combining the advantages of fluorescence and ^{19}F NMR spectroscopy. The ^{19}F isotope was chosen for its natural abundance, high sensitivity, significant chemical shift dispersion, and absence in biological systems, which greatly aids in the detection of structures in cell-based systems.[46,47] Specifically, 5-fluorobenzofuran-modified 2′-deoxyuridine (FBFdU) functions as an effective dual-purpose probe for interrogating human telomeric DNA repeat *in vitro* and in cellular environment using fluorescence and ^{19}F NMR.[42] This dual-app probe when incorporated into ONs marginally affects the native folding and nicely reports the changes in the microenvironment.[46,47] Notably, this probe was also utilized to elucidate the proviral DNA GQ structures of the HIV-1 LTR region, in a cell-free environment and in an *ex vivo* model using oocytes egg extract.[48] As an expansion to our probe's utility, the probe design was extended to the RNA analogue and employed as a nucleic acid structure sensor. Recently, 5-fluorobenzofuran-modified-uridine (FBFU **1**) analog was used to detect the metal ion-induced conformational change of internal ribosome entry site (IRES) RNA motif of hepatitis C virus HCV.[49]

Building on the effectiveness of these nucleoside probes, we extended its applicability to investigate viral RNA GQs, particularly in the context of the Nsp3 GQ. Recent studies using biophysical and biochemical tools have identified the formation of this GQ and demonstrated its stabilization by small molecules like TMPyP4, which in turn leads to reduced viral

replication in animal models.[22] While these findings highlight the therapeutic potential of targeting the Nsp3 GQ, a systematic understanding of its structural features and polymorphism could be highly beneficial for advancing the drug development framework. In this regard, we leveraged the unique properties of FBFU. Its dual functionality combines fluorescence sensitivity with the specificity of ^{19}F NMR, providing a simple and robust platform for probing the structural features and ligand interactions of the Nsp3 GQ. The FBFU probe detected the formation and structural polymorphism of Nsp3 GQ with distinct and resolved spectral signatures. Importantly, ^{19}F NMR enabled the identification of co-existing higher-order GQ structures, while the fluorescent component facilitated the estimation of binding affinities between Nsp3 GQ and small molecule stabilizers. This approach builds on existing studies while offering valuable insights that could guide the development of targeted strategies.

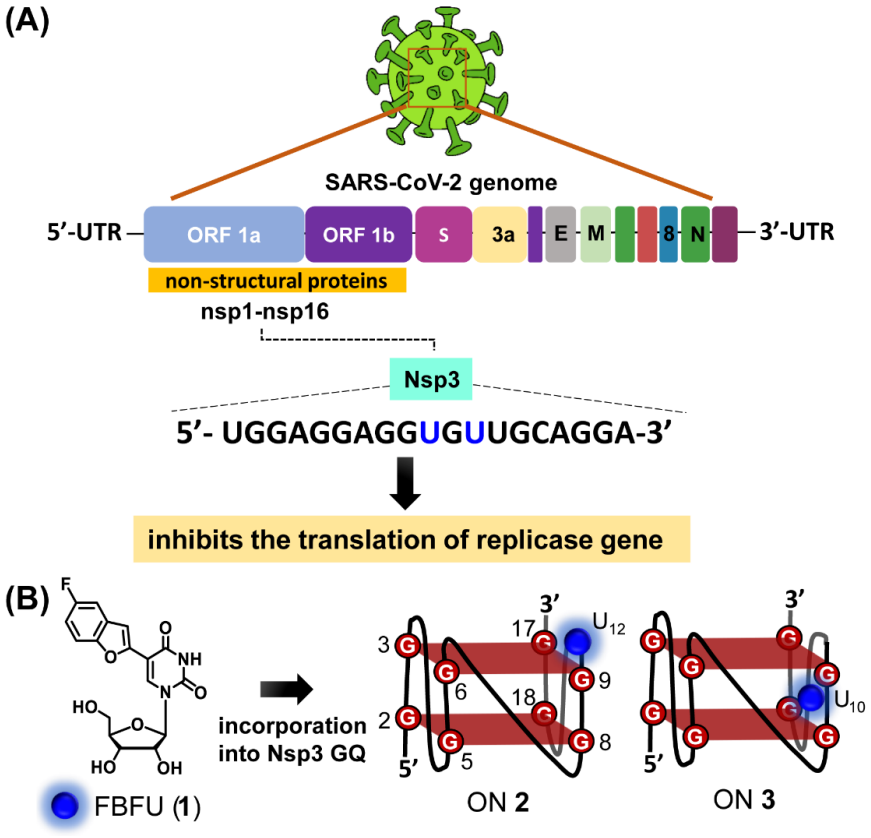


Figure 1. (**A**) SARS-CoV-2 genome representing structural proteins, non-structural proteins, and accessory proteins. (**B**) Chemical structure of microenvironment sensitive probe, FBFU (**1**) and its incorporation into SARS-CoV-2 Nsp3 GQ.

4.2 Results and Discussion

4.2.1 Synthesis of FBFU-modified SARS-CoV-2 Nsp3 ONs

The G-rich segment of Nsp3 predominantly forms a two-tetrad intramolecular parallel GQ structure.[22,23] To probe the structural features, the next key factor we considered was the positioning of FBFU (**1**) within SARS-CoV-2 Nsp3 GQ-forming ONs. Compared to G-quartets, the loop residues vary significantly in the orientation type, composition, and interactions with adjacent bases among different GQs.[50,51] Hence, we anticipate that placing the FBFU (**1**) probe in the loop region replacing the natural uridine would have minimum impact on the folding process and would also provide the structural information (Figure 1). The nucleoside FBFU was placed in the U_{12} and U_{10} positions in the third lateral loop of the SARS-CoV-2 Nsp3 G-rich ON sequences (19-mer) **2** and **3**, respectively, *via* solid-phase ON synthesis using the phosphoramidite (Figure 2, Table 1). Modified ONs were purified by using denaturing PAGE. Isolated ONs were subjected to RP-HPLC to determine their purity and ESI-MS to ascertain their integrity (Figure 3 and 4) (Table 2).

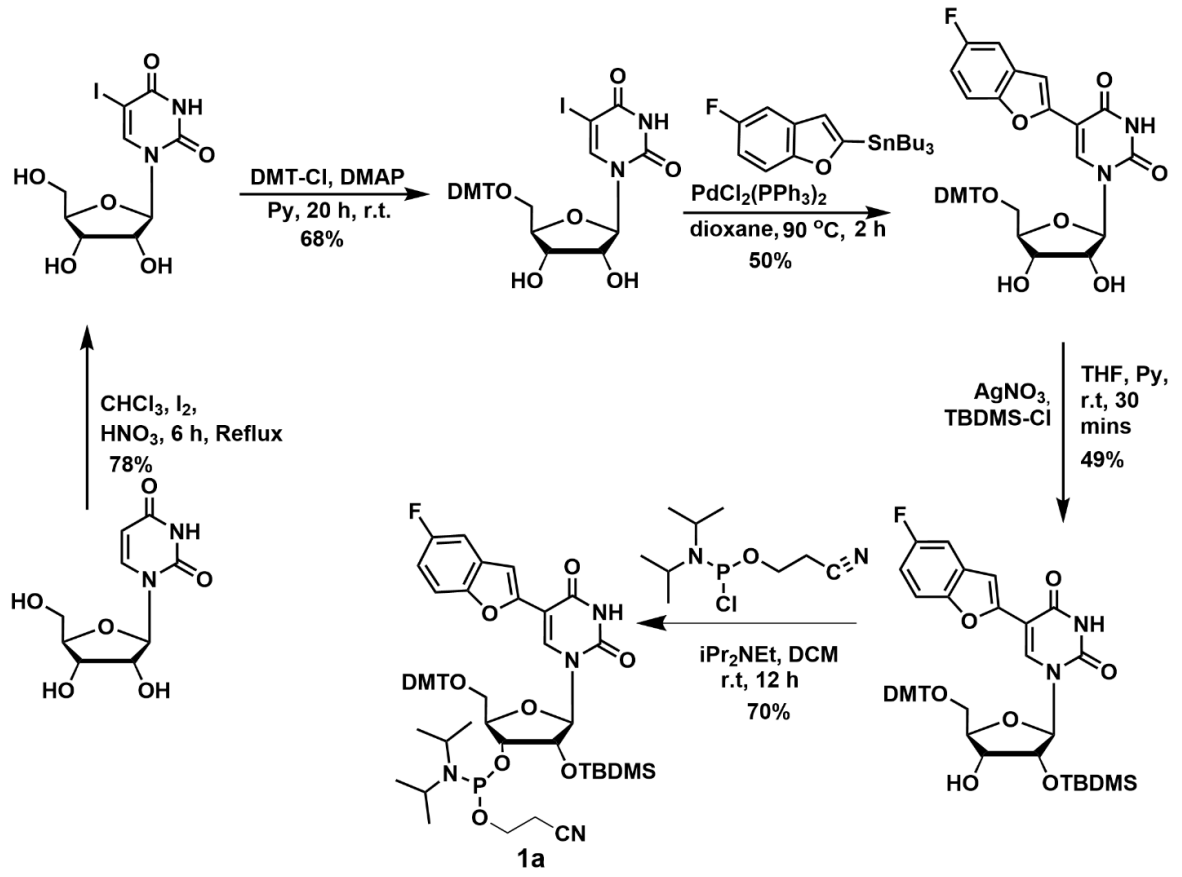


Figure 2. Synthesis of modified phosphoramidite **1a.**[49]

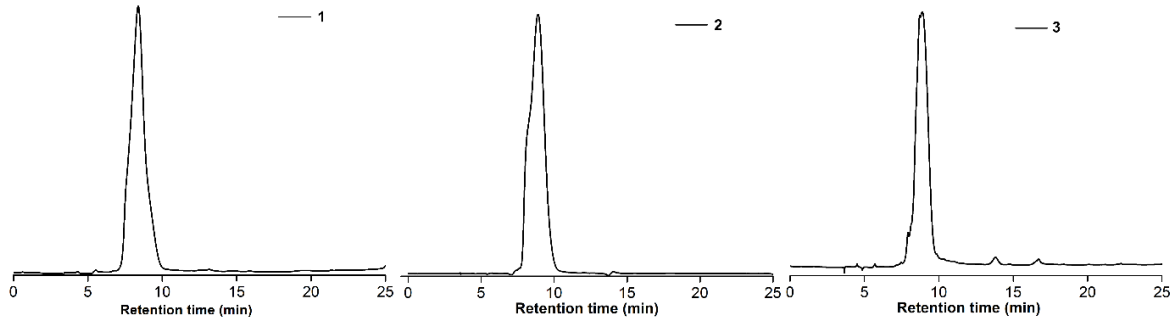
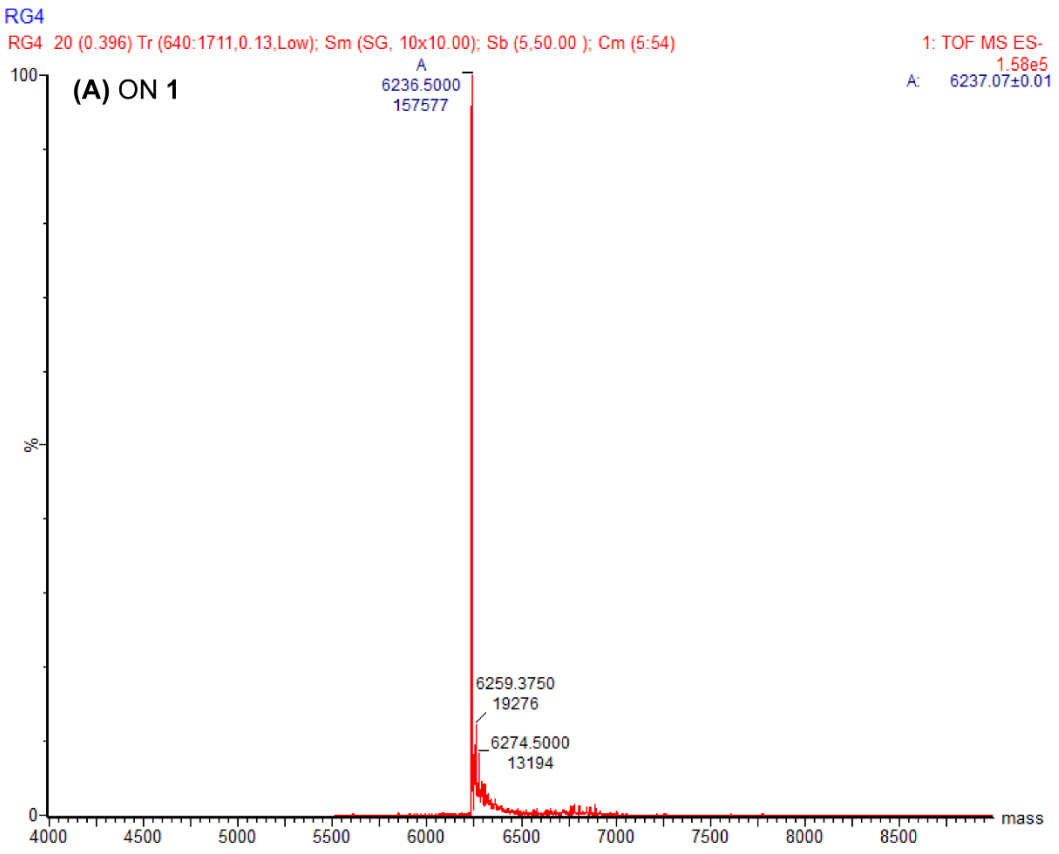


Figure 3. The purity of PAGE purified ONs containing the modification was analyzed by RP-HPLC at 260 nm using Luna C18 column (250 x 4.6 mm, 5 micron). Mobile phase A= 100 mM triethylammonium acetate (pH=7.3) and B=acetonitrile. Gradient: 0–100 % B in 30 min with a flow rate of 1 mL/min was used for ONs **1**–**3**. The chromatogram peaks appear to be broad because of the existence of higher-order structures and could not be separated (*vide infra*).

Table 1: Native and FBFU-modified SARS-CoV-2 Nsp3 GQ. Modification position is highlighted in bold and underlined.

ON	5′----------3′
1	UGGAGGAGGUGUUGCAGGA
2	UGGAGGAGGUG**U**UGCAGGA
3	UGGAGGAGG**U**GUUGCAGGA



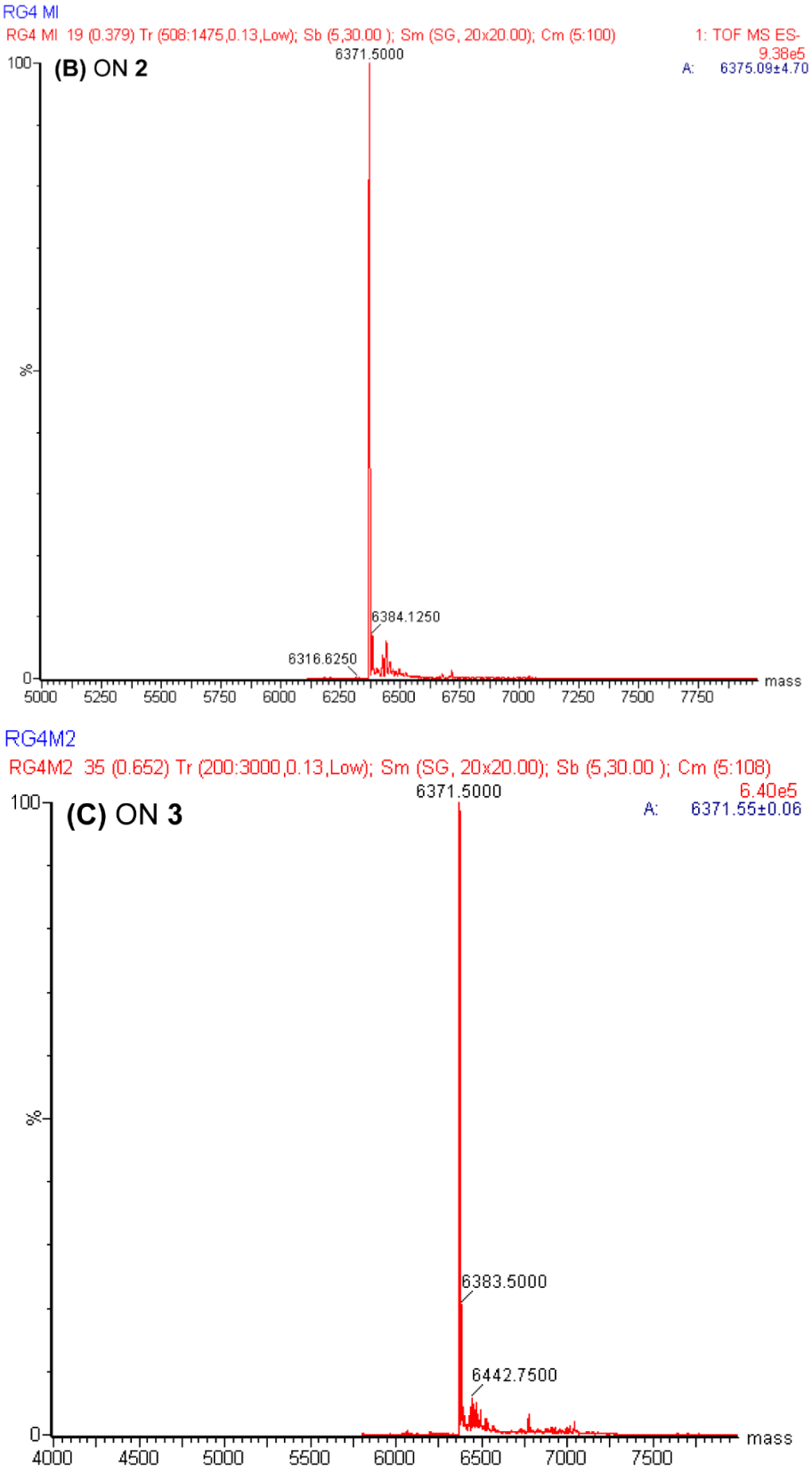


Figure 4. ESI-MS of purified SARS-CoV-2 Nsp3 ONs. See Table 2 for details.

Table 2. ε_{260} and mass data of control and modified SARS-CoV-2 Nsp3 ONs

ON sequence	ε_{260} [M^{-1} cm^{-1}][a]	calculated (g/mol)	found (g/mol)
1	199300	6236.75	6236.50
2	199860	6370.50	6371.25
3	199860	6370.50	6371.50

[a]Molar absorption coefficient (ε) of modified ONs was determined by using OligoAnalyzer 3.1. The molar absorptivity of modified nucleoside FBFU (ε_{260}= 10360 M^{-1} cm^{-1}) was used in place of U.

4.2.2 Minimally perturbing nature of the probe FBFU (1)

The formation of Nsp3 GQ was first evaluated using CD analysis. Both control unmodified ON **1** and modified ONs **2** and **3** display a positive peak at ~260 nm and a negative dip at ~240 nm, depicting the parallel GQ topology. Notably, a hump at ~310 nm can also be observed, which suggests the possible prevalence of higher-order structure.[21] Importantly, the spectra also infer that the modification in the loop region minimally perturbs the native structure (Figure 5A). Also, thermal melting analysis revealed that the forward (heating) and backward (cooling) cycles of the GQ formed by ONs **1**−**3** display a hysteresis of ~10 °C. It suggests relatively slow folding and unfolding of GQ structures, following a different $T_{1/2}$ transition, signifying multiple structures.[52–54] In addition, the T_m for both control ON **1** and modified ONs **2** and **3** are found to be similar, emphasizing the minimal perturbing nature of the modification to the native GQ structure (Figure 5B) (Table 3). To further substantiate the folding of the GQs, thermal differential spectra (TDS) were obtained. The absorbance of the GQs formed by ONs **1**−**3** was initially measured at a lower temperature (20 °C) and then at a temperature above its T_m (70 °C). A plot of molar extinction difference *vs* wavelength demonstrates hypochromicity at 295 nm, featuring the formation of the parallel GQ for all the ONs **1**−**3**. (Figure 6).

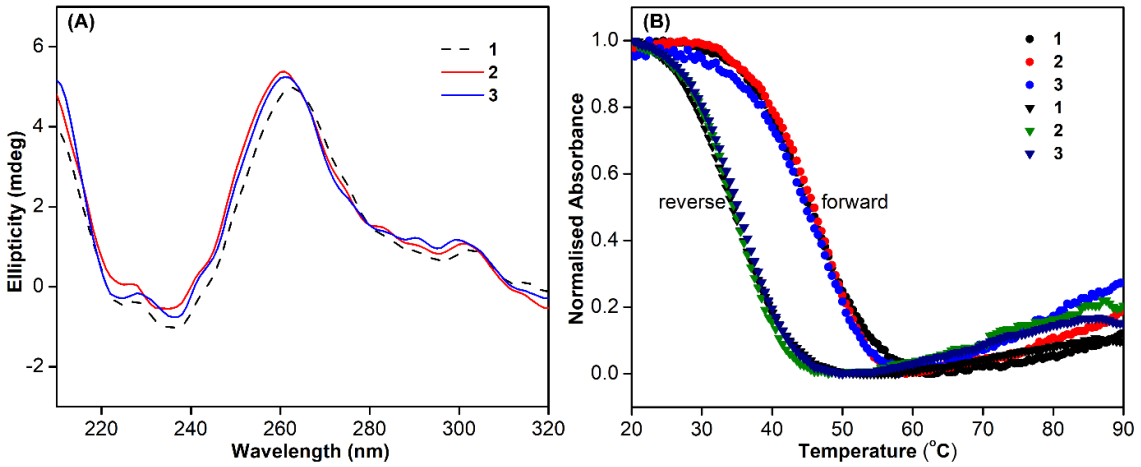


Figure 5. (**A**) CD spectra of control ON **1** (10 µM) and modified ONs **2** and **3** (10 µM) in 100 mM Tris-HCl containing 100 mM KCl (**B**) UV-thermal melting profiles for the same at 295 nm (5 µM) in 100 mM Tris-HCl containing 100 mM KCl. See Table 3 for T_m values.

Table 3. T_m values of modified and control unmodified ONs

ON sequence	**Forward** °C	**Reverse** °C
1	44.5 ± 1.5	35.0 ± 1.3
2	46.5 ± 0.5	35.0 ± 1.8
3	46.6 ± 0.7	34.1 ± 2.0

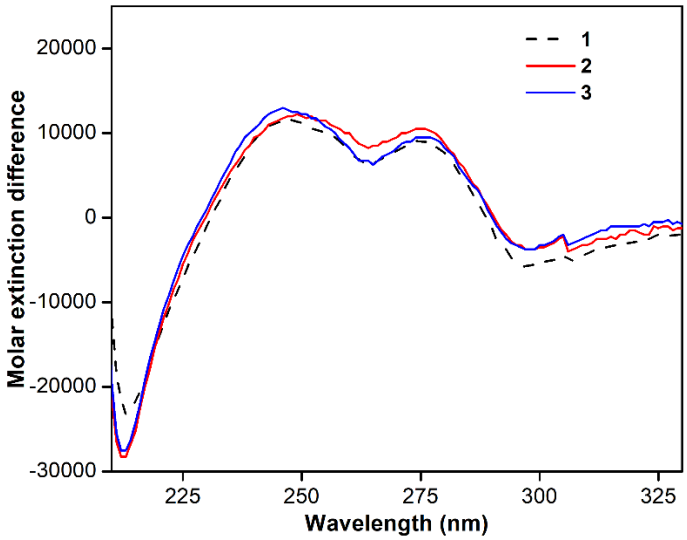


Figure 6. (**A**) TDS spectra of ONs **1**−**3** (5 µM) in 10 mM Tris-HCl (pH 7.4) containing 100 mM of KCl (**B**) CD spectra of ON **1** (5 µM) in different concentrations of KCl (0−100) mM.

Although these biophysical tools detect signals corresponding to GQ formation, several challenges complicate the accurate identification and study of these structures. First, the

reliance on CD spectroscopy to identify GQs is problematic, as a strong positive peak at 260 nm, often used as proof of GQ formation, is also characteristic of A-form RNA duplexes.[55] This overlap can lead to the misinterpretation of RNA hairpins as intramolecular GQs. Second, while most GQs fold stably under ~100 mM K^+ conditions, their stabilization and folding are highly dependent on the type and concentration of ions present. Studies have shown that some predicted GQs, like the GQ formed by the N gene of SARS-CoV-2, show minimal changes in CD spectra ellipticity despite decreasing K^+ concentrations.[56] This suggests that all GQs do not respond uniformly to ionic changes, highlighting the need to examine the behaviour of individual GQs to understand their unique structural properties and stability under different ionic environments. To further investigate the role of ionic conditions, both control unmodified ON **1** and modified ONs **2** and **3** were annealed in the absence and presence of different concentrations of K^+ ions. At 0 mM KCl the CD spectra of ONs did not show prominent bands corresponding to a GQ form, suggesting the formation of a non-GQ structure (black lines, Figure 7A–C). As the concentration of the K^+ is enhanced from (25−100) mM, there is an increment in the intensity of the ellipticity at 260 nm (a small dip at ~240 nm is also seen), suggesting the progressive stabilization of the GQ structure (Figure 7A–C).

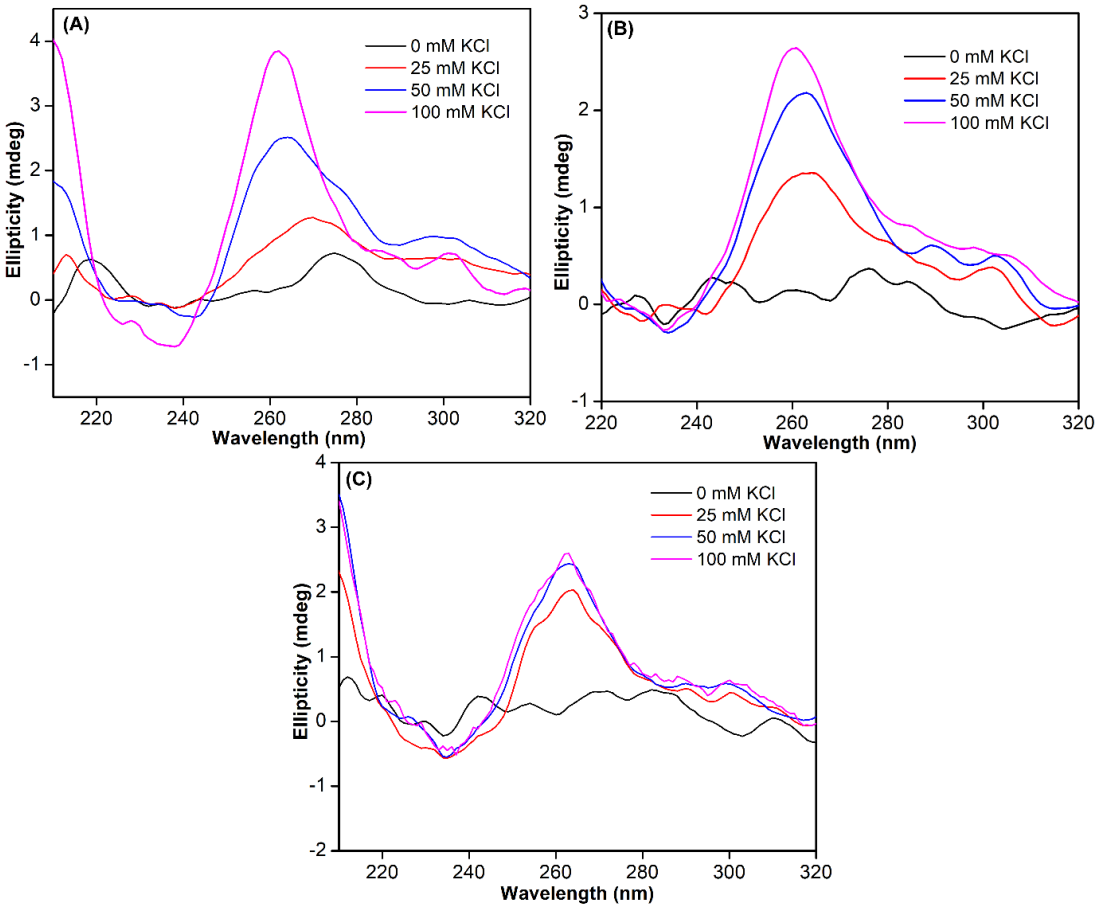


Figure 7. CD spectra in different concentrations of KCl (0−100) mM. (**A**) ON **1** (**B**) ON **2** (5 µM) **(C)** ON **3** (5 µM)

4.2.3 Probing SARS-CoV-2 Nsp3 GQs using the responsiveness of FBFU

Fluorescence intensity of FBFU changes with the polarity of the environment and generally get enhanced and red shifted in the polar environment but it is also reduced by the stacking interaction of the nearby base or the induced electron transfer by the adjacent guanine bases.[42] These environment changes can be possibly associated with the folding process of the SARS-CoV-2 GQs, and hence, we utilized the property of the probe to detect the formation of the GQs in different ionic conditions. FBFU-modified Nsp3 ON **2** and **3** were annealed in Tris-HCl buffer (pH 7.4) containing varying concentrations of K^+ ions, excited at 330 nm, and steady-state emission spectra were recorded. In the absence of KCl, the possible non-GQ structure exhibited an intense fluorescence band with an emission around 430 nm. Increasing the K^+ ion concentration from 25−100 mM resulted in a progressive decrease in the fluorescence intensity, accompanied by a slight blue shift in the emission maximum, as a result of the formation of a GQ structure (Figure 8A and 8B). In a control experiment, the fluorescence of the free nucleoside analog FBFU **1** remained largely unaffected by the variations in K^+ ion concentration (Figure 8C). Corroborating CD profiles, these results indicate that the observed quenching in fluorescence intensity as a function of increasing K^+ ion concentration is due to the formation of a stable GQ structure. Based on the photophysical properties of the ribonucleoside analog within ONs, we speculate that the probe incorporated at the U_{12} and U_{10} position (third loop) in ON **2** and **3** comes closer to the top G-tetrad upon GQ formation, which manifests in guanine-induced fluorescence quenching (Figure 8A and 8B).

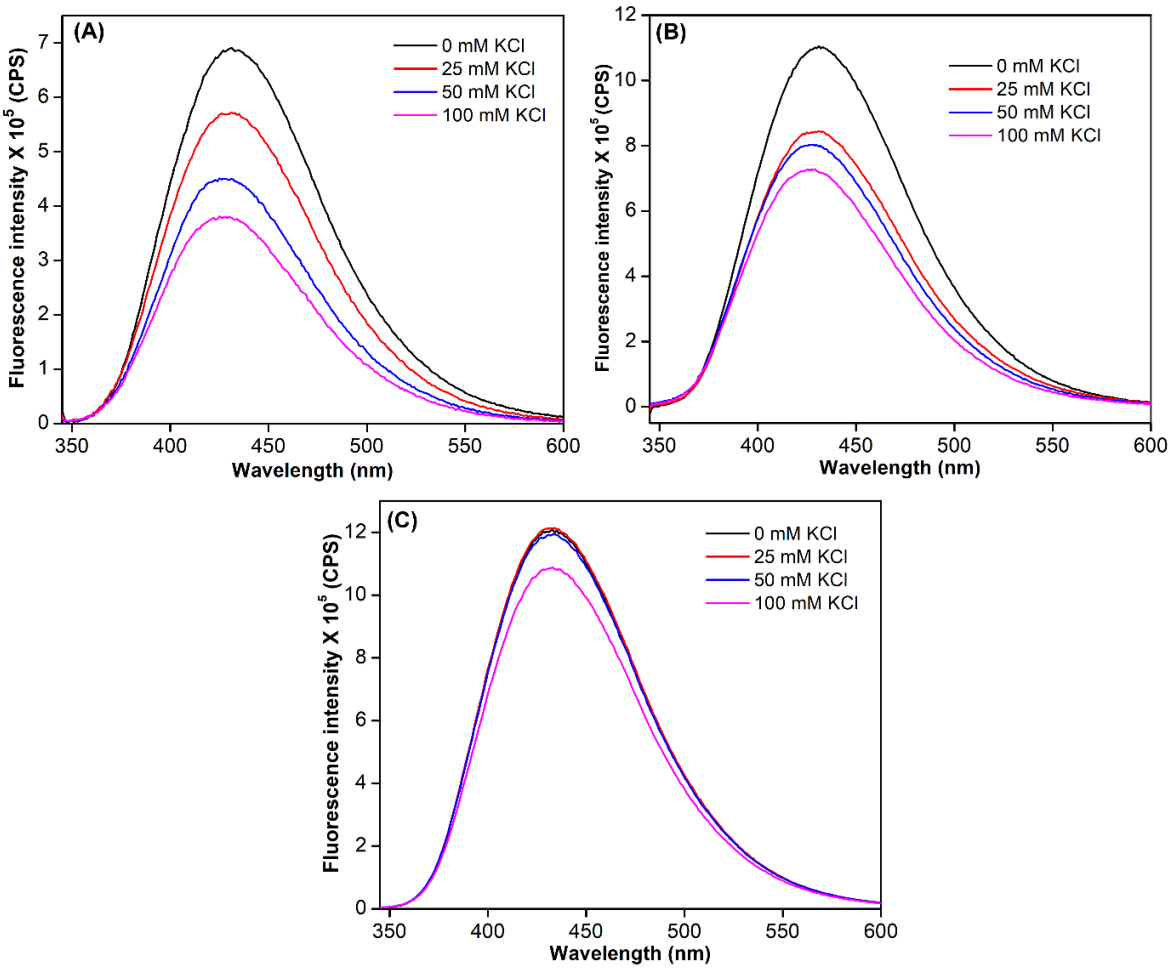


Figure 8. Steady-state fluorescence spectra at different KCl concentrations of (**A**) modified ON **2** (0.5 µM) (**B**) modified ON **3** (0.5 µM) (**C**) nucleoside FBFU (**1**) (5 µM).

4.2.4 Probing SARS-CoV-2 Nsp3 GQs using ^{19}F NMR

While the aforementioned techniques confirm the formation of GQ structures, they do not provide information about the coexisting forms, thus offering limited insights into their structural polymorphism. Thus, we opted for ^{19}F NMR signatures to gain a clearer understanding of the structural polymorphism and coexisting forms of GQs. ON **2** was first annealed in Tris-HCl buffer without KCl, and the resulting ^{19}F NMR spectrum displayed a peak at −122.18 ppm, indicative of a non-GQ form. This is also evident from the ^{1}H NMR as signified by the absence of imino proton signals in 10−12 ppm range (Figure 9).[57] ON **2** in the presence of 25 mM KCl produced two distinct peaks – one broader peak at −122.31 ppm and another sharper peak at −121.86 ppm (Figure 9). This signifies the presence of multiple GQ structures resembling the parallel form (*vide supra*, see CD spectra Figure 5A). Subsequent increments in K+ ions (50 mM and 100 mM) resulted in the retention of these two signals, albeit with slight chemical shifts (Figure 9), which was further corroborated by the ^{1}H NMR,

indicating the formation of the GQ structures at different KCl concentrations. ^{19}F NMR of ON **3**, recorded in 100 mM K+, exhibited two broad peaks. The ^{19}F NMR of the nucleoside exhibited minimal changes across different K^+ conditions, suggesting that the slight peak variations observed in the spectra during GQ formation result from structural transitions (Figure 10). Despite ^{1}H NMR indicating the formation of the GQ structure, the broad nature of the ^{19}F NMR signals rendered the sequence not much suitable for further experimental analysis (Figure 11).

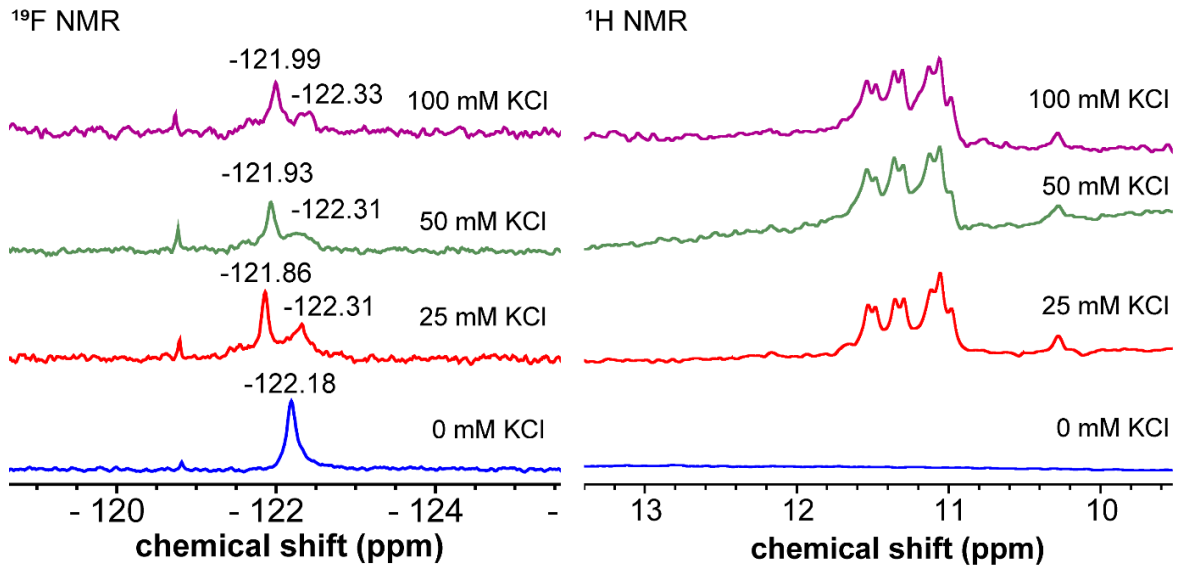


Figure 9. ^{19}F NMR and ^{1}H NMR spectra of modified ON **2** (45 µM) at different KCl concentrations.

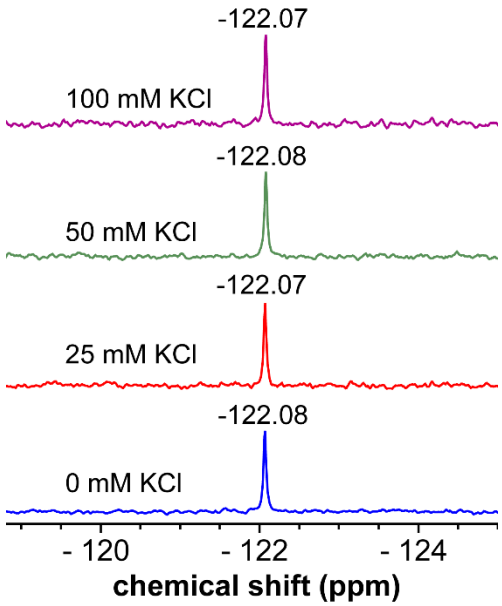


Figure 10. ^{19}F NMR of the nucleoside FBFU (**1**)

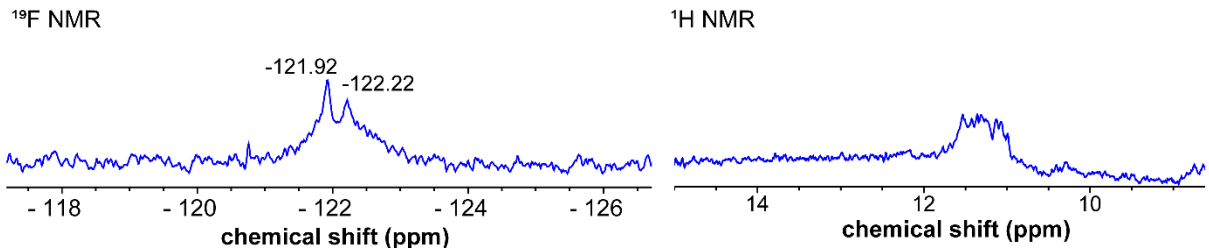


Figure 11. ^{19}F and ^{1}H NMR of ON **3** (45 µM).

In general, parallel GQ structures can self-assemble into higher-order GQ structures by the stacking of two or more GQ structures via end-to-end stacking interaction.[58] However, it remains unclear whether RNA sequences with GG motifs, like those in the Nsp3 G-rich segment, can independently fold into an intramolecular two-tetrad GQ. Alternatively, they may form intermolecular oligomers, a question of particular interest given their frequent occurrence in biologically significant RNA regions.[59] To assess the structural distribution of different GQ structures, a temperature-dependent ^{19}F NMR experiment was performed in a buffer containing 100 mM KCl (Figure 11). At 5 °C and 15 °C, although the formation of GQ structures was evident, both ^{19}F and ^{1}H spectra were very broad. We speculated that these broadened signals could arise due to the formation of monomeric as well as different assemblies of higher-order GQ structures at lower temperatures, which may be interconverting. As the temperature was increased to 25 °C, we observed a clear peak at −121.9 ppm. Interestingly, a further increase in temperature (37 °C) produced two distinct peaks at −121.9 ppm and −122.2 ppm. At 45 °C, the peak at −121.9 ppm diminished and concurrently the peak at −122.2 ppm increased, likely due to the conversion of a higher-order GQ structure into a monomeric form.[58] Further, the ^{19}F NMR of the nucleoside FBFU (1) was recorded at different temperatures, revealing marginal downfield peak shifts. These shifts correspond to the temperature-dependent changes and alludes to the slight peak shifts for the GQ structures varied with temperature (Figure 12). Although the ^{1}H NMR spectra at different temperatures confirmed the formation of GQ structures, they lacked detailed insights into structural features. These finer details were obtained using the responsiveness of our ^{19}F NMR-labeled nucleoside probe, which revealed that, near physiological temperature and K^+ ion concentration, the Nsp3 ON forms both monomeric and higher-order GQ structures.

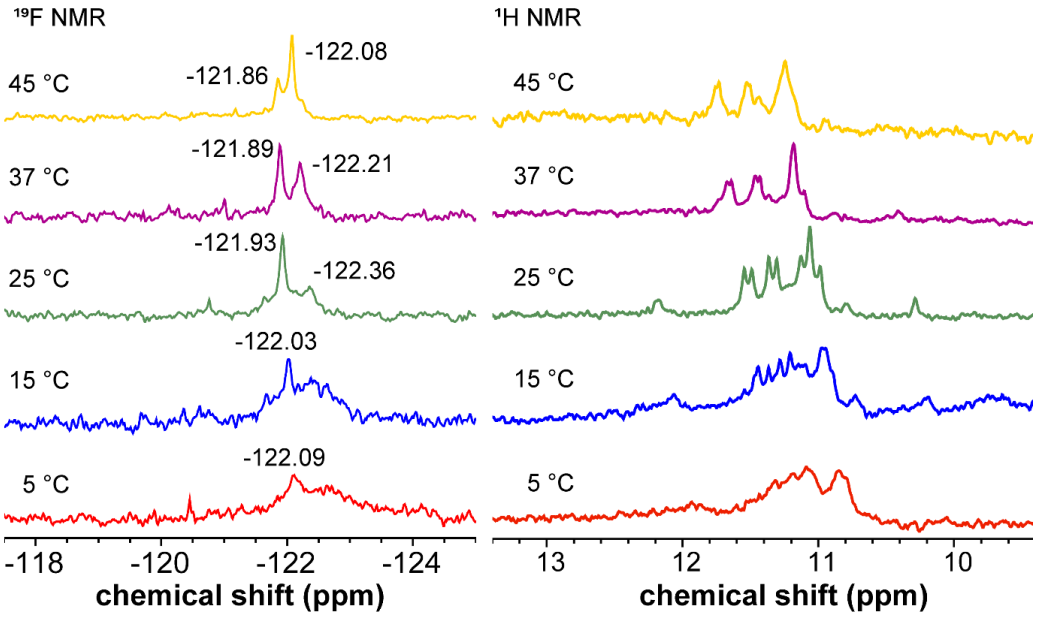


Figure 12. ^{19}F NMR and ^{1}H NMR spectra of modified ON **2** (45 µM) at different temperatures.

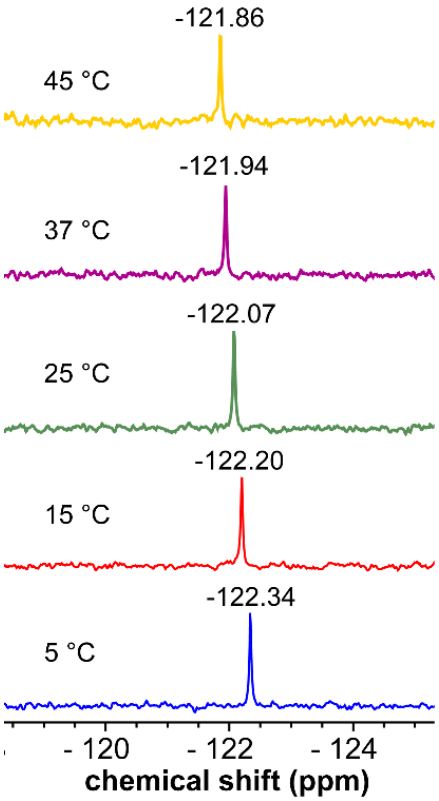


Figure 13. ^{19}F NMR of FBFU (**1**) in 100 mM KCl at different temperatures.

Further evidence for the formation of higher-order structures was obtained from electrophoretic mobility shift assay (EMSA) under non-denaturing conditions. GQs of unmodified Nsp3 ON **1** and modified ONs **2** and **3** (19-mer) annealed in a buffer containing 100 mM KCl were resolved using native polyacrylamide gel electrophoresis at 5 °C and gel

bands were analyzed by UV-shadowing and also by using Stains-All reagent (Figure 12). Appropriate control ON sequences forming duplex (**C•C1**, 34 base-pairs), monomeric GQ (**C2**, 20-mer), dimeric GQ (**C3**, 20-mer) structures, and a longer sequence forming GQ structures (**C4**, 65-mer) were used for comparison (Table 4). In general, the monomeric GQ form runs faster as compared to the dimeric and multimeric GQ structures.[60] ONs **1**–**3** exhibited two bands (lanes 2–4, Figure 11B). The faster migrating band had mobility similar to the control monomeric GQ structure of ON **C2** of similar sequence length, and hence, this band could be considered as the monomeric GQ structure of ONs **1**–**3** (compare lanes 2–4 with lane 5). The slower moving band runs similar to the 34 base-pair duplex and the GQ form of the 65-mer G-rich sequence **C4** (compare lanes 2–4 with lane 1 and lane 7). Therefore, the slower migrating band of ONs **1**–**3** could be associated to a higher-order structure, wherein three or more parallel structures would have assembled one-above the other. Taken together, this study underscores the utility of FBFU in detecting co-existing GQ structures under equilibrium conditions, which otherwise would be difficult to elucidate by most other biophysical tools.

Table 4. Sequence of control ONs used in EMSA.

ON	5′----------3′
C1	GGGAGGCGTGGCCTGGGCGGGACTGGGG
C	CCCCAGTCCCGCCCAGGCCACGCCTCCC
C2	TTGGGTGGGTGGGTGGGT[59]
C3	GGGTGGGTGGGTGGGT[59]
C4	TTTTTGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGTTTTTCTGCATATAAGCAGCTGCTTTTTGCC

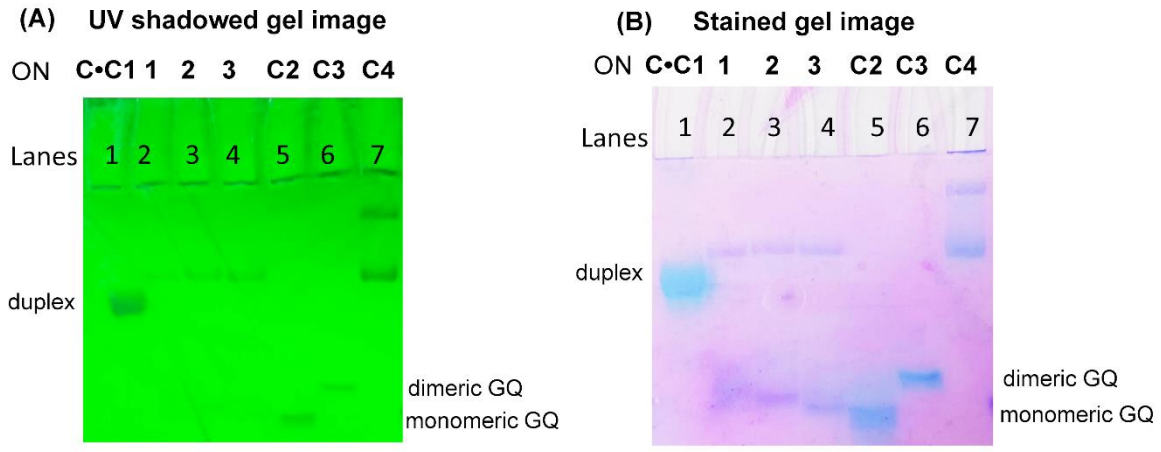


Figure 14. (**A**) UV-shadowed image of an 18% native PAGE, showing the migration of ON sequences under varying conditions. Lane 1: control duplex (**C•C1**), Lane 2–4: ON **1**–**3**, Lane 5–7: ON **C2**, **C3**, **C4**. (**B**) Gel stained with Stains-All dye, confirming the presence of dimeric and monomeric GQ structures, along with the duplex form.

4.2.5 Probing ligand binding interaction of Nsp3 GQs

Structurally distinct ligands, such as BRACO19 and TMPyP4, have been widely used to stabilize GQs. For Nsp3 GQs, their interactions with the ligands have been studied using CD and T_m experiments.[22] Since the binding affinity of the ligands was not estimated *in vitro*, we aimed to assess the dissociation constant of the ligand binding using the fluorescence of FBFU. When a ligand binds to a GQ structure, the fluorescence of FBFdU located in the loop region decreases significantly due to its proximity to the ligand.[42] We envisioned that the FBFU would also behave the same way, and hence, it was utilized in devising a fluorescence method to estimate the binding affinity of Nsp3 RNA ON **2** GQs with two commercially available binders namely, TMPyP4 and BRACO19 (Figure 13A). Upon increasing the concentration of both the ligands, there is a dose-dependent quenching in the fluorescence intensity with marginal changes in the emission maximum (Figure 14A and 14B). Normalized fluorescence intensity plotted against ligand concentration and fitted to the Hill equation produced a sigmoidal curve. TMPyP4 exhibited higher binding affinity with an apparent K_d of 0.18 ± 0.02 μM as compared to BRACO19, which showed a K_d of 0.39 ± 0.05 μM (Figure 13B). The ligand interaction by TMPyP4 or BRACO19 could not be detected by ^{19}F NMR, possibly due to the absence of a distinct signature peak for the GQ structure in ^{1}H NMR (Figure 17). This lack of a clear ^{1}H NMR signature for the GQ conformation might stem from the dynamic nature of GQ structures, which can lead to broadening or overlapping of peaks, making specific resonances difficult to detect. These limitations in detecting the GQ structure hinder the ability to observe ligand-induced changes. Such challenges are well-documented in studies that emphasize the difficulty of resolving GQ structures and their interactions with ligands in solution NMR due to dynamic and exchange processes (Figure 17).[61,62]

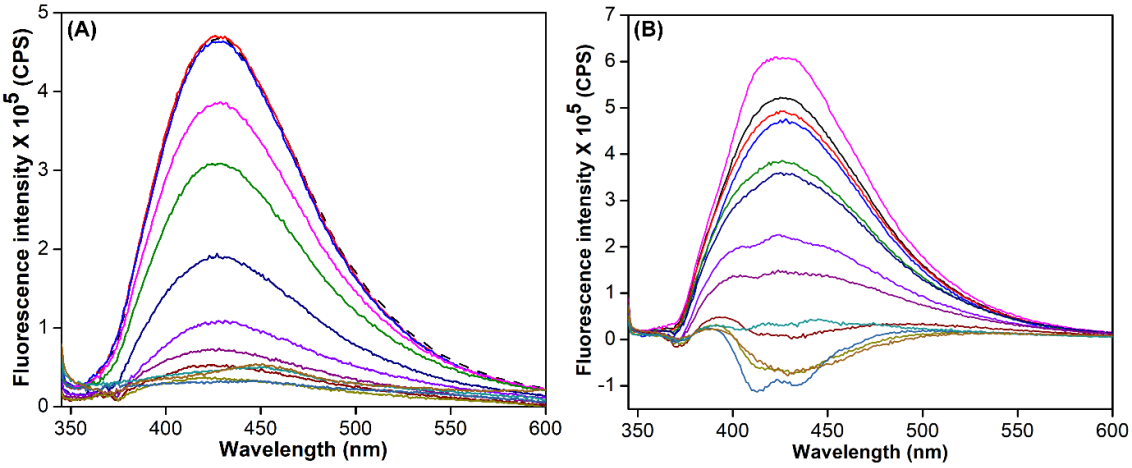


Figure 15. Emission spectra for the titration of modified SARS-CoV-2 Nsp3 ON **2** (0.5 µM) with increasing concentration of (**A**) TMPyP4 (12.5 nM–2.5 µM) and (**B**) BRACO19 (10 nM–2.5 µM). Samples were excited at 330 nm with an excitation and emission slit widths of 7 nm and 8 nm, respectively. The black line represents the spectrum of ON **2** without any ligand. See experimental details.

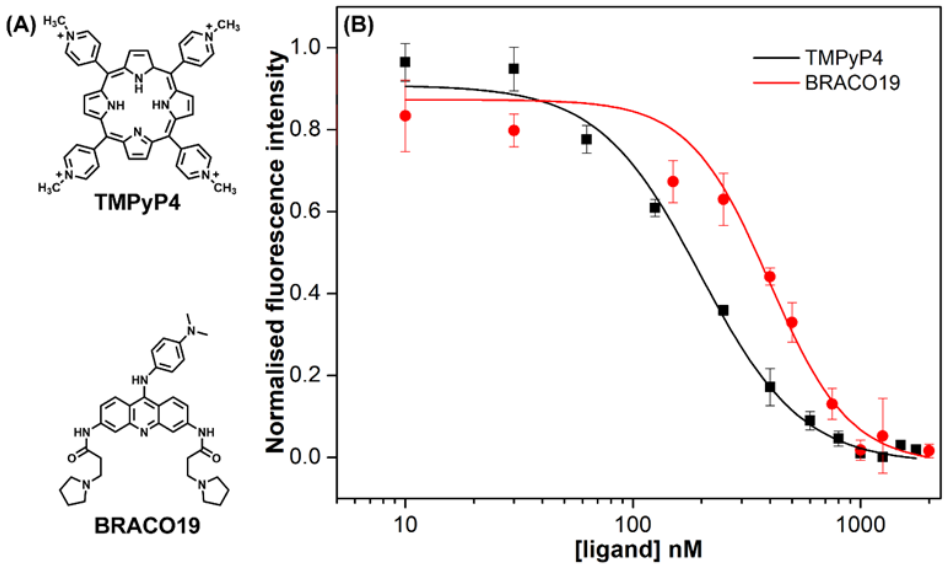


Figure 16. (**A**) Chemical structures of GQ-binding ligands used in this study. (**B**) Curve fits for the binding of ligands TMPyP4 and BRACO19 to RNA ON **2** (0.5 µM).

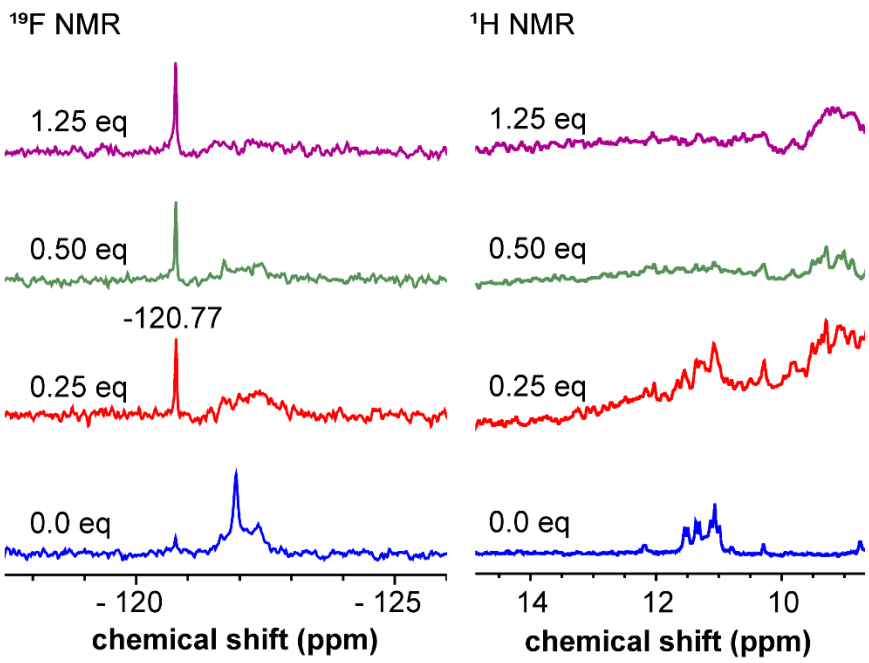


Figure 17. ^{19}F NMR titration of ON **2** with TMPyP4.

4.3 Conclusions

With the frequent emergence of SARS-CoV-2 variants that could compromise the effectiveness of existing vaccines and antiviral drugs, targeting viral GQs presents a promising alternative therapeutic approach. In this context, we have studied the structural dynamics and recognition properties of the highly conserved SARS-CoV-2 Nsp3 genome. The probe successfully detected the multiple structures of the GQ, providing better insights into the structural distribution of the higher-order and monomeric GQs. Also, ligand binding was estimated, demonstrating a high affinity of TMPyP4 and BRACO19 with Nsp3 GQ. We envision that such an understanding would present a new dimension in designing specific drug molecules considering the presence of the multiple GQ structures.

4.4 Experimental Section

4.4.1 Materials: 5-Flurobenzofuran modified uridine (**1**) and it's phosphoramidite substrates (**1a**) were synthesized following a reported procedure.[49] *N,N,N',N'*-tetramethylethylenediamine, *n*-butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) dichloride, *tert*-butyldimethylsilyl chloride, silver nitrate, *N,N*-diisopropylethylamine, anhydrous DMSO, Triethyl amine trishydrofluoride were purchased from Sigma-Aldrich. 2-cyanoethyl-*N,-N*-diisopropylchlorophosphoramidite was purchased from Entegris. TBDMS-protected ribonucleoside phosphoramidites for RNA synthesis were

obtained from Glen Research and Innovasynth. Solid supports for RNA synthesis were procured from Glen Research. Desalting cartridges Sep-Pak Classic C18 cartridges were brought from (Waters Corporation). BRACO19 hydrochloride and all reagents (Bio-Ultra grade) used in the preparation of buffers were purchased from Sigma-Aldrich. TMPyP4 was procured from Merck-Millipore. Autoclaved Millipore water was used for the preparation of all buffer solutions and in all biophysical studies.

4.4.2 Instrumentation: NMR spectra for small molecules were obtained using a Bruker AVANCE III HD ASCEND 400 MHz spectrometer and analyzed with Mnova software (Mestrelab Research). Mass spectrometry data were acquired using an ESI-MS Waters Synapt G2-Si Mass Spectrometer. RNA oligonucleotides (ONs) were synthesized on a K&A DNA/RNA Synthesizer H6. Reverse-phase HPLC (RP-HPLC) analysis was conducted with an Agilent Technologies 1260 Infinity HPLC system. Absorption spectra were measured on a Shimadzu UV-2600 spectrophotometer, while fluorescence measurements for the ONs were performed on a Horiba Scientific Fluoromax-4 spectrophotometer. UV-thermal melting studies were carried out using a Cary 3500 multicell UV-Vis spectrophotometer, and circular dichroism (CD) spectra were recorded on a JASCO J-815 CD spectrometer. NMR spectra for the ONs were recorded on a Bruker AVANCE III HD ASCEND 600 MHz spectrometer with a Cryo-Probe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT) and processed using Bruker TopSpin software.

4.4.3 Solid phase RNA synthesis: Control ON (**1**) and FBFU (**1**) modified RNA ONs **2** and **3** were synthesized on a 1 μmole scale (1000 Å CPG solid support) with K&A H-6 synthesizer. A coupling time of 6 min was set for each natural base phosphoramidite while modified phosphoramidite FBFU (**1a**) underwent double coupling of 12 min (total 24 min) each. After each oxidation step, a second capping step was performed to maximize the removal of water molecules that can interfere with the next coupling reaction. Post-synthesis, the ONs were cleaved from the solid support using a 1:1 mixture of 30% aqueous ammonia and 40% methylamine in water in a 2 mL screw cap vial (without rubber O-ring), heated at 65 °C (1 h), and then left at room temperature (2 h). The supernatant was transferred to 1.5 mL centrifuge tubes and allowed to get completely dried in SpeedVac. To remove the 2´-TBDMS group, the dried product was re-dissolved in anhydrous DMSO (100 μL), and $Et_3N{\cdot}3HF$ (125 μL) was added. The mixture was heated at 65 °C for 2.5 h at 900 rpm in a Thermomixer. Finally, the reaction mixture was cooled in ice-bath followed by the addition of 5 M NaOAc (25 μL, pH 7). It was vortexed for 30 sec. 750 μL of cold *n*-Butanol was then added, it was again mixed

well. The centrifuge tubes were cooled at −40 °C overnight. It was allowed to centrifuge at 13000 rpm at 4 °C for 15 min. The supernatant was carefully removed followed by washing the pellet with cold 70% EtOH (2 X 750 µL) and finally dried in SpeedVac. The pellet was re-dissolved in a mixture of autoclaved water and denaturing loading buffer. ONs were purified by denaturing PAGE (20% gel) and the product bands were identified by UV-shadowing. The bands were cut, transferred to poly prep columns (Bio-Rad), and crushed using a sterile glass rod. It was extracted with 4 mL of 0.5 M ammonium acetate buffer in a poly-prep column for 12 h. Desalting was performed using a Waters C18 cartridge. The purity of ONs was monitored by RP-HPLC (Figure S1). The integrity was verified by ESI-MS.

4.4.4 ESI-MS analysis: Negative mode ESI-MS was used to analyze RNA oligonucleotides (ONs) (~800 pmol), which were dissolved in a mixture of 50% acetonitrile, 10 mM triethylamine, and 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol in water.

4.4.5 Circular dichroism: ONs **1**–**3** (10 µM) were annealed by heating at 85 °C for 5 minutes in a solution of 10 mM Tris-HCl (pH 7.4), containing 100 mM KCl followed by gradual cooling to RT. CD spectra were recorded on a JASCO J-810 spectropolarimeter using a 1 mm path length cuvette. Each spectrum was collected in a 320–220 nm range at a scan rate of 100 nm min^{-1} with three acquisitions at 25 °C. Baseline corrected data were smoothed using the JASCO graphing software.

4.4.6 Thermal melting analysis: ONs **1**–**3** (10 µM) were annealed by heating at 85 °C for 5 minutes in a solution of 10 mM Tris-HCl (pH 7.4), containing 100 mM KCl, as mentioned above. Spectra were collected in Cary 3500 Multicell UV-Vis spectrophotometer using a temperature interval of 1 °C. Absorbance was monitored at 295 nm with a data interval of 0.5 °C. Experiments were performed in duplicate.

4.4.7 Thermal differential spectra: ONs **1**–**3** (10 µM) were annealed by heating at 85 °C for 5 minutes in a solution of 10 mM Tris-HCl (pH 7.4), containing 100 mM KCl, as mentioned above. Thermal difference spectra (TDS) were recorded in the Shimadzu UV-2600 spectrophotometer. At first, UV cuvettes containing the samples were scanned from (200–320) nm individually at 20 °C. Samples were then incubated in the sample holder while the temperature was increased to 70 °C using the temperature controller, and the individual scans were recorded. Spectra were recorded in duplicate. TDS was obtained by subtracting the absorbance spectrum at 20 °C from the absorbance spectrum at 70 °C for each sample after baseline correction in the instrument.

4.4.8 Fluorescence: ONs **1**–**3** (10 µM) were annealed by heating at 85 °C for 5 minutes in a solution of 10 mM Tris-HCl (pH 7.4), containing the absence and presence of different concentrations of potassium (25–100) mM. Spectra were recorded in triplicate in a micro-fluorescence cuvette (Hellma, path length 1.0 cm) on a Fluoromax-4 spectrofluorometer (Horiba Scientific) at 25 °C. Spectra of FBFU (**1**) in different concentrations were also recorded as mentioned. Excitation and emission slit widths of 7 nm and 8 nm are set, respectively.

4.4.9 NMR: ONs **2** and **3** (45 µM) were annealed by heating the samples at 85 °C for 5 min 10 mM Tris-HCl (pH 7.4). For ion-dependent experiments, samples were treated with increasing concentrations of KCl and incubated at 25 °C for 30 minutes. NMR spectra were acquired using a Bruker AVANCE III HD ASCEND 600 MHz spectrometer equipped with a CryoProbe (CP2.1 QCl 600S3 H/F-C/N-D-05 Z XT). ^{19}F NMR data were collected at a frequency of 564.9 MHz and referenced to an external standard, trifluorotoluene (TFT, −63.72 ppm). The acquisition parameters for ^{19}F NMR included a 12 µs excitation pulse, a spectral width of 90.32 ppm, a transmitter offset of −145 ppm, an acquisition time of 0.33 s, a relaxation delay of 1.0 s, and 5000–6000 scans, with each spectrum requiring approximately 2–2.5 hours. Data processing involved applying an exponential window function with a line broadening of 20 Hz. ^{1}H NMR spectra were recorded at 600 MHz using excitation sculpting with gradients for water suppression, with a total of 1200 scans per spectrum.

4.4.10 EMSA: SARS-CoV-2 Nsp3 ON **1** and modified ONs **2**, **3**, control ONs **C2** and **C3** were annealed in 100 mM KCl at 85 °C for 5 min. The native PAGE and running buffer were prepared by adding 10 mM Tris-HCl and 100 mM KCl. Sucrose was added to each sample in the loading buffer before loading. The gel was run at 25 W in a cold room for 5 h. Gel was stained with Stains- All dye and destained with 70% EtOH.

4.4.11 Ligand binding studies of modified SARS-CoV-2 ONs by fluorescence: SARS-CoV-2 GQ structures of ON **2** (0.5 µM), were formed by heating the samples at 85 °C for 5 min in 10 mM Tris-HCl buffer (pH 7.4) buffer containing 100 mM KCl. The samples were incubated with increasing concentrations of the ligands (TMPyP4 and BRACO19) at 25 °C for 1 h. Samples were excited at 330 nm with excitation and emission wavelength slit widths of 7 nm and 8 nm, respectively. The final volume of each sample solution was kept at 200 µL. Fluorescence experiments were performed in triplicate in a micro-fluorescence cuvette at 25 °C. An appropriate blank containing a ligand in a buffer was subtracted from each reading for the corresponding ligand concentration. Normalized fluorescence intensity (FN) against ligand

concentration was plotted and fitted to a Hill equation to determine the apparent K_d values. Fitted graphs were prepared using OriginPro 8.5 software.

4.5 References

1. Guo, Y. R.; Cao, Q. D.; Hong, Z. S.; Tan, Y. Y.; Chen, S. D.; Jin, H. J.; Tan, K. S.; Wang, D. Y.; Yan, Y. The origin, transmission, and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil. Med. Res*., **2020**, *7*, 11.
2. Dhama, K.; Khan, S.; Tiwari, R.; Sircar, S.; Bhat, S.; Malik, Y. S.; Singh, K. P.; Chaicumpa, W.; Bonilla-Aldana, D. K.; Rodriguez-Morales, A. J. Coronavirus disease 2019-COVID-19. *Clin. Microbiol. Rev.*, **2020**, *33*, e00028-20
3. Li, J.; Lai, S.; Gao, G. F.; Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **2021**, *600*, 408–418
4. Chen, Y.; Liu, Q.; Guo, D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.*, **2020**, *92*, 418–423.
5. Kim, D.; Lee, J. Y.; Yang, J. S.; Kim, J. W.; Kim, V. N.; Chang, H. The architecture of SARS-CoV-2 transcriptome. *Cell,* **2020,** *181*, 914–921.e10.
6. Zu, S.; Deng, Y. Q.; Zhou, C.; Li, J.; Li, L.; Chen, Q.; Li, X. F.; Zhao, H.; Gold, S.; He, J.; Li, X.; Zhang, C.; Yang, H.; Cheng, G.; Qin, C. F. 25-Hydroxycholesterol is a potent SARS-CoV-2 inhibitor. *Cell Res.* **2020**, *30*, 1043–1045.
7. Qiao, J.; Li, Y. S.; Zeng, R. et al. SARS-CoV-2 M (pro) inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371*, 1374–1378.
8. Yin, W.; Luan, X.; Li, Z. Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. *Nat. Struct. Mol. Biol.* **2021**, *28*, 319–325.
9. White, K. M.; Rosales, R.; Yildiz, S. et al. Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A. *Science* **2021**, 371, 926–931.
10. Yuan, S.; Yin, X.; Meng X. et al. Clofazimine broadly inhibits coronaviruses including SARS-CoV-2. *Nature* **2021**, *593*, 418–423.
11. Thakur, S.; Sasi, S.; Pillai, S. G.; Nag, A.; Shukla, D.; Singhal, R.; Phalke, S.; Velu, G. S. K. SARS-CoV-2 mutations and their impact on diagnostics, therapeutics, and vaccines. *Front. Med.* **2022**, *9*, 815389–815389.
12. Carabelli, A. M.; Peacock, T. P.; Thorne, L. G. et al. SARS-CoV-2 variant biology: immune escape, transmission, and fitness. *Nat. Rev. Microbiol*. **2023**, *21*, 162–177.
13. Xu, J.; Huang, H.; Zhou, Xiang. G-quadruplexes in neurobiology and virology: functional roles and potential therapeutic approaches. *JACS Au* **2021**, *1*, 2146−2161
14. Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palù, G.; Brazzale, A. R.; Richter, S. N.; Toppo, S. G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Comput. Biol.* **2018**, 14, e1006675.
15. Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palu, G.; Richter, S. N. A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. *J. Med. Chem.* **2013**, *56*, 6521–6530.
16. Wang, S. R.; Min, Y. Q.; Wang, J. Q.; Liu, C. X.; Fu, B. S.; Wu, F.; Wu, L. Y.; Qiao, Z. X.; Song, Y. Y.; Xu, G. H.; et al. A highly conserved G-rich consensus sequence in hepatitis C virus core gene represents a new anti-hepatitis C target. *Sci. Adv.* **2016**, *2*, e1501535.

17. Majee, P.; Pattnaik, A.; Sahoo, B. R.; Shankar, U.; Pattnaik, A. K.; Kumar, A.; Nayak, D. Inhibition of Zika virus replication by G-quadruplex-binding ligands. *Mol. Ther. Nucleic Acids* **2021**, *23*, 691–701.
18. Ruggiero, E.; Zanin, I.; Terreri, M.; Richter, S. N. G-quadruplex targeting in the fight against viruses: an update. *Int. J. Mol. Sci*. **2021**, *22*, 10984.
19. Lavezzo, E.; Berselli, M.; Frasson, I. Perrone, R.; Palu, G.; Brazzale, A. R.; Richter, S.N.; Toppo, S. G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Comput. Biol*. **2018**, *14*, e1006675.
20. Gussow, A. B.; Auslander, N.; Faure, G.; Wolf, Y. I.; Zhang, F.; Koonin, E. V. Genomic determinants of pathogenicity in SARS CoV-2 and other human coronaviruses. *Proc. Natl. Acad. Sci. U S A*. **2020,** *117*, 15193–15199.
21. Jahirul Islam, M.; Nawal Islam, N.; Siddik Alom, M.; Kabir, M.; Halim, M. A. A review on structural, non-structural, and accessory proteins of SARS-CoV-2: Highlighting drug target sites. *Immunobiology* **2023**, *228*, 152302.
22. Qin, G.; Zhao, C.; Liu, Y.; Zhang, C.; Yang, G.; Yang, J.; Wang, Z.; Wang, C.; Tu, C.; Guo, Z.; Ren, J.; Qu, X. RNA G-quadruplex formed in SARS-CoV-2 used for COVID-19 treatment in animal models. *Cell Discov*. **2022**, *8*, 86.
23. Belmonte-Reche, E.; Serrano-Chacón, I.; Gonzalez, C.; Gallo, J.; Bañobre-López, M. Exploring G and C-quadruplex structures as potential targets against the severe acute respiratory syndrome coronavirus 2. *Bioinformatics* **2020**, preprint. doi:10.1101/2020.08.19.257493.
24. Liu, G.; Du, W.; Sang, X.; Tong, Q.; Wang, Y.; Chen, G, et al. RNA G-quadruplex in TMPRSS2 reduces SARS-CoV-2 infection. *Nat. Commun.* **2022**, *13*, 1444.
25. Ji, D.; Juhas, M.; Tsang, C. M.; Kwok, C. K.; Li, Y.; Zhang, Y. Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief Bioinform.* **2021**, *22*, 1150–60.
26. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X. Targeting RNA G-quadruplex in SARS-CoV-2: a promising therapeutic target for COVID-19? *Angew Chem. Int. Ed. Engl.* **2021**, *60*, 432–8.
27. Wang, Y.; Grunewald, M.; Perlman, S. Coronaviruses: an updated overview of their replication and pathogenesis. In: Maier, H., Bickerton, E. (eds) Coronaviruses. *Methods Mol. Biol.* **2020**, *2203*.
28. Song, J.; Perreault, J. P.; Topisirovic, I.; Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **2016**, *4*, 1–16. e1244031
29. Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.*, 2018, **46**, 9264–9275.
30. Lacroix, L.; Séosse, A.; Mergny, J.-L. Fluorescence-based duplex–quadruplex competition test to screen for telomerase RNA quadruplex ligands. *Nucleic Acids Res.*, **2011**, *39*, e21–e31.
31. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the human telomere in K^+ solution: An intramolecular (3 + 1) G-quadruplex scaffold *J. Am. Chem. Soc.*, **2006**, *128*, 9963–9970.
32. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **2002**, *417*, 876 –880.
33. Biffi, G.; Di Antonio, M.; Tannahill, D. *et al.* Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature Chem*. **2014**, *6*, 75–80.
34. Zheng, B. X.; Long, W.; She, M. T.; Wang, Y.; Zhao, D.; Yu, J.; Siu-Lun Leung, A.; Hin Chan, K.; Hou, J.; Lu, Y. J.; Wong, W. L. A Cytoplasm-specific fluorescent ligand

for selective imaging of RNA G-quadruplexes in live cancer cells. *Chem. Eur. J.* **2023**, *29*, e202300705.
35. Chen, X. C.; Chen, S. B.; Dai, J.; Yuan, J. H.; Ou, T. M.; Huang, Z. S.; Tan, J. H. Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angewandte Chemie* **2018,** *130*, 4792–4796.
36. Robinson, J.; Stenspil, S. G.; Maleckaite, K.; Bartlett, M.; Di Antonio, M.; Vilar, R.; Kuimova, M. K. Cellular visualization of G-quadruplex RNA via fluorescence- lifetime imaging microscopy. *J. Am. Chem. Soc*., **2024**, *146*, 1009–1018.
37. Yuan, J. H.; Shao, W.; Chen, S. B.; Huang, Z. S.; Tan, J. H. Recent advances in fluorescent probes for G-quadruplex nucleic acids. *Biochem. Biophys. Res. Commun.*, **2020**, *531*, 18–24.
38. Ahmed, A. A.; Angell, R.; Oxenford, S.; Worthington, J.; Williams, N.; Barton, N.; Fowler, T. G.; O'Flynn, D. E.; Sunose, M.; McConville, M.; Vo, T.; Wilson, W. D.; Karim, S. A.; Morton, J. P.; Neidle, S. Substituted naphthalenediimide compounds bind selectively to two human quadruplex structures with parallel topology. *ACS Med. Chem. Lett.*, **2020**, *11*, 991–999.
39. Kumar, S.; Pany, S. P. P.; Sudhakar, S.; Singh, S. B.; Todankar, C. S.; Pradeepkumar, P. I. Targeting parallel topology of G-quadruplex structures by indole-fused quindoline scaffolds. *Biochemistry*, **2022**, *61*, 2546–2559.
40. Takahashi, S.; Kotar, A.; Tateishi-Karimata, H.; Bhowmik, S.; Wang, Z. F.; Chang, T. C.; Sato, S.; Takenaka, S.; Plavec, J.; Sugimoto, N. Chemical modulation of DNA replication along G-quadruplex based on topology-dependent ligand binding. *J. Am. Chem. Soc.*, **2021**, *143*, 16458–16469.
41. Sun, H.; Sun, R.; Yang, D.; Li, Q.; Jiang, W.; Zhou, T.; Bai, R.; Zhong, F.; Zhang, B.; Xiang, J.; Liu, J.; Tang, Y.; Yao, L. A cyanine dye for highly specific recognition of parallel G-quadruplex topology and its application in clinical RNA detection for cancer diagnosis. *J Am Chem Soc*. **2024,** *146*, 22736–22746.
42. Manna, S.; Sarkar, D.; Srivatsan, S. G. A dual-app nucleoside probe provides structural insights into the human telomeric overhang in live cells. *J. Am. Chem. Soc.*, **2018**, *140*, 12622–12633.
43. 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.
44. Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. Fluorescent Nucleoside Analogs: Probes for Investigating Nucleic Acid Structure and Function. *Isr. J. Chem.* **2013**, *53*, 366–378.
45. Manna, S.; Srivatsan, S. G. Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.* **2018**, *8*, 25673–25694.
46. Cheng, M.; Cheng, Yu.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.*, **2018**, *46*, 9264–9275.
47. Jana, J.; Vianney, Y. M.; Schröder, N.; Weisz, K. Guiding the folding of G-quadruplexes through loop residue interactions. *Nucleic Acids Res.*, **2022**, *50*, 7161–7175.
48. Roy, S.; Majee, P.; Sudhakar, S.; Misra, S.; Kalia, S.; Pradeepkumar, P. I. Structural elucidation of HIV-1 G-quadruplexes in a cellular environment and their ligand binding using responsive ^{19}F-labeled nucleoside probes. *Chem. Sci.*, **2024**, *15*, 7982–7991.

49. Manna, S.; Sontakke, V. A.; Srivatsan, S. G. Incorporation and utility of a responsive ribonucleoside analogue in probing the conformation of a viral RNA motif by fluorescence and ^{19}F NMR spectroscopy. *Chembiochem* **2022**, *23*, e202100601.
50. Cheng, M.; Cheng, Yu.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.*, **2018**, *46*, 9264–9275.
51. Jana, J.; Vianney, Y. M.; Schröder, N.; Weisz, K. Guiding the folding of G-quadruplexes through loop residue interactions. *Nucleic Acids Res.*, **2022**, *50*, 7161–7175.
52. Kabbara, A.; Vialet, B.; Marquevielle, J.; Bonnafous, P.; Mackereth, C. D.; Amrane, S. RNA G-quadruplex forming regions from SARS-2, SARS-1 and MERS coronoviruses. *Front Chem*. **2022**, *10*, 1014663.
53. Murat, P.; Singh, Y.; Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions. *Chem Soc Rev*. **2011**, *40*, 5293–5307.
54. Basu, P.; Kejnovská, I.; Gajarský, M.; Šubert, D.; Mikešová, T.; Renčiuk, D.; Trantírek, L.; Mergny, J. L.; Vorlíčková, M. RNA G-quadruplex formation in biologically important transcribed regions: can two-tetrad intramolecular RNA quadruplexes be formed? *Nucleic Acids Res.*, **2024**, gkae927. Advance online publication.
55. Vorlícková, M.; Kejnovská, I.; Bednárová, K.; Renciuk, D.; Kypr, J. Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, **2012**, *24*, 691–698.
56. Bezzi, G.; Piga, E. J.; Binolfi, A.; Armas, P. CNBP Binds and Unfolds In Vitro G-Quadruplexes Formed in the SARS-CoV-2 Positive and Negative Genome Strands. *Int. J. Mol. Sci.* **2021**, *22*, 2614.
57. Adrian, M.; Heddi, B.; Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods* **2012**, *57*, 11–24.
58. Islam, B.; Stadlbauer, P.; Vorlickova, M.; Mergny, J.-L.; Otyepka, M.; Sponer, J. Stability of two-quartet G-quadruplexes and their dimers in atomistic simulations. *J. Chem. Theory Comput.* **2020**, *16*, 3447– 3463,
59. Do, N. Q.; Phan, A. T. Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry* **2012**, *18*, 14752–14759.
60. Krafčík, D.; Ištvánková, E.; Džatko, Š.; Víšková, P.; Foldynová-Trantírková, S.; Trantírek, L. Towards Profiling of the G-Quadruplex Targeting Drugs in the Living Human Cells Using NMR Spectroscopy. *Int. J. Mol. Sci.*, **2021**, *22*, 6042.
61. Lin, C.; Dickerhoff, J.; Yang, D. NMR Studies of G-quadruplex structures and G-quadruplex-interactive compounds. *Methods Mol Biol.* **2019**, *2035*, 157–176.
62. Kolesnikova, S.; Hubálek, M.; Bednárová, L.; Cvačka, J.; Curtis, E. A. Multimerization rules for G-quadruplexes, *Nucleic Acids Res.*, **2017**, *45*, 8684–8696,

List of publications

1. **Roy, S.;** Majee, P.; Sudhakar, S.; Misra, S.; Kalia, J.; Pradeepkumar, P. I.; Srivatsan, S. G. Structural elucidation of HIV-1 G-quadruplexes in a cellular environment and their ligand binding using responsive ^{19}F-labeled nucleoside probes. *Chem. Sci.*, **2024**,*15*, 7982–7991.
2. Rout, B.; **Roy, S**.; Srivatsan, S. G. 5-Fluoro-2′-deoxyuridine as an efficient ^{19}F NMR reporter for G-quadruplex and i-motif structures.
3. Khatik, S. Y.; **Roy, S**.; Srivatsan, S. G. Synthesis and enzymatic incorporation of a Dual-App Nucleotide Probe That Reports Antibiotics-Induced Conformational Change in the Bacterial Ribosomal Decoding Site RNA. *ACS Chem. Biol.* **2024**, *19*, 3, 687–695.
4. Pandey, A.; **Roy, S**.; Srivatsan, S. G. Probing the Competition between Duplex, G-Quadruplex and i-Motif Structures of the Oncogenic c-Myc DNA Promoter Region. *Chem Asian J*. **2023**, *18*, e202300510.
5. **Roy, S.;** Srivatsan, S. G. Probing viral G-quadruplexes of SARS-CoV-2 with ^{19}F dual-app nucleoside analog. (*Manuscript under preparation*)
6. **Roy. S**.; Srivatsan, S. G. Probing nucleic acid structures using ^{19}F NMR. (Review article) (*Manuscript under preparation*)

Copyrights and permission

Structural elucidation of HIV-1 G-quadruplexes in a cellular environment and their ligand binding using responsive ^{19}F-labeled nucleoside probes

S. Roy, P. Majee, S. Sudhakar, S. Mishra, J. Kalia, P. I. Pradeepkumar and S. G. Srivatsan, *Chem. Sci.*, 2024, **15**, 7982 **DOI:** 10.1039/D4SC01755B

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2024, **15**, 7982

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 15th March 2024
Accepted 23rd April 2024

DOI: 10.1039/d4sc01755b

rsc.li/chemical-science

Structural elucidation of HIV-1 G-quadruplexes in a cellular environment and their ligand binding using responsive ^{19}F-labeled nucleoside probes†

Sarupa Roy,[a] Priyasha Majee,[b] Sruthi Sudhakar,[b] Satyajit Mishra,[c] Jeet Kalia,[cd] P. I. Pradeepkumar *[b] and Seergazhi G. Srivatsan *[a]

Understanding the structure and recognition of highly conserved regulatory segments of the integrated viral DNA genome that forms unique topologies can greatly aid in devising novel therapeutic strategies to counter chronic infections. In this study, we configured a probe system using highly environment-sensitive nucleoside analogs, 5-fluoro-2′-deoxyuridine (FdU) and 5-fluorobenzofuran-2′-deoxyuridine (FBFdU), to investigate the structural polymorphism of HIV-1 long terminal repeat (LTR) G-quadruplexes (GQs) by fluorescence and ^{19}F NMR. FdU and FBFdU, serving as hairpin and GQ sensors, produced distinct spectral signatures for different GQ topologies adopted by LTR G-rich oligonucleotides. Importantly, systematic ^{19}F NMR analysis in *Xenopus laevis* oocytes gave unprecedented information on the structure adopted by the LTR G-rich region in the cellular environment. The results indicate that it forms a unique GQ-hairpin hybrid architecture, a potent hotspot for selective targeting. Furthermore, structural models generated using MD simulations provided insights on how the probe system senses different GQs. Using the responsiveness of the probes and *Taq* DNA polymerase stop assay, we monitored GQ- and hairpin-specific ligand interactions and their synergistic inhibitory effect on the replication process. Our findings suggest that targeting GQ and hairpin motifs simultaneously using bimodal ligands could be a new strategy to selectively block the viral replication.

Introduction

HIV-1 is one of the most lethal retroviruses, which induces a chronic infection by etching the host cell genome with a proviral DNA that is reverse transcribed from its RNA genome. Established treatments use a cocktail of drugs having different modes of action to control the disease progression.[1] However, due to persistence of latent reservoirs, drug-resistance and promiscuity of the viral polymerase, it is very hard to eradicate the virus completely from the host system.[2,3] One of the current ways to counter latency involves awakening dormant viruses and simultaneously inhibiting viral replication with antiviral agents.[4,5] It is also hypothesized that targeting certain structural and functional segments of the integrated viral DNA genome could complement the above strategy and help in curing the disease.[6] An important and a highly conserved gene segment that could be suitable for this purpose is the long terminal repeat (LTR) of the HIV-1 promoter region.[7,8]

The initiation of HIV-1 transcription is navigated by the promoter region 5′-LTR, which is composed of U3, R and U5 regions.[9] The U3 region consists of three functional segments including the highly conserved core-binding site of NF-κB and Sp1 transcription factors,[10] which harbors contiguous G-rich tracts capable of forming G-quadruplex (GQ) structures namely LTR-II, LTR-III and LTR-IV, and HIVpro1 and HIVpro2 (Fig. 1A).[11–13] Notably, LTR-III and LTR-IV form GQs *in vitro* in a mutually exclusive manner. While LTR-III adopts a unique architecture made of a hybrid-type GQ juxtaposed with a three G-C paired hairpin motif, LTR-IV attains a parallel GQ topology with a T-bulge (Fig. 1B).[14,15] However, the entire G-rich region majorly forms the GQ-hairpin form like the LTR-III motif, and the parallel form of LTR-IV is induced when it binds to ligands or protein factors.[14–16] Importantly, the LTR GQ region represents an evolutionary conserved element across all primate lentiviruses, and the balance between different GQ structures is implicated in the propagation and latency of the virus.[17,18]

[a]*Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune, Dr Homi Bhabha Road, Pune 411008, India. E-mail: srivatsan@iiserpune.ac.in*

[b]*Department of Chemistry, Indian Institute of Technology Bombay, Mumbai 400076, India. E-mail: pradeep@chem.iitb.ac.in*

[c]*Department of Biological Sciences, Indian Institute of Science Education and Research (IISER) Bhopal, Bhopal Bypass Road, Bhauri, Bhopal 462066, India*

[d]*Department of Chemistry, Indian Institute of Science Education and Research (IISER) Bhopal, Bhopal Bypass Road, Bhauri, Bhopal 462066, India*

† Electronic supplementary information (ESI) available: Experimental details, CD, T_m, fluorescence, MD simulation, mass, NMR spectra and gel images. Experimental procedure for NMR measurements of ONs in a cellular environment, and ligand interaction with LTR GQs and their effects on the *Taq* DNA polymerase activity are described. See DOI: https://doi.org/10.1039/d4sc01755b

ACS chemical biology

pubs.acs.org/acschemicalbiology

Articles

Synthesis and Enzymatic Incorporation of a Dual-App Nucleotide Probe That Reports Antibiotics-Induced Conformational Change in the Bacterial Ribosomal Decoding Site RNA

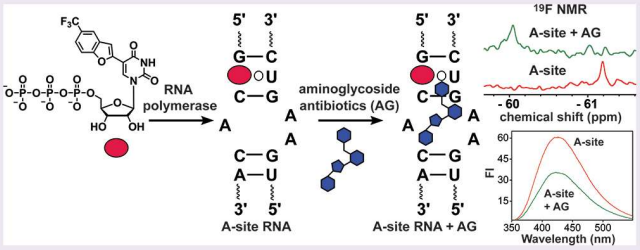
Saddam Y. Khatik, Sarupa Roy, and Seergazhi G. Srivatsan*

Cite This: ACS Chem. Biol. 2024, 19, 687−695

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Natural nucleosides are nonfluorescent and do not have intrinsic labels that can be readily utilized for analyzing nucleic acid structure and recognition. In this regard, researchers typically use the so-called "one-label, one-technique" approach to study nucleic acids. However, we envisioned that a responsive dual-app nucleoside system that harnesses the power of two complementing biophysical techniques namely, fluorescence and ^{19}F NMR, will allow the investigation of nucleic acid conformations more comprehensively than before. We recently introduced a nucleoside analogue by tagging trifluoromethyl-benzofuran at the C5 position of 2′-deoxyuridine, which serves as an excellent fluorescent and ^{19}F NMR probe to study G-quadruplex and i-motif structures. Taking forward, here, we report the development of a ribonucleotide version of the dual-app probe to monitor antibiotics-induced conformational changes in RNA. The ribonucleotide analog is derived by conjugating trifluoromethyl-benzofuran at the C5 position of uridine (TFBF-UTP). The analog is efficiently incorporated by T7 RNA polymerase to produce functionalized RNA transcripts. Detailed photophysical and ^{19}F NMR of the nucleoside and nucleotide incorporated into RNA oligonucleotides revealed that the analog is structurally minimally invasive and can be used for probing RNA conformations by fluorescence and ^{19}F NMR techniques. Using the probe, we monitored and estimated aminoglycoside antibiotics binding to the bacterial ribosomal decoding site RNA (A-site, a very important RNA target). While 2-aminopurine, a famous fluorescent nucleic acid probe, fails to detect structurally similar aminoglycoside antibiotics binding to the A-site, our probe reports the binding of different aminoglycosides to the A-site. Taken together, our results demonstrate that TFBF-UTP is a very useful addition to the nucleic acid analysis toolbox and could be used to devise discovery platforms to identify new RNA binders of therapeutic potential.



INTRODUCTION

RNA takes part in several essential cellular processes ranging from protein synthesis to catalysis and gene regulation and editing. In order to perform its function, RNA adopts complex secondary and tertiary structures, which are influenced by posttranscriptional modifications, changes in environmental conditions, and interaction with proteins, metabolites, and external ligands.[1−3] Chemical tools and biophysical techniques based on fluorescence, NMR, and X-ray crystallography are widely employed to study the structure, dynamics, and recognition properties of RNA.[4−6] While fluorescence spectroscopy aids in the rapid analysis of RNA in solution and cells,[5] NMR[7] and X-ray[8] crystallography techniques provide information at the atomic level. These methods require fluorophore-labeled, isotope-enriched, and heavy atom-labeled RNA molecules as basic components of nucleic acids are nonemissive and lack intrinsic labels suitable for such analyses.[9−11] In this context, environment-sensitive nucleoside analogs inserted into RNA oligonucleotides (ONs) serve as useful tools to analyze their structure and recognition.[12−20] More recently, ^{19}F NMR has emerged as a valuable biophysical handle to study nucleic acid conformations as the ^{19}F atom is 100% naturally abundant and its chemical shift is highly environment-sensitive.[21,22] Chemically synthesized ^{19}F-labeled ONs have been employed in studying RNA structures and their interaction with proteins and ligands.[23−26]

Further, the ability of fluorine-labeled probes to exhibit unique spectral signatures for different structures and the absence of fluorine in biological systems (no background signal) has aided the determination of noncanonical structures adopted by certain ON sequences in the cellular milieu.[23,27−29]

The majority of tools that have been employed to investigate nucleic acids use a type of label that is compatible with only a given biophysical technique. We envisioned developing an

Received: November 7, 2023
Revised: February 6, 2024
Accepted: February 13, 2024
Published: February 26, 2024

https://doi.org/10.1021/acschembio.3c00676
ACS Chem. Biol. 2024, 19, 687−695

Bioorg. Med. Chem. Lett. 117 (2025) 130060

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

5-Fluoro-2′-deoxyuridine as an efficient ^{19}F NMR reporter for G-quadruplex and i-motif structures

Bhakti P. Rout [a], Sarupa Roy [a], Seergazhi G. Srivatsan [*]

Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune, Dr. Homi Bhabha Road, Pune 411008, India

ARTICLE INFO

Keywords:
Nucleoside probe
G-quadruplex
I-motif
Telomere
Environment-sensitive probe
^{19}F NMR

ABSTRACT

DNA sequences that are composed of multiple G- and C-tracts can potentially form non-canonical structures called G-quadruplex (GQ) or i-motif (iM), respectively. Such sequences are found at the ends of chromosomes (telomeric repeats) and in the promoter region of several genes that cause cancer. Despite extensive studies, distinguishing different GQ and iM topologies is not easy. In this work, we have used one of the conservatively modified nucleoside analogs, namely 5-fluoro-2′-deoxyuridine (FdU) to study different GQ and iM structures of the human telomeric (H-Telo) DNA repeat sequence using ^{19}F NMR technique. The probe is minimally perturbing and distinguishes different GQ topologies by providing unique ^{19}F signatures. Our findings suggest that the telomeric repeat assumes hybrid-type GQ structures in intracellular ionic conditions as opposed to a parallel form predicted by using synthetic cellular crowding mimics. Further, with the incorporation of the probe into a C-rich H-Telo DNA ON, we were able to study the transition from iM structure to a random coil structure. Taken together, FdU is a promising probe, which could be used to determine the structure of non-canonical nucleic acid motifs *in vitro* and potentially in the native cellular environment.

Non-canonical nucleic acid structures, namely G-quadruplexes (GQs) and i-motifs (iMs) are important motifs due to their conserved location in the genome[1,2] and their functional role in regulating gene expression.[3–7] GQs are formed when two or more G-quartets, assembled *via* Hoogsteen base pairing, stack with each other in the presence of cations like K^+ and Na^+. Initially, it was thought that this class of structures is a mere structural aberration found only in *in-vitro* conditions but extensive investigation in the last decade clearly indicated their presence in the cellular environment.[8,9] Notably, computational predictions backed up by experimental evidence indicate the presence of GQs in important regions of the genome such as telomeres and DNA promoter regions and untranslated regions of mRNA, which are implicated in cancer progression.[10,11] Dysfunction of these regulatory elements due to mutations, and compromised replication of these regions profoundly impact cellular processes.[12] The complementary C-rich strand can adopt iM structures *via* intercalation of hemi-protonated C^+–C base pairs.[13] Despite the fact that C-rich and G-rich strands co-exist, iMs have received less attention compared to GQs. This is because GQs form under physiological conditions but iM are generally formed at acidic pH. However, recent studies suggest that negative super helicity, molecular crowding, length of the C-tract and loop composition can potentially induce the formation of iMs under physiological conditions.[14] In support of these studies, iM formation in cells was reported using an antibody fragment (iMab).[15] Similar to GQs, iMs have been reported to take part in the regulation of gene expression.[16] Consequently, many GQ and iM binders have been developed that efficiently affect the replication and transcription process of disease-causing genes, and hence, targeting these non-canonical structures is envisioned as an alternative strategy to treat aging-related diseases and cancers.[11a,17,18] However, a very few ligands have progressed to clinical trials because of the structural polymorphism exhibited by GQ- and iM-forming sequences.

In order to advance therapeutic strategies, several chemical probes and tools have been developed to map various GQ and iM structures.[19,20] Typical biophysical methods that are used to investigate these structures are CD, fluorescence, NMR, and X-ray crystallography.[20–23] These methods provide valuable structural information when one species is formed, but are not effective when multiples topologies are formed by a sequence. Single-molecule analysis techniques using force-based (optical and magnetic tweezers) and fluorescence are somewhat

* Corresponding author.
E-mail address: srivatsan@iiserpune.ac.in (S.G. Srivatsan).
[a] These authors contributed equally.

https://doi.org/10.1016/j.bmcl.2024.130060
Received 30 September 2024; Received in revised form 10 November 2024; Accepted 1 December 2024
Available online 3 December 2024

ACES Asian Chemical Editorial Society

doi.org/10.1002/asia.202300510

CHEMISTRY
AN ASIAN JOURNAL
Research Article

www.chemasianj.org

Special Collection

Probing the Competition between Duplex, G-Quadruplex and i-Motif Structures of the Oncogenic c-Myc DNA Promoter Region

Akanksha Pandey,[a] Sarupa Roy,[a] and Seergazhi G. Srivatsan*[a]

Development of probe systems that provide unique spectral signatures for duplex, G-quadruplex (GQ) and i-motif (iM) structures is very important to understand the relative propensity of a G-rich-C-rich promoter region to form these structures. Here, we devise a platform using a combination of two environment-sensitive nucleoside analogs namely, 5-fluorobenzofuran-modified 2′-deoxyuridine (FBF-dU) and 5-fluoro-2′-deoxyuridine (F-dU) to study the structures adopted by a promoter region of the c-Myc oncogene. FBF-dU serves as a dual-purpose probe containing a fluorescent and ^{19}F NMR label. When incorporated into the C-rich sequence, it reports the formation of different iMs *via* changes in its fluorescence properties and ^{19}F signal. F-dU incorporated into the G-rich ON reports the formation of a GQ structure whose ^{19}F signal is clearly different from the signals obtained for iMs. Rewardingly, the labeled ONs when mixed with respective complementary strands allows us to determine the relative population of different structures formed by the c-Myc promoter by the virtue of the probe's ability to produce distinct and resolved ^{19}F signatures for different structures. Our results indicate that at physiological pH and temperature the c-Myc promoter forms duplex, random coil and GQ structures, and does not form an iM. Whereas at acidic pH, the mixture largely forms iM and GQ structures. Taken together, our system will complement existing tools and provide unprecedented insights on the population equilibrium and dynamics of nucleic acid structures under different conditions.

Introduction

Apart from double helical structure of DNA, G-quadruplex (GQ) and i-motif (iM) structures formed by G-rich and C-rich sequences present in the genome serve as regulatory elements in cellular processes.[1–4] While, GQs are assembled by stacking of tetrads formed by four guanine bases through Hoogsteen base pairing, iMs are formed by intercalation of two hemiprotonated cytosine$^+$-cytosine base-paired strands.[5–7] Sequences capable of forming these structures are widely found in telomeric overhang and promoter regions including that of many genes that cause cancer and neurodegenerative diseases.[8–10] In terms of structure, both G-rich and C-rich sequences form multiple topologies of GQ and iM *in vitro*, which depend on the sequence composition and environmental conditions.[11–16] However, their existence in the cellular milieu was highly debated until recently. Antibodies BG4 and D1, which specifically bind to GQ topologies provided compelling evidence for the formation of GQs in cells.[17,18] Further, in-cell NMR and fluorescence studies also supported these findings.[19–22] Several biochemical investigations directed towards understanding the functional role of GQs not only indicated that these motifs play important roles in the regulation of gene expression but also validated their importance as a potential target for drug discovery.[23] Further, a notable number of small molecule ligands that interact and stabilize GQs suppress the protein expression of oncogenes that harbor such motifs in the promoter region.[24] In contrast, studies on iMs are limited as these structures mostly form at acidic pH.[25] However, recent findings suggest that negative superhelicity, molecular crowding and certain modifications could support the formation of iMs *in vitro* at near physiological conditions.[25,26] Interestingly, transfection followed by in-cell proton NMR analysis of preformed iMs indicated the stabilization of such structures in live cells.[27,28] The confirmatory evidence for the existence of iMs formed by endogenous nucleic acids was obtained using a structure-specific antibody called iMab.[29] Moreover, the ability of small molecule ligands and proteins to bind C-rich domains and regulate the expression of certain oncogenes suggests that iM could complement GQ as a therapeutic target, particularly in promoter regions where G-rich and C-rich segments coexist.[30,31]

Circular dichroism (CD), thermal melting, X-ray crystallography, fluorescence and NMR are typically employed to interrogate the structure, dynamics and binding affinity of small molecules to GQs and iMs.[11,32–35] In most of these studies, single-stranded G-rich and C-rich oligonucleotides (ONs) are used to evaluate the formation and folding dynamics of individual structures. However, in order to conceive a strategy that would enable the targeting of both these structures, it is

[a] *A. Pandey, S. Roy, S. G. Srivatsan*
Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune
Dr. Homi Bhabha Road, Pune 411008 (India)
E-mail: srivatsan@iiserpune.ac.in

Supporting information for this article is available on the WWW under https://doi.org/10.1002/asia.202300510

This manuscript is part of an Indo-German Workshop (IGW) special collection.