A nucleoside probe containing a ^{19}F label serves as an efficient NMR probe to detect different G-quadruplex and i-motif topologies

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfillment of the requirements for the BS-MS Dual Degree Programme

by

Bhakti Prasad Rout

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

April, 2020

Supervisor: Prof. S.G. Srivatsan

Certificate

This is to certify that this dissertation entitled '**A nucleoside probe containing a ^{19}F label serves as an efficient NMR probe to detect different G-quadruplex and i-motif topologies'** towards the partial fulfillment of the BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Bhakti Prasad Rout at Indian Institute of Science Education and Research under the supervision of Dr. S. G. Srivatsan, Professor, Department of Chemistry, during the academic year 2019-2020.

Prof. S. G. Srivatsan

Committee:

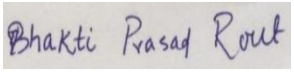
Dr. S. G. Srivatsan

Dr. Raghavendra Kikkeri

This thesis is dedicated to all my teachers.

Declaration

I hereby declare that the matter embodied in the report entitled '**A nucleoside probe containing a ^{19}F label serves as an efficient NMR probe to detect different G-quadruplex and i-motif topologies'** are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. S. G. Srivatsan and the same has not been submitted elsewhere for any other degree.



Bhakti Prasad Rout

Date: 29th March 2020.

Acknowledgment

It gives me immense pleasure to express my heartiest gratitude towards all those without whom this work of mine would not have been possible.

First of all, I am sincerely thankful to my guide prof. Dr. S. G. Srivatsan, whose continuous encouragement, motivation & guidance throughout my research year made me accomplish my work. I would like to thank my official lab mentor, Akanksha and my unofficial lab mentor, Saddam for nurturing me with different lab procedures and contributing their valuable experience from time to time proved to be plateful in the completion of this work. I also want to express my sincere obligations towards all other lab members Manisha, Jerrin, Pulak, Sarupa, Indu and Swagata for creating a jovial atmosphere that helped in carrying out a full year of rigorous research work.

I would also like to thank Rupam from Dr. Jayant Udgaonkar's lab whose help was priceless in achieving the mass spectrums of the synthesized Oligonucleotides.

I would also like to oblige Sandeep, Nitin, and Chinmay, the NMR experts in the NMR research center for rendering inestimable help in carrying out the ^{19}F and ^{1}H experiments. Sandeep and Nitin also made me competent in using the NMR machine in carrying out 600MHz experiments independently.

I am indebted to Mrutyunjay for all his help during the project year.

I would like to thank my baba and mama for having their blissful hands upon my head throughout my life & my elder brother for his unceasing encouragement and motivation. And last but not least my friends, heartthrobs are the ones who made me believe in me & my knowledge during times of distress all these years.

Contents:

List of Tables:

List of Figures:

- **Figure 9**: CD study of FiMHT in buffers of different pH
- **Figure 10**: Determination of ipH of FiMHT by using circular dichroism (CD) studies
- **Figure 11 & 12**: Thermal melting studies of H-Telo DNA ONs
- **Figure 13**: ^{19}F NMR signatures of H-Telo DNA ONs in different ionic conditions at 298 K
- **Figure 14**: ^{19}F NMR signatures of parallel conformation of H-Telo DNA GQs in intracellular buffer containing 40% PEG
- **Figure 15**: ^{19}F NMR of C-rich H-Telo DNA ON FiMHT in buffers of different pH

Annexure

- Mass spectra of purified H-Telo DNA ONs FGQH1, FGQH2, FGQ1 and FiMHT
- ^{1}H, ^{13}C NMR spectra of compound 1(DMT-FdU)
- ^{1}H, ^{31}P NMR spectra of DMT-FdU phosphoramidite substrate 2

Abbreviations:

- DNA: Deoxyribonucleic acid
- RNA: Ribonucleic acid
- GQ: G-quadruplex
- i-motif: intercalated motif
- ON: Oligonucleotide
- H-Telo: Human-Telomeric
- EPR: Electron paramagnetic resonance spectroscopy
- PEG: Polyethylene Glycol
- CD: Circular Dichroism
- FdU: 5-Fluoro-2'-deoxy uridine
- RP-HPLC: Reverse Phase High-Performance Liquid Chromatography
- IC buffer: Intracellular buffer
- NMR: Nuclear Magnetic Resonance
- mmol: millimole
- M: Molar
- mg: milligram
- mL: milliliter
- MHz: Megahertz
- HRMS: High Resolution Mass spectrometry
- DMSO: Dimethyl sulfoxide
- ACN: Acetonitrile
- TEAA buffer: Triethylammonium acetate buffer
- w.r.t: with respect to

Abstract

G-rich and C-rich DNA sequences capable of forming G-quadruplex (GQ) and i-motif structures are found in the telomeres and promoter region of several genes that cause cancer. Despite extensive studies *in vitro*, determining their topologies in the native cellular environment is not easy. In this work, we have used one of the conservatively modified nucleoside analogs, namely 5-fluoro-2'-deoxyuridine (FdU) to study the GQ and i-motif structures of the human telomeric (H-Telo) DNA repeat sequence. The probe was found to be minimally perturbing and could distinguish different GQ topologies by providing unique signatures. Previous reports using cellular crowding mimics have indicated the formation of parallel GQ conformation. However, we found out that the H-Telo DNA ONs did not assume parallel conformation in intracellular buffer conditions using our probe. Further, with the incorporation of the probe into a C-rich H-Telo DNA ON, we were able to study the transition from i-motif to random coil structure. Besides, we were also able to validate the ipH of the C-rich sequence using CD studies. Taken together, 5-fluoro-2'-deoxyuridine is a promising probe, which can be used to determine the structure of non-canonical nucleic acid motifs in the native cellular environment.

1. Introduction

DNA can form both canonical and non-canonical structures. Canonical structures include the different types of double-helical conformations such as A-form, B-form, and Z-form (Figure 1B). All three conformations are characterized by Watson-Crick base pairing where A is paired with T and G is paired with C through hydrogen bonding (Figure 1A). B-form is commonly found double-helical conformation in all the biological systems while A-form is found in dehydrated samples of DNA such as those in crystallographic experiments and the Z-form is found in an environment having very high salt concentrations. Non-canonical structures include hairpin, bulge, pseudoknot, tetraplexes such as G-quadruplexes and i-motifs, etc.[1-4] The tetraplexes such as GQ and i-motifs are characterized by Hoogstein base pairing (Figure 2A). The structural diversity found in nucleic acids is associated with a multitude of functions they carry out in the biological system. They perform a plethora of functions such as storage and transfer of genetic information from one generation to another generation, catalysis and regulation of gene expression, etc.[5] Non-canonical structures, namely G-quadruplexes and i-motifs (Figure 2B) have gained particular importance due to their conserved location in the genome[6,7] and their functional role in regulating gene expression including that of oncogenes.[8-12]

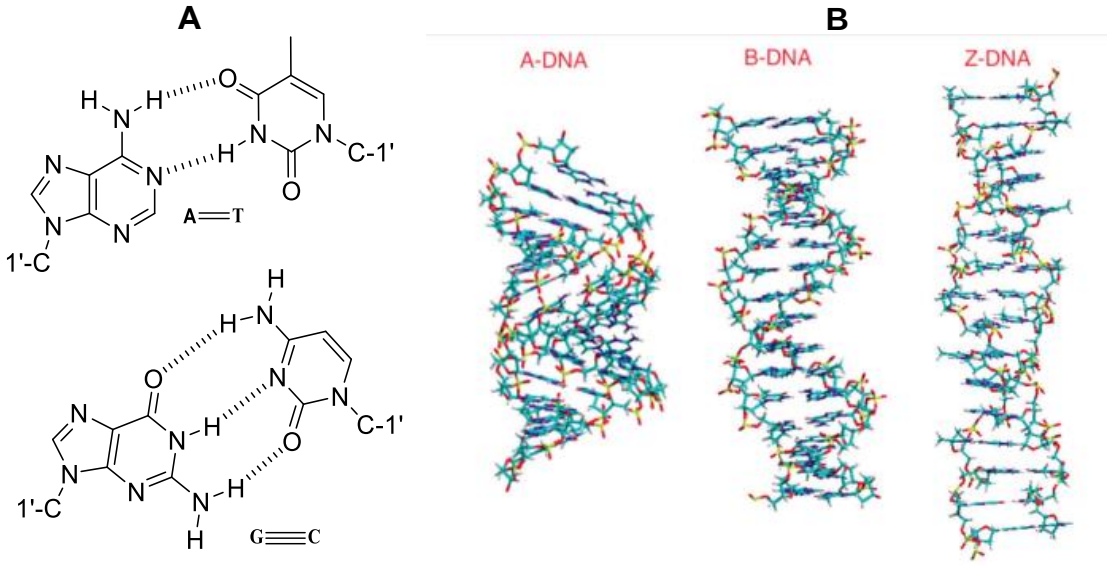


Figure 1: (**A**) A:T and G:C Watson-Crick base pairing. (**B**) Different forms of DNA such as A-form, B-form, and Z-form where the individual strands are connected via Watson-Crick base pairing.[13]

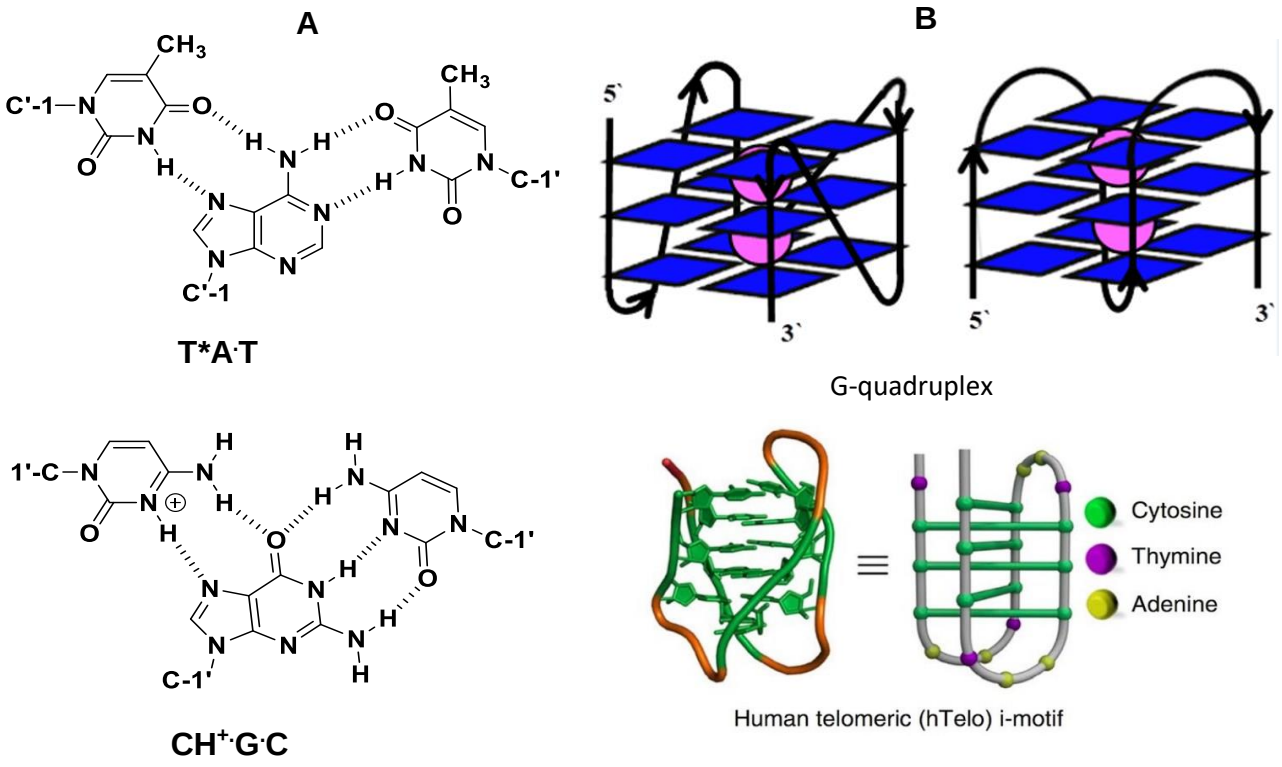


Figure 2: (**A**) Hoogstein base pairs. (**B**) H-Telo G-quadruplexes[14] and i-motif structures stabilized through Hoogstein base-pairing.[15]

1.1. G-quadruplex structure and function

G-quadruplexes (GQs) are formed when two or more G-quartets stack with each other in the presence of metal cations like K^+, Na^+, Li^+ or NH_4^+. The G-quartets are a group of four guanine nucleotides located on one plane interacting with each other through Hoogstein base pairing. Initially, it was thought that these are mere structural aberrations found only in *in vitro* conditions but extensive investigations in the last two decades have clearly indicated their presence in cellular conditions.[16,17] Computational predictions backed up by experimental evidence indicate the presence of GQs in the functional regions of genome like telomeres and promoter regions of DNA and untranslated regions of mRNA, which are known to cause cancer.[16,17] Studies have suggested that GQ-forming regions of the genome are found to be conserved across species especially vertebrates[6,7] and found to be playing an important role in regulating important cellular processes such as replication, transcription, translation, telomere maintenance, etc.[16,18,19] For example, a

mutation from G→A in the purine-rich strand complementary to the nuclease hypersensitivity element III (1) (present in the upstream region to the promoter region of proto-oncogene c-MYC) affects the rate of transcription significantly. Thus, small molecules, which can act as GQ-binders and affect the important regulatory processes such as transcription, translation, replication, etc. have been proposed as an alternative strategy to treat diseases such as aging-related diseases and cancer.[20,21] However, there are a very few clinically available small molecules binders because of a very fundamental challenge associated with structural diversity of GQs in cellular conditions. This structural diversity (Figure 3) is due to the diversity in sequences, ionic conditions and molecular crowding conditions found in cells.[22-26] Thus, there is a need to develop specific GQ binders that will have the ability to bind to only a particular GQ topology leaving others unaffected.[27]

Various biophysical techniques like circular dichroism, NMR, EPR, and fluorescence spectroscopy have been used to probe the structural diversity of GQs in different ionic conditions and molecular crowding conditions trying to mimic the native environment in the cell.[28-31] Human telomeric (H-Telo) DNA Oligonucleotide (ON) is a very well-studied system consisting of $(TTAGGG)_n$ repeats. It is known that H-Telo DNA adopts multiple conformations in K^+ solution as found out using 2-D NMR.[32] It adopts an anti-parallel conformation in NaCl, parallel conformation in $SrCl_2$ and in presence of molecular crowding agents such as PEG.[30]

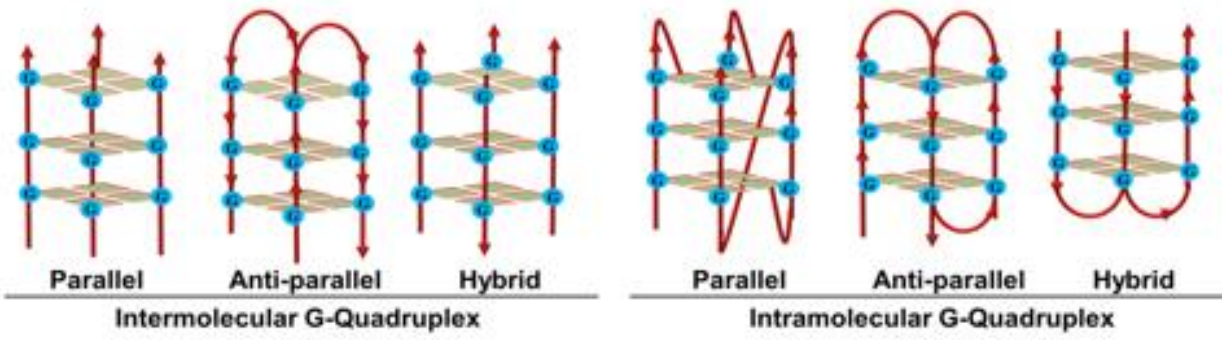


Figure 3: Structural diversity of G-quadruplexes based on the strand stoichiometry, strand directionality, and arrangement of connecting loops.[33]

1.2. i-Motif structure and function

i-Motifs are formed by C-rich strands with stacked intercalation of hemi-protonated C-neutral C base pairs (CH^+-C, Figure 4.[34] The i-motifs are found in the regions which are complementary to G-rich regions in the genome which are known to cause cancer.[3,20] Despite this fact, i-motifs have been given much less attention compared to GQs. This is because GQs form under physiological conditions but i-motifs are generally formed at acidic pH[34] and evidence for GQ formation in cellular conditions is available now while this is yet to be unambiguously established in the case of i-motifs.[35] However, some studies suggest that negative super helicity, molecular crowding, change in the loop composition can potentially induce the formation of i-motifs under physiological conditions.[36] Similar to GQs, i-motifs have been reported to take part in the regulation of gene expression. For example, a study suggests that the C-rich strand directly upstream of the BCL2 promoter element has the ability either to form a hairpin or an i-motif structure present in the form of an equilibrium mixture.[37] The study demonstrates how small molecules can shift the equilibrium and affect the transcription. The pH sensitivity of i-motifs has been applied in mapping pH changes in live cells, in designing nanomachines and as switches in logic operations.[38] All these results suggest their potential application as therapeutic targets similar to the GQs. However, there are not many viable small molecule binders that target i-motifs specifically.

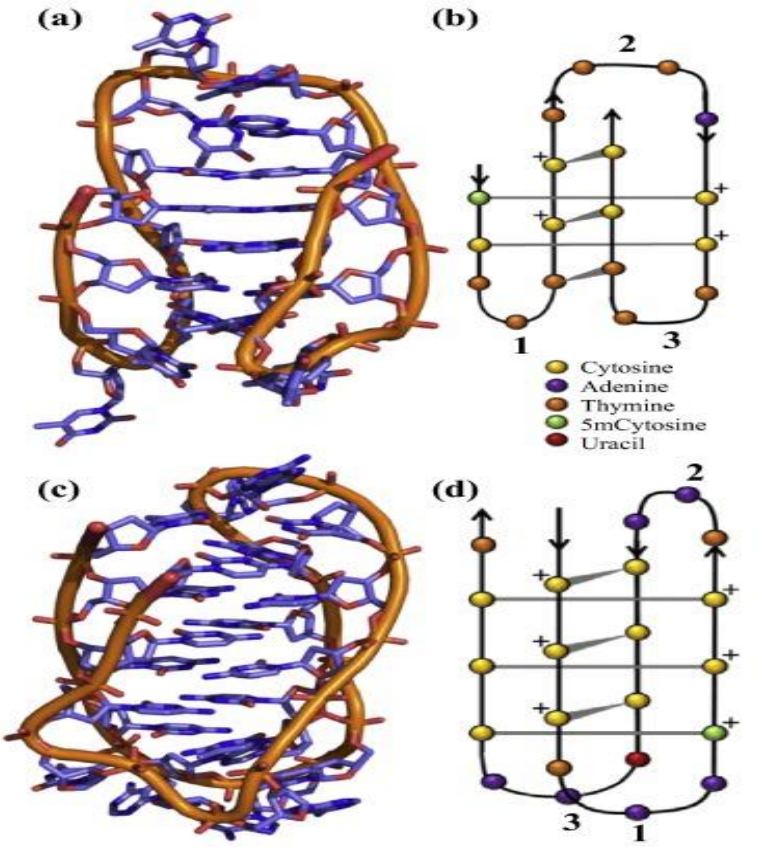


Figure 4: Structural diversity in i-motifs. NMR structure of i-motif-forming oligonucleotide sequences. The two i-motifs show differences in intercalation topologies, strand polarity and loop composition.[39] Figure adapted from Reference 39.

1.3. Challenges associated with the study of GQ and i-motif structures

Numerous biophysical methods, namely circular dichroism (CD), fluorescence and X-ray crystallography and NMR are used in probing the structure and small molecular binding properties of GQ and i-motifs.[40] Methods like fluorescence, NMR and X-ray use ONs labeled with an appropriate probe or probes because natural nucleosides do not contain intrinsic labels that are compatible with the above-said techniques. So, several chemical probes have been developed to analyze these non-canonical structures. However, the majority of probes are not compatible with both *in vitro* and in-cell analysis, and most importantly, the probes poorly distinguish different topologies of GQ and i-motif structures. Therefore, development of chemical probes that can differentiate different nucleic acid structures, namely duplex, GQ and i-motif *in vitro* and can also be used in cell-based assays is highly desired.

Recently, Srivatsan and co-workers introduced a novel dual-app nucleoside probe[41] (5-fluorobenzofuran 2'-deoxyuridine), containing a fluorophore and ^{19}F NMR label. The probe is an excellent GQ sensor and enabled the determination of the structure of H-Telo DNA overhang in live frog oocytes by ^{19}F NMR. This report elucidated the power of ^{19}F NMR in the structural analysis of nucleic acids. We next wanted to harness the power of ^{19}F isotope in probing the relative population of GQ/i-motif/duplex forms adopted by H-Telo DNA ONs in *in vitro* as well as in live cells such as mammalian cells and frog oocytes. For this, it is important to use a probe that will provide unique signatures for GQ, i-motif and duplex forms. In this regard, we decided to use one of the conservatively modified nucleoside analogs, namely 5-fluoro-2'-deoxy uridine (FdU) to study different forms of H-Telo G-rich and C-rich ON sequences.

Herein, we report the synthesis and incorporation of 5-fluoro-2'-deoxyuridine and its phosphoramidite substrate into H-Telo G-rich and C-rich ONs. The nucleoside analog is practically non-perturbing and distinguishes different GQ topologies by providing unique signatures. Further, this probe incorporated into C-rich H-Telo sequence clearly reported the transformation of a random unfolded structure to i-motif structure as a function of pH. Taken together, 5-fluoro-2'-deoxyuridine is a promising probe to determine the structure of nucleic acids, in this case, non-canonical nucleic acid motifs in native cellular environment.

2. Experimental Section

2.1. Materials

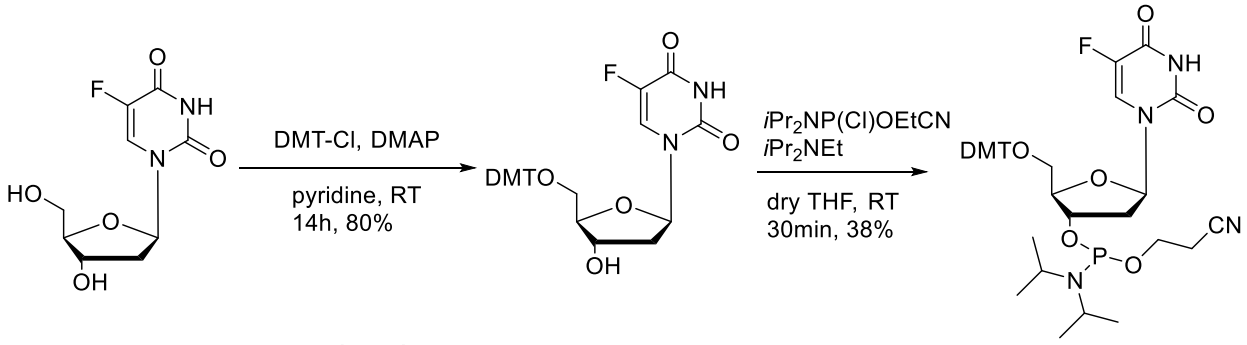
5-Fluoro-2'-deoxyuridine was obtained from Biosynth Carbosynth. 4,4'-dimethoxytrityl chloride, pyridine, *N,N*-dimethyl aminopyridine (DMAP), *N,N*-diisopropylethylamine (DIPEA), PEG 200 and all reagents used for buffer preparation (Bio-ultra grade) were obtained from Sigma Aldrich. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite was obtained from Alfa Aesar. H-Telo DNA ONs CONGQ1, CONGQH1 and CONGQH2 were obtained from Integrated DNA Technology (IDT) or Eurofins Genomics India Pvt. Ltd. Phosphoramidites of natural bases for solid-phase synthesis were obtained from ChemGenes corporation or from Sigma Aldrich. Solid supports for DNA solid-phase synthesis were obtained from Glenn Research. Other solid-phase reagents were ordered

either from ChemGenes corporation or Sigma Aldrich. DNA ONs were purified by PAGE (20% gel) using urea as the denaturant and desalting was carried out using a Sep-Pak Classic C18 cartridge (Waters Corporation). Autoclaved water was used for all biophysical analysis.

2.2. Instruments

DNA ONs synthesis was performed using ABI applied Biosystems 392 DNA/RNA synthesizer. High-Performance Liquid Chromatography (HPLC) was carried out using Agilent Technologies 1260 Infinity HPLC. Absorbance measurements were carried out using Shimadzu UV-2600 spectrophotometer. UV thermal melting analysis was carried out with the help of Cary 300BIO UV-Vis Spectrophotometer. CD analysis was done using JASCO J-815 CD spectrometer. NMR of small molecules was obtained using Bruker AVANCE III HD ASCEND 400 MHz spectrometer and processed using Mnova form Mestrelab research. Similarly, NMR of ONs in different salt conditions was obtained using Bruker AVANCE III HD ASCEND 600 MHz spectrometer using QCIF cryoprobe and processed using Bruker Topspin software.

2.3. Synthesis of 5'-*O*-DMT-protected 5-fluoro-2'-deoxyuridine 1(DMT-FdU) and DMT-FdU-phosphoramidite 2



Scheme 1. Synthesis of DMT-FdU-phosphoramidite substrate **2**.

2.3.1. Synthesis of 5'-*O*-DMT-protected FdU 1 (DMT-FdU)

FdU (350 mg, 1.4 mmol, 1.0 eq), 4,4'-dimethoxytrityl chloride (DMT-Cl) (723 mg, 2.1 mmol, 1.5 eq) and 4-(dimethylamino)pyridine (DMAP) (87 mg, 0.7 mmol, 0.5 eq) were

taken in a 50 mL round-bottomed flask and dried under vacuum for 1 h. 12 mL of dry pyridine was added to the above reaction mixture following which it was kept for stirring at room temperature (~25 °C) for 14 h. The reaction mixture was extracted with 100 mL (×2) of ethyl acetate and 80 mL (×2) of sodium bicarbonate solution. The extracted organic layer was passed through sodium sulfate to remove traces of water. Silica gel column chromatography was conducted to purify the compound **1**. Yield 80% (780 mg). ^{1}H NMR (400 MHz, $CDCl_3$): δ(ppm) 7.78 (d, 1H, J = 6.8 Hz), 7.41-7.39 (m, 2H), 7.32-7.20 (m, 7H), 6.85-6.82 (dd, 4H (meta in 4-(OMe)C_6H_4)), 6.33-6.29 (m, 1H, J_1 = 6.5Hz, J_2 = 5.9 Hz), 4.54-4.51 (m, 1H, J_1 = 3.2 Hz, J_2 = 4.0 Hz, J_3 = 3.2 Hz), 4.08-4.05 (q, 1H, J_1 = J_2 = J_3 = 3.2 Hz), 3.78 (s, 6H), 3.66-3.44 (m, 2H), 2.50-2.44 (m, 2H); ^{13}C NMR (400 MHz, $CDCl_3$): δ(ppm) 158.5, 144.0, 135.1, 129.8, 127.8, 126.9, 124.1, 123.8, 113.2, 87.0, 86.2, 85.3, 76.8, 76.5, 71.7, 63.1, 55.1, 52.7, 45.3, 40.1; HRMS: m/z Calculated for $C_{30}H_{29}FN_2NaO_7$ $[M+Na]^+$ = 571.1900, found: 571.1897.

2.4. Synthesis of DMT-FdU-phosphoramidite substrate 2

200 mg (0.4 mmol, 1.0 eq) of compound **1** was taken in a 10 mL round-bottom flask and kept in vacuum for 30 minutes. It was dissolved in 4 mL of dry tetrahydrofuran (THF) under N_2 atmosphere. *N,N*-Diisopropylethylamine (DIPEA, 0.19 mL, 1.1 mmol, 3.0 eq) was added and the solution was stirred under N_2 atm at 0 °C. After 10 minutes of stirring at 0 °C, 0.163 mL (0.7 mmol, 2.0 eq) of 2-cyanoethyl *N,N*-diisopropylaminochloro phosphoramidite was added using a gas-tight syringe and stirred at 0 °C. The reaction was seen to be complete within half an hour. A work up was carried out using 40 mL ethyl acetate and 40 mL (×2) sodium bicarbonate solution. It was followed by a brine wash and finally passed through sodium sulfate to remove traces of water. A silica gel column chromatography was performed to elute a mixture of two diastereomers at 60% ethyl acetate:petether (containing 0.4% Triethylamine) solution. Yield 38% (104 mg). ^{1}H NMR (400 MHz, $CDCl_3$): δ(ppm) 7.88 (d, 1H, J = 6.1 Hz), 7.82 (d, 0.65H, J = 6.1 Hz), 7.42-7.39 (m, 3.3H), 7.32-7.22 (m, 11.7H), 6.86-6.81 (m, 6.62H), 6.35-6.32 (m,1.65H), 4.68-4.66 (m, 1.65H), 4.21-4.05 (m, 1.65H), 3.79 (s, 9.93H), 3.76 (m, 6.73H), 3.42-3.38 (m, 3.3H), 2.65-2.60 (m, 1.65H), 2.64-2.46 (m, 3.44H), 2.33-2.26 (m, 1.72H), 1.20-1.10 (m, 19.8H);

^{31}P NMR (400 MHz, $CDCl_3$): δ(ppm) 149.64, 149.38; HRMS: m/z Calculated for $C_{39}H_{47}FN_4O_8P$ $[M+H]^+$ = 749.3100, found: 749.3072.

2.5. ^{19}F NMR of FdU different solvents

150 μM samples of FdU in different solvents containing 15% DMSO-d_6 were prepared and 19 NMR was recorded using trifluorotoluene (TFT) -63.72 ppm as reference.

2.6. Synthesis of H-Telo G-rich and C-rich DNA ONs

The following four H-Telo DNA ON (3 G-rich ONs and 1 C-rich ON) sequences were synthesized with the modified nucleoside phosphoramidite incorporated at the positions shown, namely compound **2**:

FGQ1: (5' AGG GTT AGG G**2**T AGG GTT AGG G 3')

FGQH1: (5' TTG GGT TAG GG**2** TAG GGT TAG GGA 3')

FGQH2: (5' TAG GGT TAG GG**2** TAG GGT TAG GGT T 3')

FiMHT: (5' CCC TAA CCC **2**AA CCC TAA CCC TAA 3')

The ONs with the incorporated modification were synthesized (1 μmole scale, 1000 Angstrom CPG solid support) following the standard procedure for solid-phase synthesis. After solid-phase synthesis, synthesized ONs were deprotected from their solid supports and other protecting groups by treating with 28-30% Ammonium Hydroxide solution for 16 h at 55 °C and 30 minutes at 70 °C. The solution was then dried using speed vac and the residue was then purified by denaturing Polyacrylamide gel electrophoresis (20% gel). The gel band w.r.t. the ON was identified using UV shadowing, following which it was cut and crushed. The ON was extracted using 4 mL ammonium acetate solution (0.05 M) for 12 h and finally desalted using Sep-Pak classic C18 cartridges (Waters) before being taken for quantification. Diluted samples of the ONs were prepared and the absorbance was at 260 nm was recorded to calculate the amount of synthesized ONs. The purity of the ONs was then checked using RP-HPLC (Figure 6) and identity confirmed by mass spectrometry (Table 1 & Figure 16 in Annexure).

HPLC conditions: Mobile phase A = 50 mM TEAA buffer (pH 7.5), mobile phase B = ACN. Flow rate = 1 mL/min. Gradient = 0−100% B in 30 min. HPLC analysis was carried out using a Luna C18 column (250 x 4.6 mm, 5 micron).

2.7. CD studies

Modified and unmodified GQ-forming ONs (8 μM) were heated (90 °C for 5 minutes) in sodium phosphate buffer (10 mM, pH 7.0) comprising 100 mM of NaCl or in potassium phosphate buffer (10 mM, pH 7.0) comprising 100 mM of KCl or in Tris-HCl buffer (50 mM, pH 7.0) comprising 150 mM of $SrCl_2$ or in Intracellular (IC) buffer (25 mM potassium phosphate buffer, 85 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 130 nM $CaCl_2$, pH 7.4) or in IC buffer containing 40% of PEG (pH 7.4). Similarly, samples of modified and unmodified i-motif ONs were formed in sodium acetate buffer (30 mM) of different pH values such as 5.0, 5.5, 5.8 or sodium phosphate buffer (30 mM) with pH values 6.0, 6.1, 6.3, 6.6, 7.0, & 7.4 all containing 100 mM NaCl. The samples were cooled and CD was recorded in at least duplicate. Similarly, duplex samples were formed by annealing FdU-modified ONs (5.0 μM) with corresponding complementary ONs (5.5 μM) in the above different buffer solutions.

2.8. T_m determination

Modified and control ONs (1 μM) were annealed similarly as in CD experiments. The measurements were then carried out using Cary 300 Bio UV-Vis spectrophotometer. The temperature was raised at a rate of 1 °C/min from 20°C to 95 °C and then cooled at the same rate until it reached 20 °C. The absorbance was measured at 295 nm for G-quadruplex samples and at 260 nm for duplex and i-motif samples.

2.9. ^{19}F and ^{1}H NMR of modified GQ- and i-motif-forming ONs

NMR samples with modified and unmodified ONs were annealed similarly as in CD experiments in different ionic conditions as mentioned below:

- FGQ1 (50 μM) in buffer (pH 7.0) containing 100 mM NaCl
- FGQ1 (50 μM) in buffer (pH 7.0) containing 100 mM KCl
- FGQ1 (50 μM) in buffer (pH 7.0) containing 150 mM $SrCl_2$
- FGQ1 (50 μM) in IC buffer (pH 7.4)
- FGQ1 (50 μM) in IC buffer containing 40%PEG (pH 7.4)
- FGQH1 (50 μM) in IC buffer (pH 7.4)

- FGQH1 (50 μM) in IC buffer containing 40%PEG (pH 7.4)
- FGQH2 (50 μM) in IC buffer (pH 7.4)
- FGQH2 (50 μM) in IC buffer containing 40%PEG (pH 7.4)
- FiMHT (50 μM) in Sodium Acetate buffer (30 mM, pH 5.0) containing 100 mM NaCl
- FiMHT (50 μM) in Sodium phosphate buffer (30 mM, pH 7.4) containing 100 mM NaCl

Note: All NMR samples contained 20 % D_2O.

3. Results and Discussions

3.1. Repurposing 5-fluoro-2'deoxy uridine (FdU) for GQ and i-motif analysis

FdU is also called as floxuridine and belongs to a class of drugs called antimetabolites. It is prescribed for the treatment of cancers. It acts as a substrate analog and inhibits thymidylate synthase, which causes an imbalance in the cellular concentration of dNTP leading to induced cell death.[42] FdU has three advantages[43]- (i) fluorine label at position 5 is one of the conservative modifications and is less likely to affect the structure of nucleic acids, (ii) ^{19}F is 100% abundant and its chemical shift is highly sensitive to its local environment, and (iii) absence of fluorine in the biological systems would eliminate background signal, which is a problem with other NMR nuclei. Based on these considerations, we decided to repurpose FdU in studying the structure and recognition abilities of GQ and i-motif structures *in vitro* and in native cellular settings.

3.2. Sensitivity of FdU to changes in polarity and viscosity of the solvent medium

FdU incorporated into ON sequences can experience different microenvironment due to changes in interaction with neighboring bases. In order to study its sensitivity at the nucleoside level, we carried out ^{19}F NMR experiments in solvents of different polarity and viscosity. It is evident from Figure 5 that FdU gives unique chemical shifts for different solvents based on polarity and viscosity. This observation encouraged us to incorporate FdU into GQ and i-motif forming H-Telo DNA ON sequences.

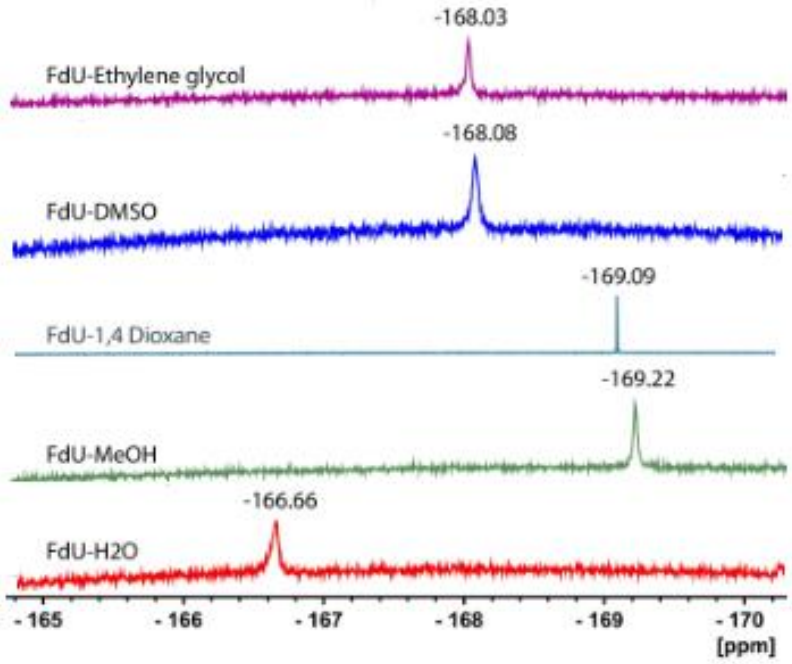


Figure 5: ^{19}F NMR of FdU in different solvents. The data presented here shows in order of increasing polarity from water to 1, 4-dioxane followed by higher viscous solvents DMSO and ethylene glycol.

3.3. Synthesis of DMT-FdU-phosphoramidite substrate 2 and its incorporation into ONs

FdU was incorporated into H-Telo G-rich and C-rich repeats by using standard solid-phase ON synthesis protocol.[44] The ON sequences, namely FGQ1, FGQH1, FGQH2 and FiMHT that were synthesized are listed in Table 1. The reason for choosing the mentioned ONs is that they adopt distinct conformations in particular ionic conditions. Thus, they give the advantage to qualitatively confirm different kinds of topologies adopted by H-Telo DNA ONs progressively starting from *in vitro* conditions to live cells such as mammalian cells and frog oocytes. The modification was introduced at the loop region, TTA in case of GQs and TAA in case of i-motifs, due to the following reasons: (i) incorporation into the G-rich/C-rich region could potentially affect the formation of G-quadruplex/i-motif structures[45] and (ii) the loop residues adopt a distinct conformation in different GQ[22]/i-motif[46] topologies. The labeled ONs were purified by gel electrophoresis (20% gel) using urea as the denaturant. The purity of the ONs was checked by reverse-phase HPLC

(Figure 6) and confirmed by mass spectrometry (mass spectra in Annexure and Table 1 next page).

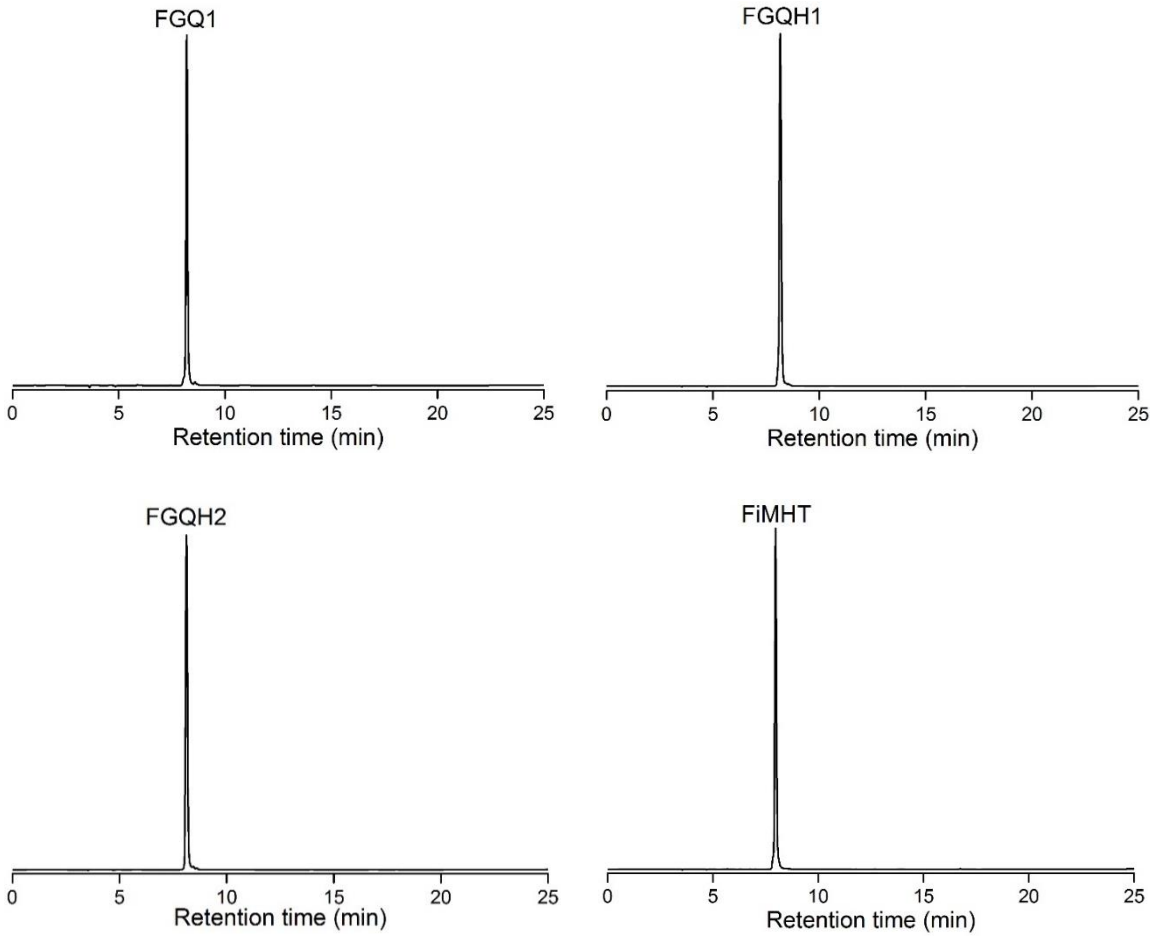


Figure 6: HPLC profiles of PAGE-purified H-Telo DNA ONs at 260 nm.

Table 1: Molar extinction coefficients and mass analysis of H-Telo DNA ONs

H-Telo DNA ON	$\varepsilon_{260}{}^{a}$ ($M^{-1}cm^{-1}$)	Calculated mass	Observed mass
FGQ1	227260	6969.16	6969.32
FGQH1	243060	7577.25	7577.69
FGQH2	251860	7881.30	7881.85
FiMHT	219160	7133.70	7133.71

Note: $\varepsilon_{260}{}^{a}$ values of modified ONs were calculated using IDT oligo-analyzer. The $\varepsilon_{260}{}^{a}$ value of modified oligonucleotide (7320 $M^{-1}cm^{-1}$) was used in the place of thymidine.

3.4. FdU modification does not affect the structure and stability of labeled GQs and i-motifs

The effect of modification on GQ and i-motif structures and their stability was evaluated by performing CD and thermal melting experiments using FdU-modified and unmodified control ONs in different ionic conditions. FGQ1 formed anti-parallel conformation in NaCl with a positive peak at ~293 nm, a strong negative peak at ~263 nm and a positive peak at ~245 nm (Figure 7a).[32] FGQ1 showed parallel conformation in $SrCl_2$ (negative peak at ~240 nm and positive peak at ~260 nm, Figure 7b).[5] FGQ1 formed hybrid type mixed parallel-antiparallel GQ structures in KCl (positive peak at nearly 290 nm and a shoulder at nearly 270 nm, Figure 7c).[32] FGQH1 showed a pattern matching with the hybrid type-1 GQ structure in the IC buffer (strong positive peak at nearly 285 nm and a small shoulder peak around 270 nm, Figure 7e).[41,47] FGQH2 presented a pattern resembling the hybrid type-2 GQ structure in the IC buffer (strong positive peak at nearly 290 nm and a shoulder peak around 266 nm, Figure 7f).[41,49] IC buffer contains a high concentration of K^+ ions, apart from other ions, and hence, FGQ1 showed a CD profile similar to that in potassium phosphate buffer containing K^+ ions alone (Figure 7d). Importantly, respective control unmodified ONs formed distinct GQ structures similar to the modified ONs in various conditions tested. These results are consistent with the literature reports for the same sequences.

H-Telo DNA ONs in a synthetic molecular crowded condition such as in the presence of PEG is known to adopt a parallel GQ topology.[30] To study the structural transformation, CD spectrum of ONs was taken in the presence of 40% PEG. Consistent with the literature reports both control and modified ONs formed parallel GQ structure (Figure 8a).[30,41] Further, control and labeled ONs annealed with the complementary ON sequence exhibited CD profiles typical for a duplex regardless of the ionic conditions (Figure 8b).

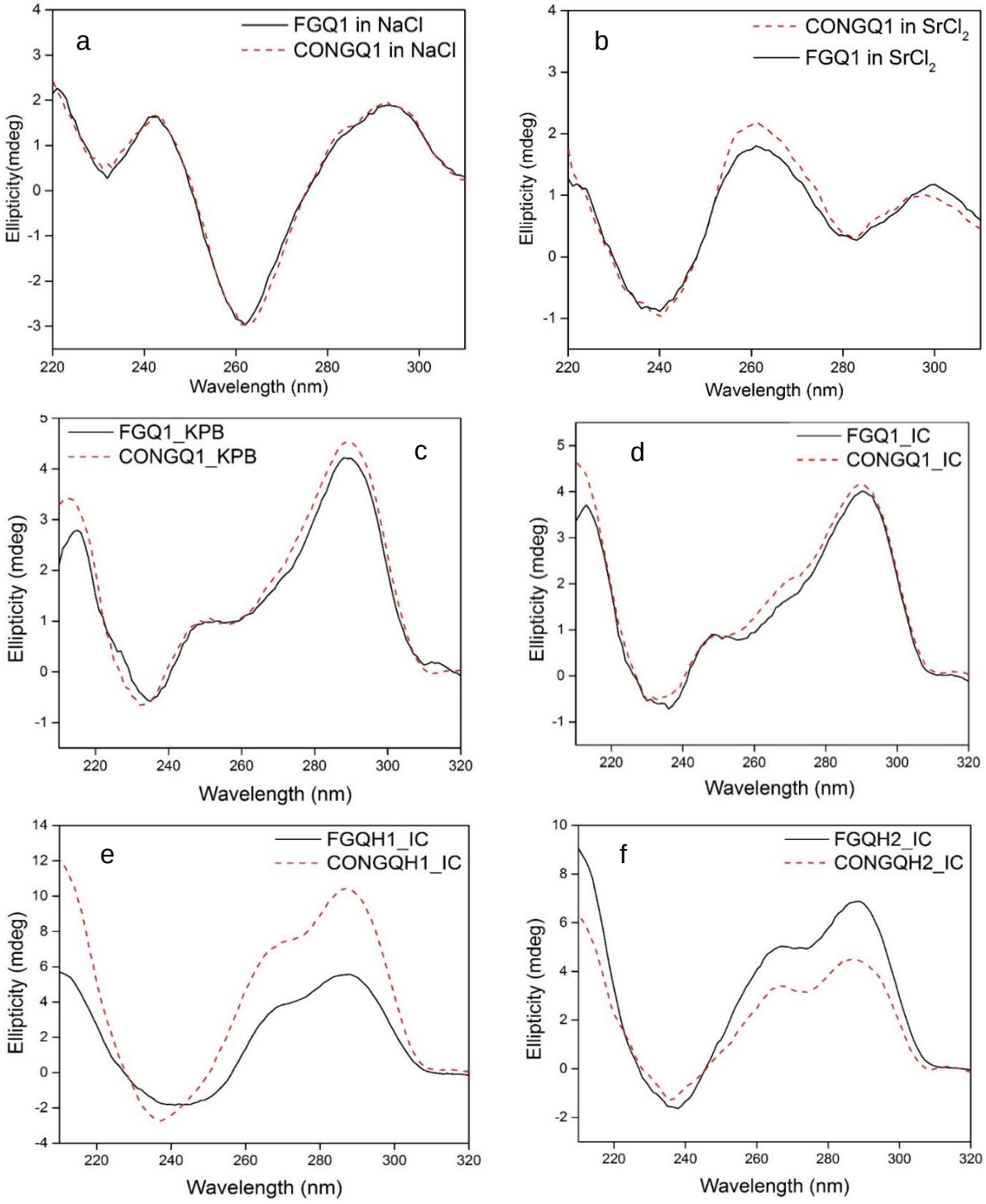


Figure 7: CD spectrum of (a) FGQ1 (5 μM) in 100 mM NaCl (b) FGQ1 (5 μM) in 150 mM $SrCl_2$ (c) FGQ1 (8 μM) in 100 mM KCl (d) FGQ1 (8 μM) in IC buffer (pH 7.4). Similarly, CD spectrum of (e) FGQH1 (8 μM) is shown in IC buffer, (f) FGQH2 (8 μM) in IC buffer.

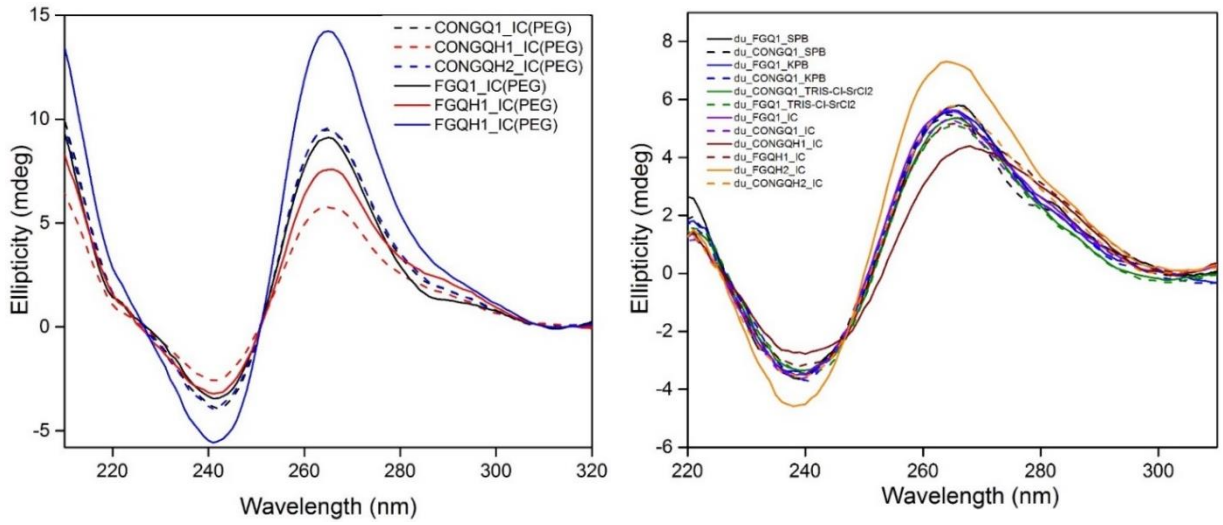


Figure 8: (a) CD study of H-Telo DNA ONs (8 μM) in IC buffer containing 40%PEG (b) CD profiles of duplexes consisting of control modified and unmodified ONs annealed with their respective complementary ON sequence.

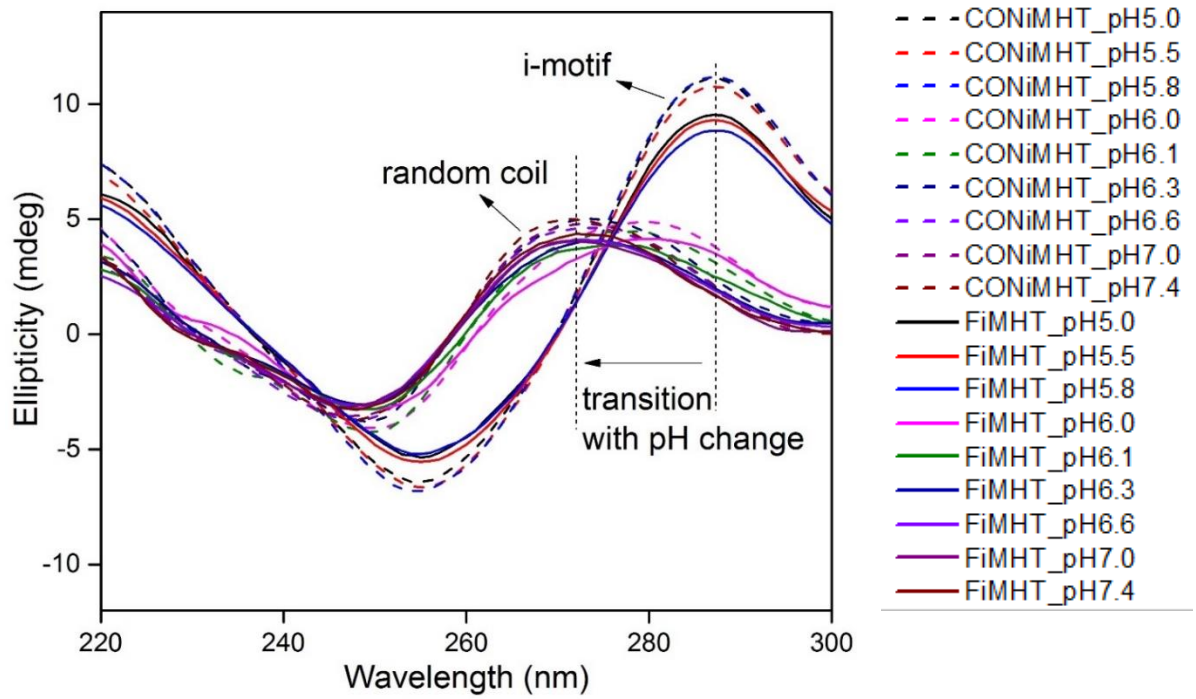


Figure 9: CD study of FiMHT in buffers of different pH.

Next, CD profile of FdU-labeled C-rich ON FiMHT and its control ON was recorded as a function of pH (Figure 9). At pH 7.4 and till pH 6.0 the ONs formed a random coil

structure (positive peak at ~272 nm and a negative peak at ~248 nm).[5] Further, reduction in pH resulted in a CD profile matching the i-motif structure (positive band centered around ~287 nm and negative band centered around ~254 nm).[5] A plot of changes in ellipticity at 287 nm versus pH gave a sigmoidal curve, which allowed to determine the *ipH* value for the transition of random unfolded structure to i-motif structure (Figure 10). The ipH value was found to similar for the control (5.95 ± 0.03) and modified sequences (5.93 ± 0.01) which was in consensus with the literature reports.[5,49]

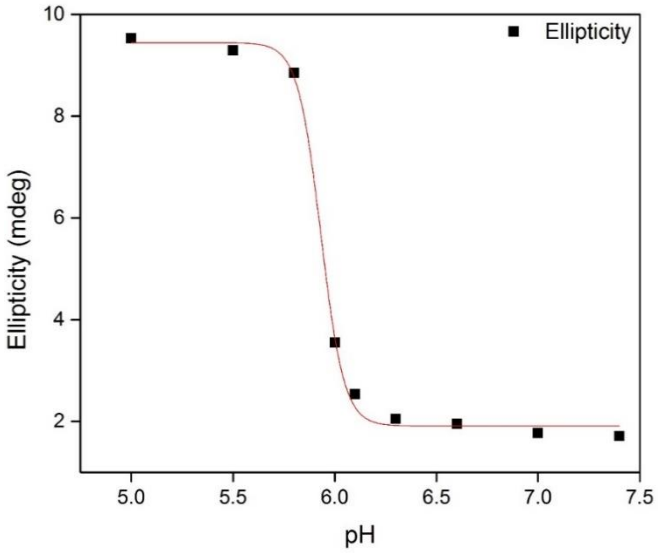


Figure 10: Determination of ipH of FiMHT by using circular dichroism (CD) studies.

The melting temperatures for GQ structures in different ionic conditions were found to be consistent with literature reports (Table 2 & 3).[41] In confirmation with CD studies, T_m values indicated that the modification had a minimal or no impact on the stability of GQs formed in different ionic conditions (Figure 11a to 11d). From a comparison of all the melting curves, the results indicate that parallel conformation is the most stable and anti-parallel the least stable with the mixed parallel anti-parallel coming in-between (Figure 11e). Similarly, we performed thermal melting experiments on i-motif forming ON sequence FiMHT in sodium acetate buffer (pH 5.0, pH 5.5) (Table 4 & Figure 12).[5] The reason for conducting the study at these pH values is that the C-rich H-Telo DNA ON FiMHT forms i-motifs under acidic conditions and as also observed in our CD experiments, they transform to random coil with increasing pH.

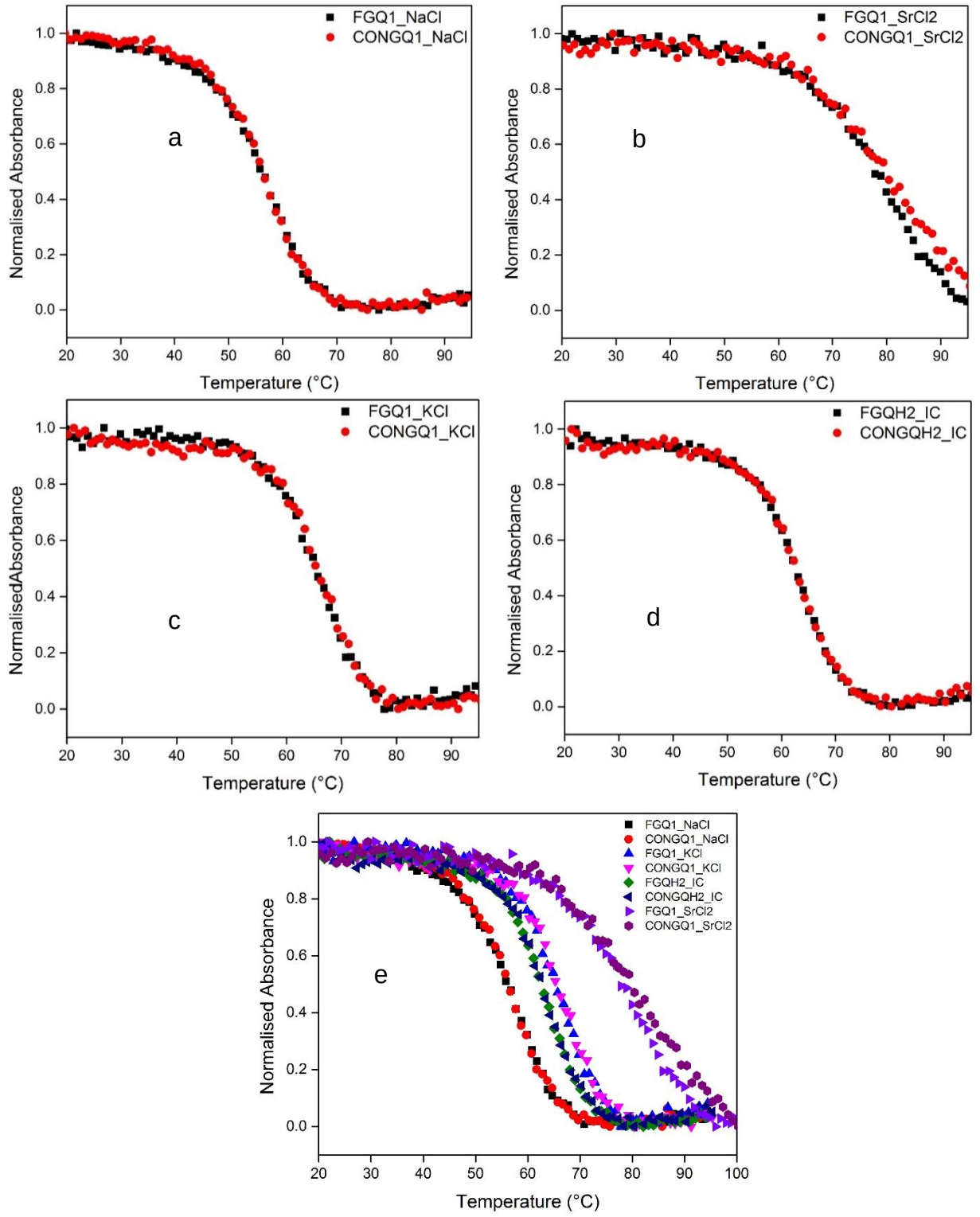


Figure 11: Thermal melting profile of (a) FGQ1 (1 μM) in 100 mM NaCl, (b) FGQ1 (1 μM) in 150 mM $SrCl_2$, (c) FGQ1 (1 μM) in 100 mM KCl, (d) FGQH2 (1 μM) in IC buffer, (e) comparison of stability of different GQ topologies.

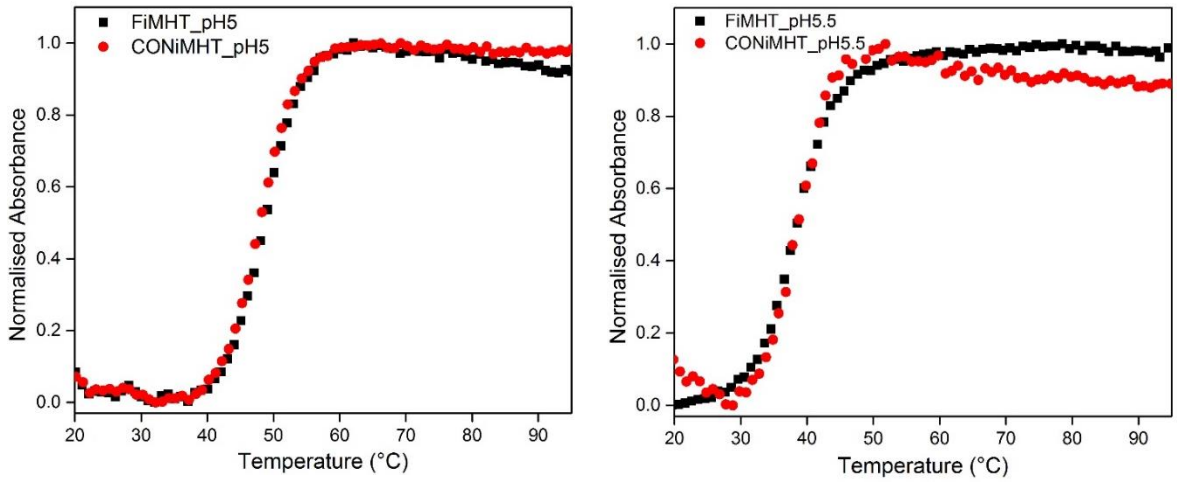


Figure 12: Thermal melting study on (a) FiMHT (1 μM) in sodium acetate buffer (30 mM) containing 100 mM NaCl (pH 5.0) (b) FiMHT (1 μM) in sodium acetate buffer (30 mM) containing 100 mM NaCl (pH 5.5).

Table 2 & 3: Calculated T_m values of H-Telo DNA G-quadruplexes

H-Telo DNA ON	T_m in $SrCl_2$	T_m in NaCl	T_m in KCl
FGQ1	84.1 ± 0.6	57.8 ± 0.4	65.7 ± 0.7
CONGQ1	84.7 ± 0.6	55.7 ± 0.6	66.3 ± 0.5

H-Telo DNA ON	T_m in IC buffer
FGQH2	64.1 ± 1.0
CONGQH2	65.3 ± 0.6

Table 4: Calculated T_m values of H-Telo DNA i-motifs

H-Telo DNA ON	pH 5.0 sodium acetate buffer (T_m)	pH 5.5 sodium acetate buffer (T_m)
FiMHT	49.0 ± 0.2	39.7 ± 0.3
CONiMHT	48.3 ± 0.4	39.7 ± 0.7

Conclusion form CD and UV-thermal melting analysis: Extensive CD experiments with control and modified i-motif and GQ -forming sequences in different ionic conditions and pH values, indicate that FdU incorporation does not affect the formation of native structure. Similar T_m values exhibited by control and modified ONs indicate the modified nucleoside does not affect the stability of GQ and i-motif structures. Inspired by these results, we next sought to obtain ^{19}F NMR signatures for different GQ and i-motif forms.

3.5. ^{19}F NMR of GQ-forming ONs

Control H-Telo DNA ON sequence (CONGQ1) is known to form anti-parallel conformation in NaCl,[32] parallel conformation in $SrCl_2$ [5] and multiple GQ structures in KCl.[32] Similarly, it is known from the literature that CONGQH1 and CONGQH2 form predominantly hybrid 1 and hybrid 2 conformations in K^+ ionic conditions.[41,47,49] ^{19}F NMR spectrum was recorded for DNA ONs annealed in the presence of different ions and the results are shown in Figure 13.

FGQ1 showed multiple peaks in KCl solution. This is in line with the reported literature[41] where it has been shown that the same ON sequence adopted multiple structures including hybrid-type GQs. FGQ1 showed a single peak in NaCl solution corresponding to an anti-parallel GQ conformation.[32] Similarly, FGQ1 showed a single broad peak in $SrCl_2$, which corresponds to the parallel GQ conformation as reported in the literature.[5] Intracellular (IC) buffer contains a high concentration of K^+ ions, apart from other ions, and hence, multiple peaks corresponding to different GQ structures were observed as in the case of a buffer containing K^+ ions alone. FGQH1, which forms predominantly a hybrid type-1 GQ structure gave a single peak in IC buffer.[41,47] However, FGQH2 showed multiple peaks in IC buffer, with a major peak at -164.98 ppm, which is possibly hybrid type-2 structure. Previously, it was reported that the same unmodified sequence adopted predominantly (~ 60% - 70%) hybrid 2 conformation.[41,48] It is important to mention here that the ONs in IC buffer did not adopt a parallel GQ structure.

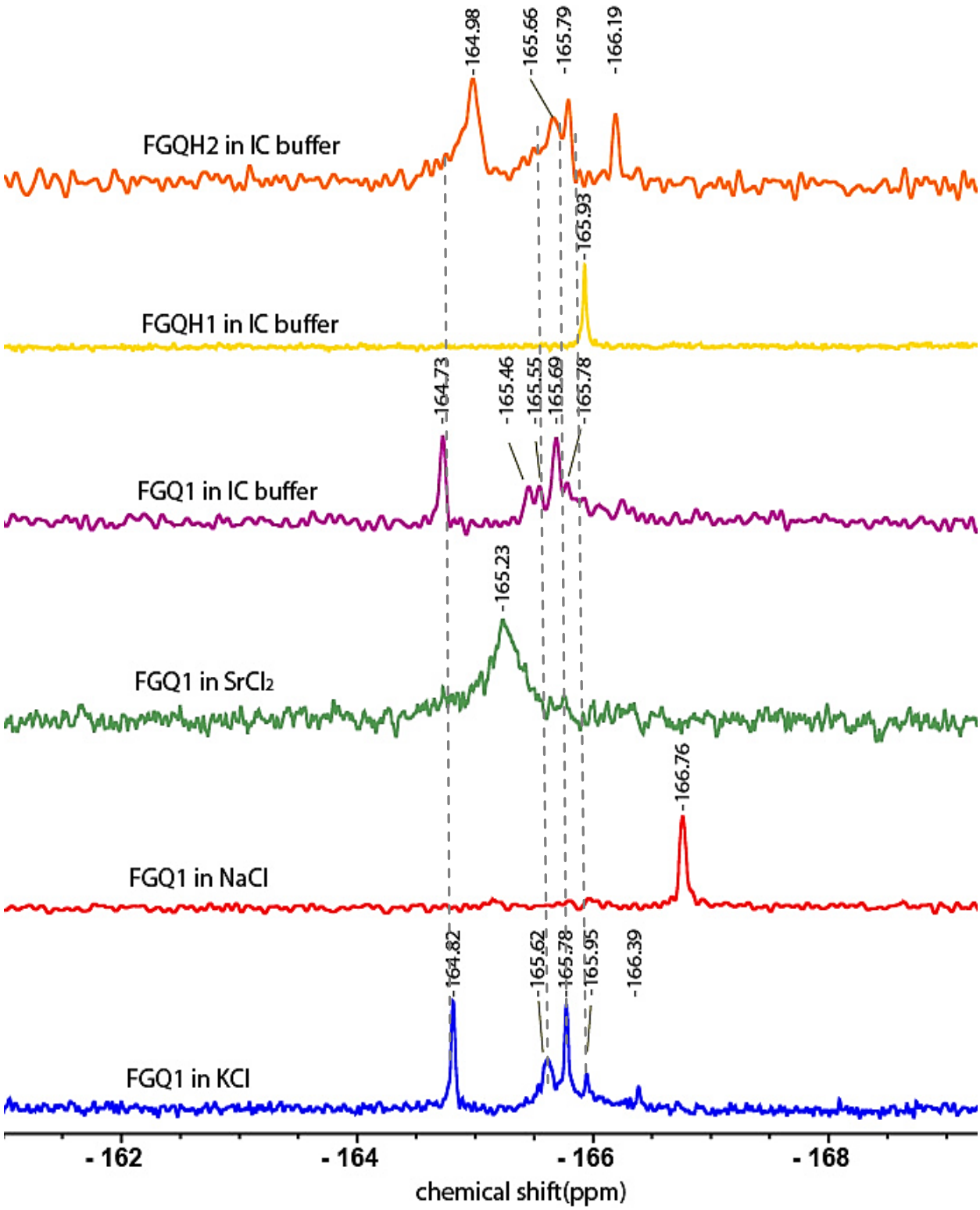


Figure 13: ^{19}F NMR signatures of H-Telo DNA ONs (50 μM) in different ionic conditions as mentioned on the left edge of the figure. All NMR experiments were carried out at 298 K or 25 °C.

Synthetic crowding agents such as PEG and deep eutectic solvents support the formation of a parallel stranded GQ structure.[30,41] To study the structural transformation by ^{19}F NMR, FGQ1, FGQH1, and FGQH2 ONs annealed in the presence of PEG were subject to NMR analysis. Interestingly, all the ONs in the presence of PEG almost completely transformed into the parallel GQ structure (Figure 14). Collectively these results indicate that FdU is an efficient GQ sensor and provides resolved ^{19}F signatures for different GQ structures, without affecting the native fold.

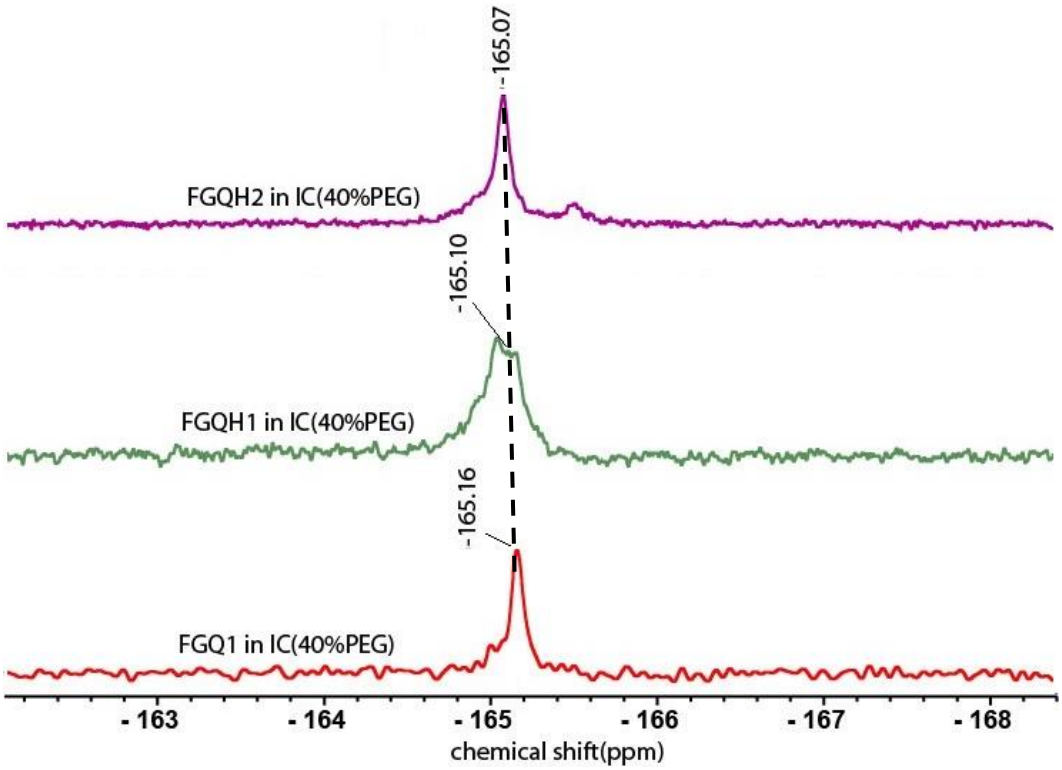


Figure 14: ^{19}F NMR signature of parallel conformation of H-Telo DNA ONs (50 μM) in IC (pH 7.4) containing 40% PEG. ONs get converted to a parallel from in the presence of PEG200. All NMR experiments were carried out at 298 K or 25 °C.

3.6. Detection of i-motif structure by ^{19}F NMR

Encouraged by the ability of FdU to detect different GQ structures, we sought to expand its utility in probing i-motif structures. i-Motif forming FiMHT ON was subjected to ^{19}F NMR analysis in buffer at different pH. At physiological pH (7.4), the ON forms a random coil and produces a signal at -165.06 ppm. At pH 5.0, this sequence adopts an i-motif and

produces a signal at -165.85 ppm (Figure 15). These results are consistent with the CD experiments, and remarkably, FdU gives distinct ^{19}F signature for random coil and i-motif structure. We plan to set up more NMR experiments at intermediate pH values in IC buffer to evaluate the progression of structural transformation from random coil to i-motif form.

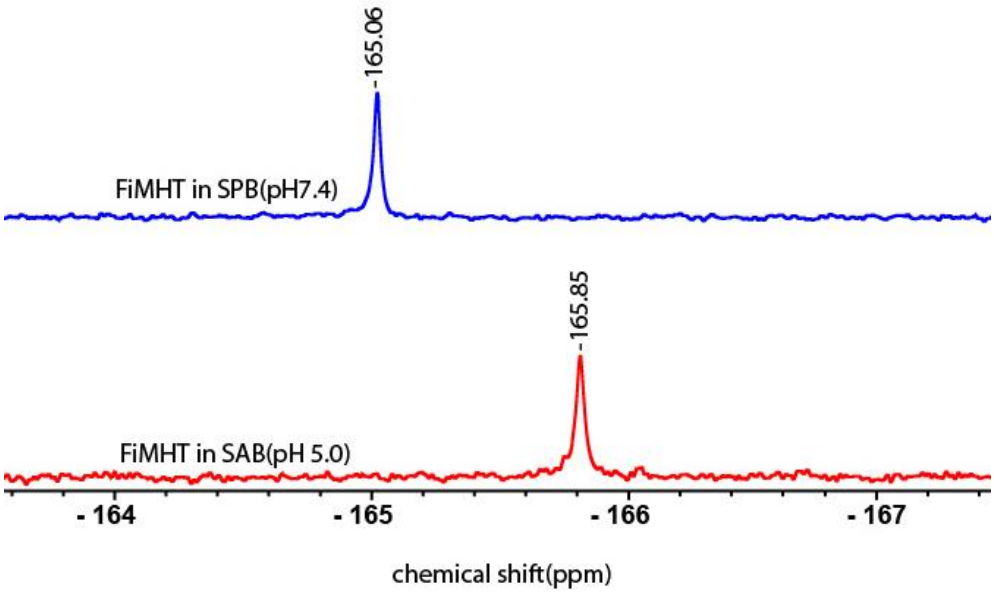


Figure 15: ^{19}F NMR of C-rich H-Telo DNA ON FiMHT (25 μM) in (a) Sodium acetate buffer (pH 5.0) containing 100 mM NaCl. (b) Sodium phosphate buffer (pH 7.4) containing 100 mM NaCl. All NMR experiments were carried out at 298 K.

4. Conclusions and future plan

We were able to design a probe that is practically non-perturbing and highly conformation-sensitive. The ^{19}F label helps us in distinguishing different GQ topologies by providing unique signatures in different ionic conditions. We found out using the probe that G-rich H-Telo DNA ONs did not assume parallel GQ structure in IC buffer conditions despite the previous findings. Further, with the incorporation of this probe in C-rich H-Telo ON, it was able to detect the transition from i-motif to random coil. Given the quality of our probe, we will be conducting further experiments in mammalian cells or frog oocytes which can help in co-relating our findings in *in vitro* conditions. Besides, we are also planning to perform experiments to elucidate the qualitative ratios of GQ/i-motif/double helix adopted by a double-stranded G-rich: C-rich H-Telo DNA in both *in vitro* and native cellular conditions.

5. References

1) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. The Double-Stranded-RNA-Binding Motif: Interference and Much More. *Nat Rev Mol Cell Biol* **2004**, *5* (12), 1013–1023.
2) Saini, N.; Zhang, Y.; Usdin, K.; Lobachev, K. S. When Secondary Comes First – The Importance of Non-Canonical DNA Structures. *Biochimie* **2013**, *95* (2), 117–123.
3) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters: G-Quadruplex and i-Motif in Oncogene Promoters. *FEBS Journal* **2010**, *277* (17), 3459–3469.
4) Belmont, P.; Constant, J. F.; Demeunynck, M. Nucleic Acid Conformation Diversity: From Structure to Function and Regulation. *Chem. Soc. Rev.* **2001**, *30* (1), 70–81.
5) 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* (22), 4141–4150.
6) Meyne, J.; Ratliff, R. L.; Moyzis, R. K. Conservation of the Human Telomere Sequence (TTAGGG)n among Vertebrates. *Proceedings of the National Academy of Sciences* **1989**, *86* (18), 7049–7053.
7) Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: Genome-Wide Database of G4 DNA Occurrence and Conservation in Human, Chimpanzee, Mouse and Rat Promoters and 146 Microbes. *Nucleic Acids Research* **2007**, *36* (Database), D381–D385.
8) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proceedings of the National Academy of Sciences* **2002**, *99* (18), 11593–11598.
9) Gomez, D.; Guédin, A.; Mergny, J.-L.; Salles, B.; Riou, J.-F.; Teulade-Fichou, M.-P.; Calsou, P. A G-Quadruplex Structure within the 5′-UTR of TRF2 MRNA

Represses Translation in Human Cells. *Nucleic Acids Research* **2010**, *38* (20), 7187–7198.

10) Alzeer, J.; Vummidi, B.; Roth, P. â€ …C.; Luedtke, N. Guanidinium-Modified Phthalocyanines as High-Affinity G-Quadruplex Fluorescent Probes and Transcriptional Regulators. *Angew. Chem. Int. Ed.* **2009**, *48* (49), 9362–9365.
11) Wang, X.-D.; Ou, T.-M.; Lu, Y.-J.; Li, Z.; Xu, Z.; Xi, C.; Tan, J.-H.; Huang, S.-L.; An, L.-K.; Li, D.; Gu, L.-Q.; Huang, Z.-S. Turning off Transcription of the *Bcl-2* Gene by Stabilizing the *Bcl-2* Promoter Quadruplex with Quindoline Derivatives. *J. Med. Chem.* **2010**, *53* (11), 4390–4398.
12) McLuckie, K. I. E.; Waller, Z. A. E.; Sanders, D. A.; Alves, D.; Rodriguez, R.; Dash, J.; McKenzie, G. J.; Venkitaraman, A. R.; Balasubramanian, S. G-Quadruplex-Binding Benzo[*a*]Phenoxazines Down-Regulate *c-KIT* Expression in Human Gastric Carcinoma Cells. *J. Am. Chem. Soc.* **2011**, *133* (8), 2658–2663.
13) Wood, B. R. The Importance of Hydration and DNA Conformation in Interpreting Infrared Spectra of Cells and Tissues. *Chem. Soc. Rev.* **2016**, *45* (7), 1980–1998.
14) Ding, Y.; Fleming, A. M.; Burrows, C. J. Case Studies on Potential G-Quadruplex-Forming Sequences from the Bacterial Orders Deinococcales and Thermales Derived from a Survey of Published Genomes. *Sci Rep* **2018**, *8* (1), 15679.
15) 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. *Nature Chem* **2018**, *10* (6), 631–637.
16) Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential. *Nat Rev Mol Cell Biol* **2017**, *18* (5), 279–284.
17) Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res* **2016**, *44* (4), 1746–1759.
18) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res* **2015**, *43* (18), 8627–8637.
19) Halder, K.; Wieland, M.; Hartig, J. S. Predictable Suppression of Gene Expression by 5′-UTR-Based RNA Quadruplexes. *Nucleic Acids Research* **2009**, *37* (20), 6811–6817.

20) Collie, G. W.; Parkinson, G. N. The Application of DNA and RNA G-Quadruplexes to Therapeutic Medicines. *Chem. Soc. Rev.* **2011**, *40* (12), 5867.

21) Neidle, S. Quadruplex Nucleic Acids as Targets for Anticancer Therapeutics. *Nat Rev Chem* **2017**, *1* (5), 0041.

22) Chen, Y.; Yang, D. Sequence, Stability, and Structure of G-Quadruplexes and Their Interactions with Drugs. *Current Protocols in Nucleic Acid Chemistry* **2012**, *50* (1).

23) Marušič, M.; Plavec, J. The Effect of DNA Sequence Directionality on G-Quadruplex Folding. *Angew. Chem. Int. Ed.* **2015**, *54* (40), 11716–11719.

24) Mondal, S.; Kundu, S. K.; Bhaumik, A. A Facile Approach for the Synthesis of Hydroxyl-Rich Microporous Organic Networks for Efficient CO_2 Capture and H_2 Storage. *Chem. Commun.* **2017**, *53* (18), 2752–2755.

25) Miyoshi, D.; Fujimoto, T.; Sugimoto, N. Molecular Crowding and Hydration Regulating of G-Quadruplex Formation. In *Quadruplex Nucleic Acids*; Chaires, J. B., Graves, D., Eds.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, **2012**; Vol. 330, pp 87–110.

26) Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. Formation and Stabilization of G-Quadruplex in Nanosized Water Pools. *Chem. Commun.* **2010**, *46* (10), 1700.

27) Panda, D.; Saha, P.; Das, T.; Dash, J. Target Guided Synthesis Using DNA Nano-Templates for Selectively Assembling a G-Quadruplex Binding c-MYC Inhibitor. *Nat Commun* **2017**, *8* (1), 16103.

28) Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3′ Sequence of d(AG $_3$ (TTAG $_3$) $_3$). *J. Am. Chem. Soc.* **2010**, *132* (48), 17105–17107.

29) Lannan, F. M.; Mamajanov, I.; Hud, N. V. Human Telomere Sequence DNA in Water-Free and High-Viscosity Solvents: G-Quadruplex Folding Governed by Kramers Rate Theory. *J. Am. Chem. Soc.* **2012**, *134* (37), 15324–15330.

30) Heddi, B.; Phan, A. T. Structure of Human Telomeric DNA in Crowded Solution. *J. Am. Chem. Soc.* **2011**, *133* (25), 9824–9833.

31) Manna, S.; Panse, C. H.; Sontakke, V. A.; Sangamesh, S.; Srivatsan, S. G. Probing Human Telomeric DNA and RNA Topology and Ligand Binding in a Cellular Model by Using Responsive Fluorescent Nucleoside Probes. *ChemBioChem* **2017**, *18* (16), 1604–1615.

32) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human Telomeric Sequence Forms a Hybrid-Type Intramolecular G-Quadruplex Structure with Mixed Parallel/Antiparallel Strands in Potassium Solution. *Nucleic Acids Research* **2006**, *34* (9), 2723–2735.

33) Mishra, S. K.; Jain, N.; Shankar, U.; Tawani, A.; Sharma, T. K.; Kumar, A. Characterization of Highly Conserved G-Quadruplex Motifs as Potential Drug Targets in Streptococcus Pneumoniae. *Sci Rep* **2019**, *9* (1), 1791.

34) (a) Guéron, M.; Leroy, J.-L. The I-Motif in Nucleic Acids. *Current Opinion in Structural Biology* **2000**, *10* (3), 326–331. (b) Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem. Int. Ed.* **2018**, *57* (8), 2165–2169.

35) (a) Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. Telomeric Repeat-Containing RNA Structure in Living Cells. *Proceedings of the National Academy of Sciences* **2010**, *107* (33), 14579–14584. (b) Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Are Stable and Detectable in Human Genomic DNA. *Nat Commun* **2013**, *4* (1), 1796. (c) 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 Research* **2014**, *42* (2), 860–869. (d) 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* (26), 8521–8525. (e) Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ren, J.; Zhao, Y.; Huang, Z.-S. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human

Telomeres. *Cell Chemical Biology* **2016**, *23* (10), 1261–1270. (f) 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 Research* **2017**, *45* (9), 5501–5511.

36) (a) Rajendran, A.; Nakano, S.; Sugimoto, N. Molecular Crowding of the Cosolutes Induces an Intramolecular I-Motif Structure of Triplet Repeat DNA Oligomers at Neutral PH. *Chem. Commun.* **2010**, *46* (8), 1299. (b) Brazier, J. A.; Shah, A.; Brown, G. D. I-Motif Formation in Gene Promoters: Unusually Stable Formation in Sequences Complementary to Known G-Quadruplexes. *Chem. Commun.* **2012**, *48* (87), 10739. (c) Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H. Tuning the PH Response of I-Motif DNA Oligonucleotides. *ChemBioChem* **2015**, *16* (11), 1647–1656. (d) Wright, E. P.; Day, H. A.; Ibrahim, A. M.; Kumar, J.; Boswell, L. J. E.; Huguin, C.; Stevenson, C. E. M.; Pors, K.; Waller, Z. A. E. Mitoxantrone and Analogues Bind and Stabilize I-Motif Forming DNA Sequences. *Sci Rep* **2016**, *6* (1), 39456.

37) Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. The Dynamic Character of the *BCL2* Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **2014**, *136* (11), 4161–4171.

38) (a) Krishnan, Y.; Simmel, F. C. Nucleic Acid Based Molecular Devices. *Angew. Chem. Int. Ed.* **2011**, *50* (14), 3124–3156. (b) Modi, S.; Nizak, C.; Surana, S.; Halder, S.; Krishnan, Y. Two DNA Nanomachines Map PH Changes along Intersecting Endocytic Pathways inside the Same Cell. *Nature Nanotech* **2013**, *8* (6), 459–467. (c) Li, T.; Famulok, M. I-Motif-Programmed Functionalization of DNA Nanocircles. *J. Am. Chem. Soc.* **2013**, *135* (4), 1593–1599. (d) Dong, Y.; Yang, Z.; Liu, D. DNA Nanotechnology Based on I-Motif Structures. *Acc. Chem. Res.* **2014**, *47* (6), 1853–1860.

39) Day, H. A.; Pavlou, P.; Waller, Z. A. E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorganic & Medicinal Chemistry* **2014**, *22* (16), 4407–4418.

40) (a) Gehring, K.; Leroy, J.-L.; Guéron, M. A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993**, *363* (6429), 561–565. (b) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57* (1), 11–24. (c) Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning. *Angew. Chem. Int. Ed.* **2012**, *51* (1), 250–253. (d) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57* (1), 64–75. (e) Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem* **2013**, *14* (5), 540–558. (f) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59* (13), 5987–6011. (g) Dembska, A.; Bielecka, P.; Juskowiak, B. PH-Sensing Fluorescence Oligonucleotide Probes Based on an i-Motif Scaffold: A Review. *Anal. Methods* **2017**, *9* (43), 6092–6106.

41) 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* (39), 12622–12633.

42) Uchikubo, Y.; Hasegawa, T.; Mitani, S.; Kim, H.-S.; Wataya, Y. Mechanisms of Cell Death Induced by 5-Fluoro-2'-Deoxyuridine (FUdR)--Necrosis or Apoptosis after Treated with FUdR--. *Nucleic Acids Symposium Series* **2002**, *2* (1), 245–246.

43) Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. 19F NMR: A Valuable Tool for Studying Biological Events. *Chem. Soc. Rev.* **2013**, *42* (20), 7971.

44) Tanpure, A. A.; Srivatsan, S. G. Synthesis and Photophysical Characterisation of a Fluorescent Nucleoside Analogue That Signals the Presence of an Abasic Site in RNA. *ChemBioChem* **2012**, *13* (16), 2392–2399.

45) (a) Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J.-L. Guanines Are a Quartet's Best Friend: Impact of Base Substitutions on the Kinetics and Stability of Tetramolecular Quadruplexes. *Nucleic Acids Research*

2007, *35* (9), 3064–3075. (b) 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* (5), 1311–1318.

46) Phan, A. T.; Guéron, M.; Leroy, J.-L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere 1 1Edited by I. Tinoco. *Journal of Molecular Biology* **2000**, *299* (1), 123–144.

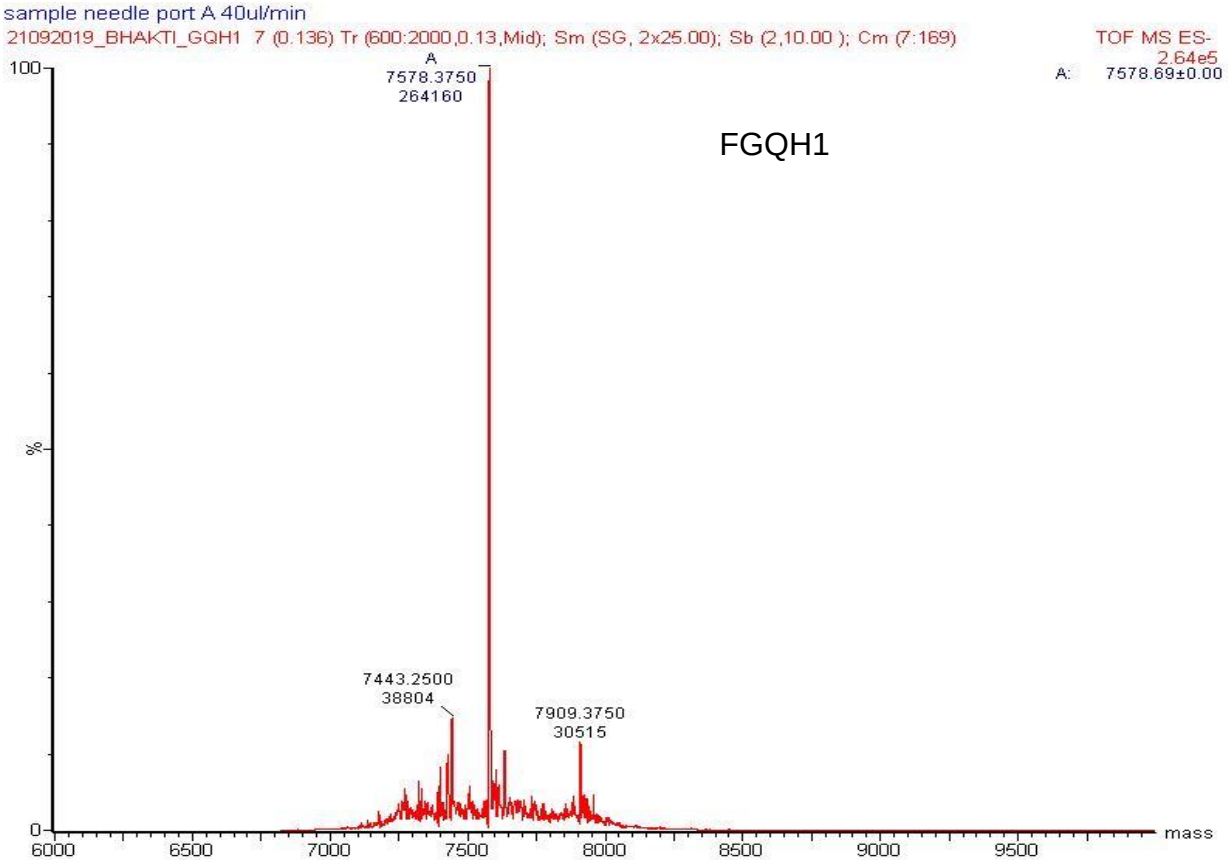
47) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K^+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128* (30), 9963–9970.

48) Phan, A. T.; Luu, K. N.; Patel, D. J. Different Loop Arrangements of Intramolecular Human Telomeric (3+1) G-Quadruplexes in K+ Solution. *Nucleic Acids Research* **2006**, *34* (19), 5715–5719.

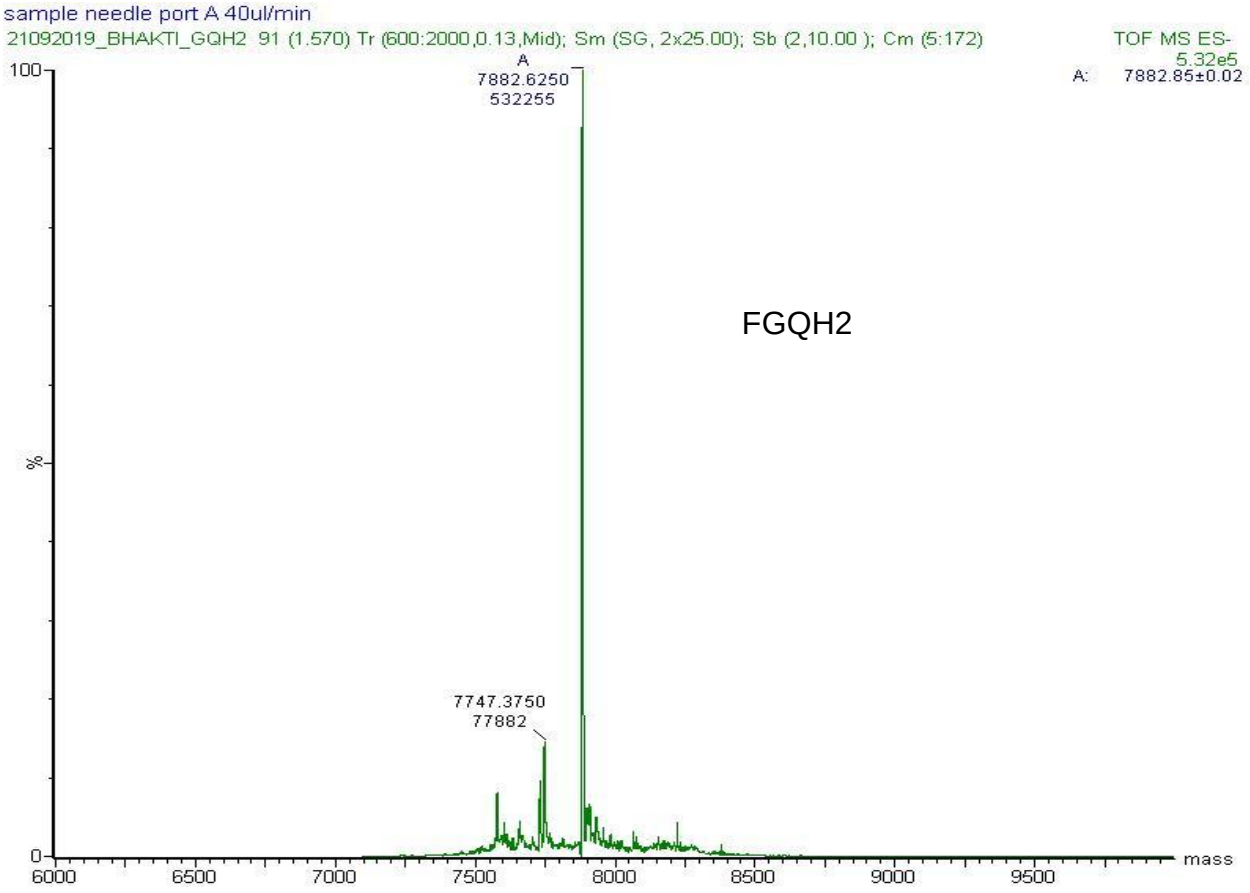
49) Mata, G.; Luedtke, N. W. Fluorescent Probe for Proton-Coupled DNA Folding Revealing Slow Exchange of *i* -Motif and Duplex Structures. *J. Am. Chem. Soc.* **2015**, *137* (2), 699–707.

6. Annexure

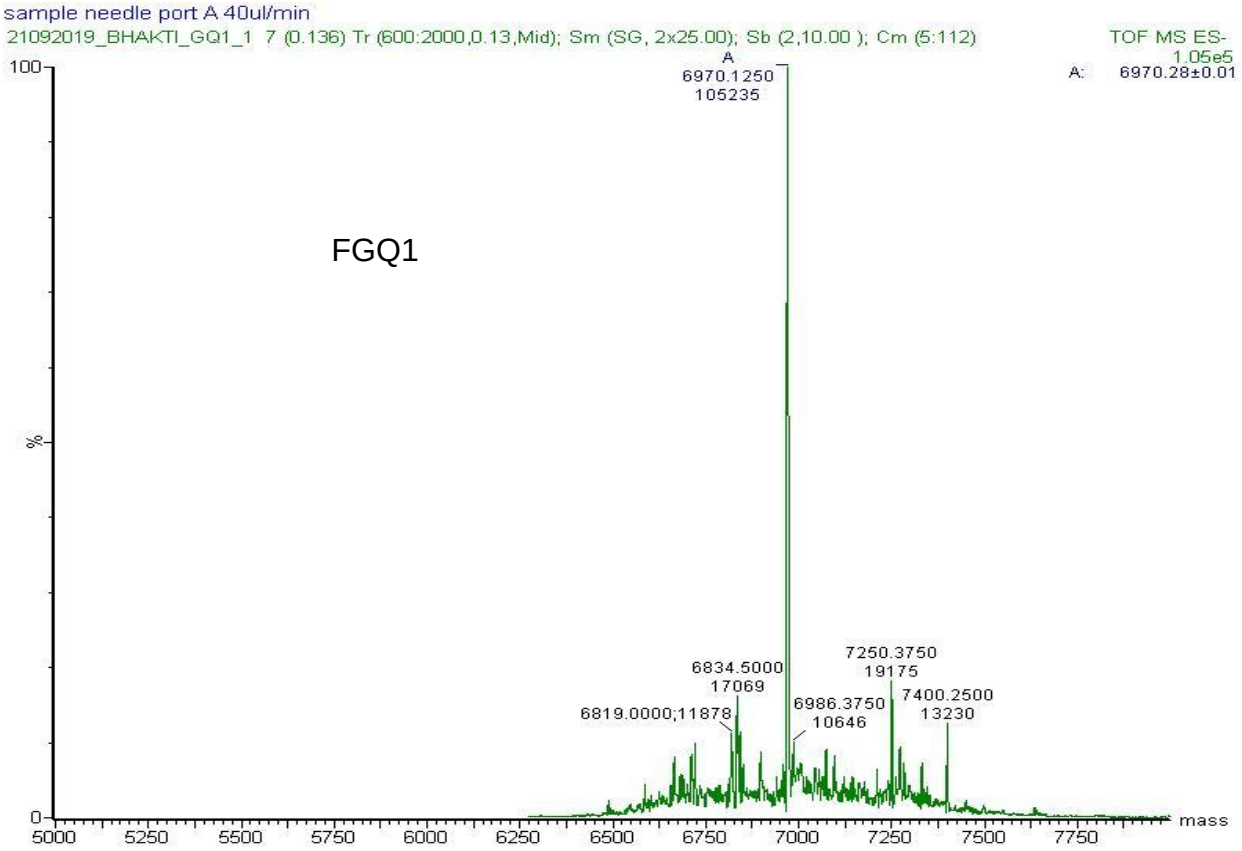
Mass spectra of FGQH1:



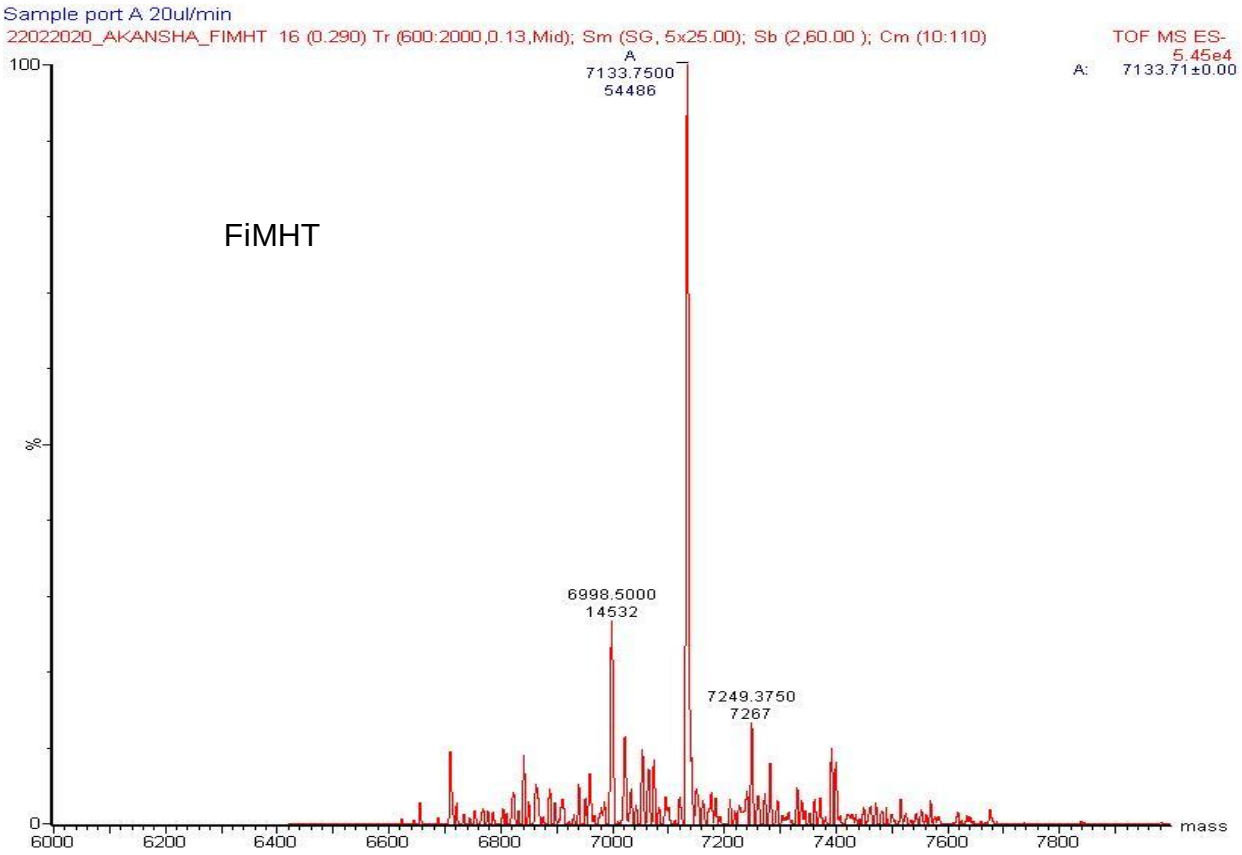
Mass spectra of FGQH2:



Mass spectra of FGQ1:

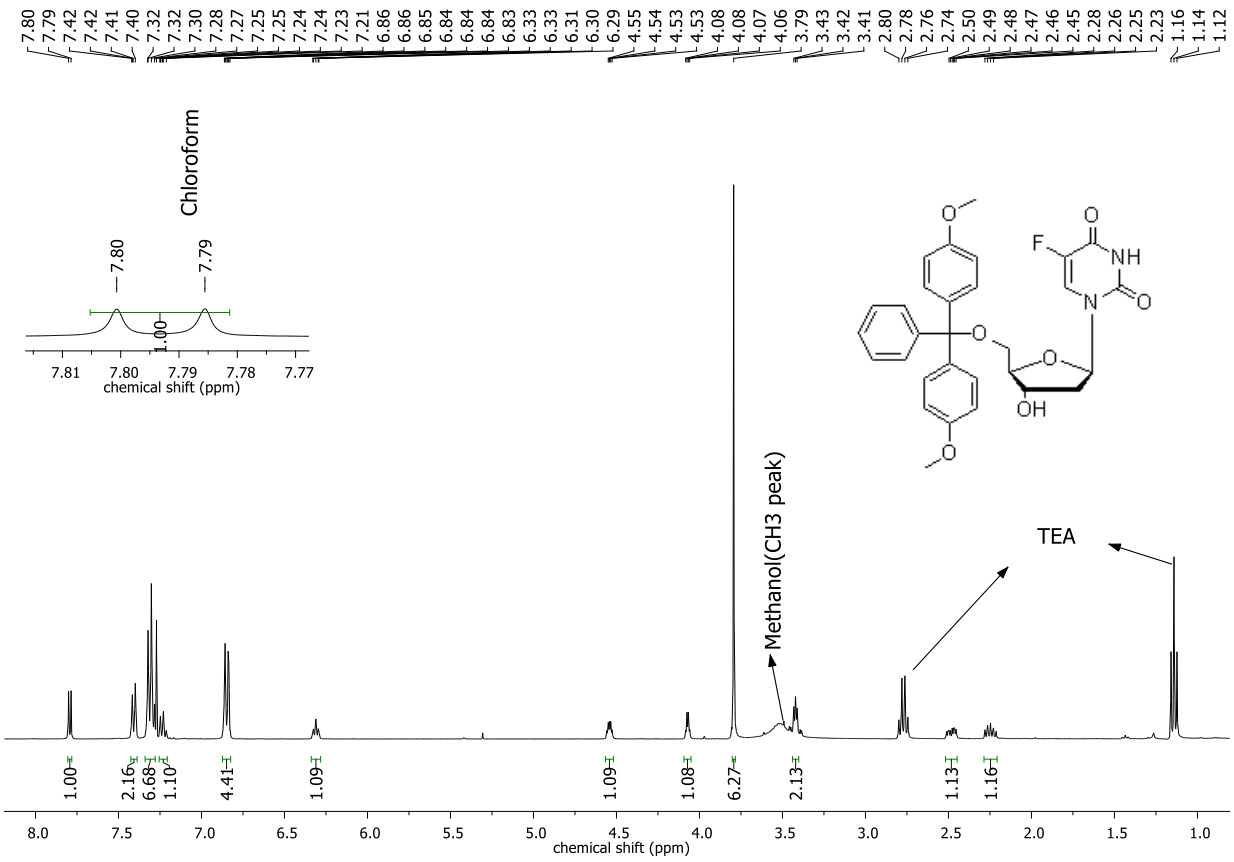


Mass spectra of FiMHT:

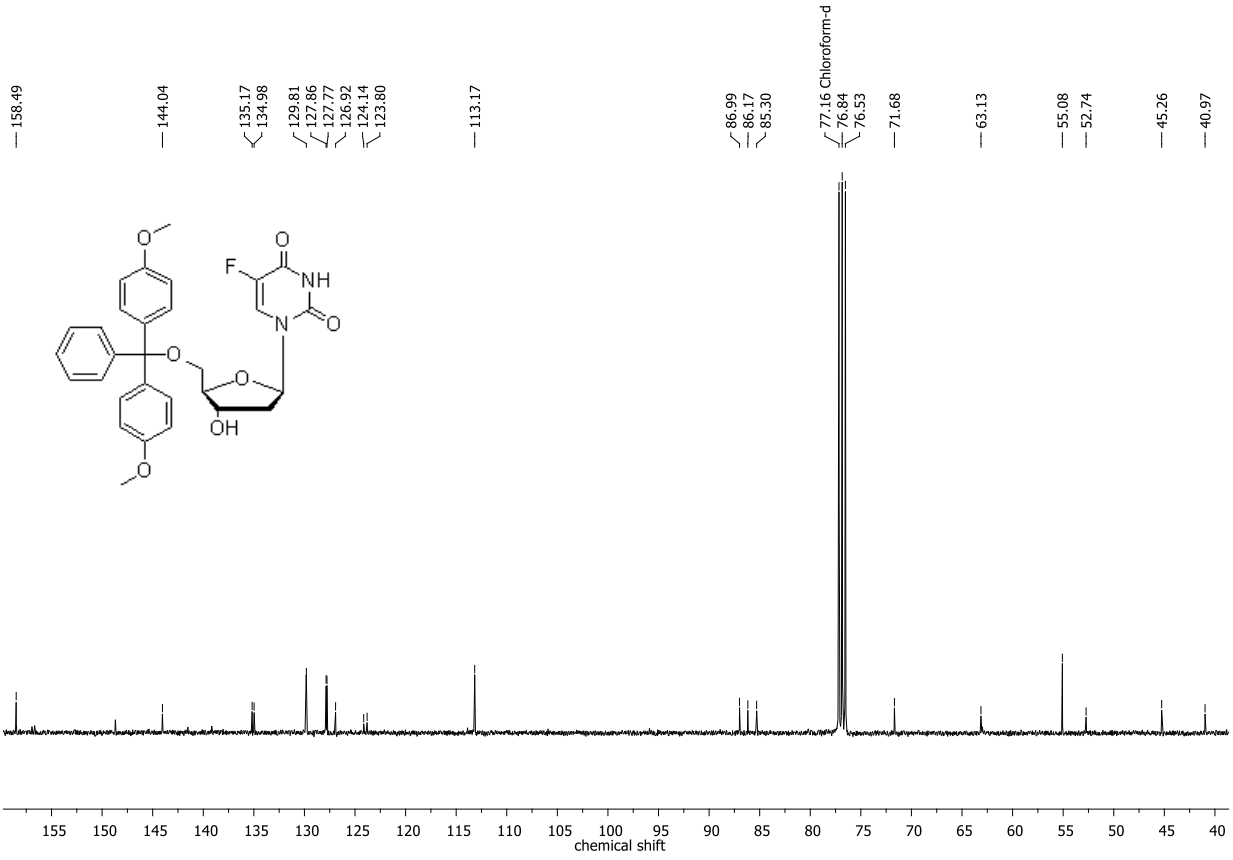


NMR Spectra:

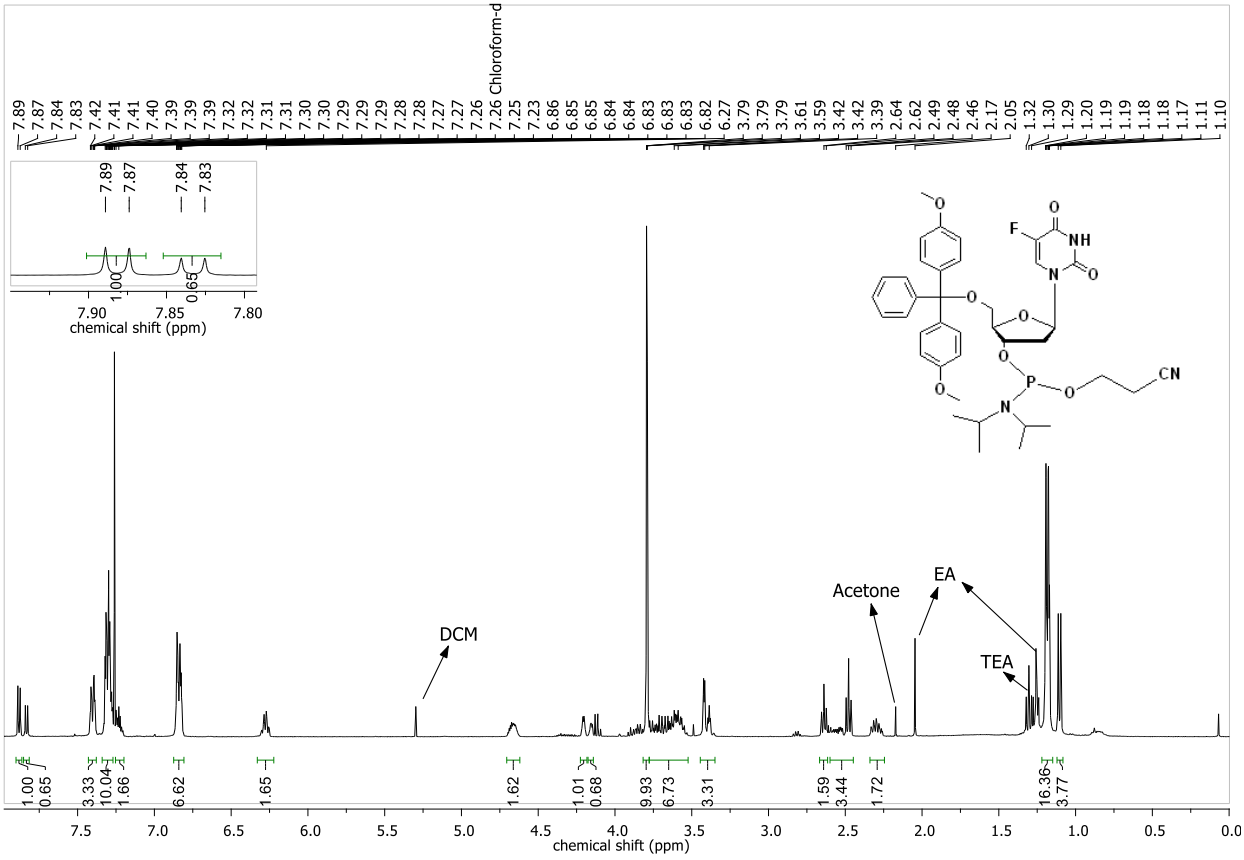
^{1}H NMR of **1(DMT-FdU)** ($CDCl_3$; 400 MHz)



^{13}C NMR of **1(DMT-FdU)** ($CDCl_3$; 400 MHz)



^{1}H NMR of **DMT-FdU-phosphoramidite substrate 2** ($CDCl_3$; 400 MHz)



^{31}P NMR of **DMT-FdU-phosphoramidite substrate 2** ($CDCl_3$; 400 MHz)