Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal

Cite this: DOI: 10.1039/d3sc00519d

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

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

Saddam Y. Khatik, [a] Sruthi Sudhakar, [b] Satyajit Mishra, [c] Jeet Kalia, [cd] P. I. Pradeepkumar [b] and Seergazhi G. Srivatsan *[a]

Received 30th January 2023
Accepted 30th April 2023

DOI: 10.1039/d3sc00519d

rsc.li/chemical-science

Paucity of efficient probes and small molecule ligands that can distinguish different G-quadruplex (GQ) topologies poses challenges not only in understanding their basic structure but also in targeting an individual GQ form from others. Alternatively, G-rich sequences that harbour unique chimeric structural motifs (*e.g.*, GQ-duplex or GQ-hairpin junctions) are perceived as new therapeutic hotspots. In this context, the epidermal growth factor receptor (EGFR) gene, implicated in many cancers, contains a 30 nucleotide G-rich segment in the promoter region, which adopts *in vitro* two unique architectures each composed of a GQ topology (parallel and hybrid-type) juxtaposed with a hairpin domain. Here, we report the use of a novel dual-app probe, C5-trifluoromethyl benzofuran-modified 2′-deoxyuridine (TFBF-dU), in the systematic analysis of EGFR GQs and their interaction with small molecules by fluorescence and ^{19}F NMR techniques. Notably, distinct fluorescence and ^{19}F NMR signals exhibited by the probe enabled the quantification of the relative population of random, parallel and hybrid-type GQ structures under different conditions, which could not be obtained by conventional CD and ^{1}H NMR techniques. Using the fluorescence component, we quantified ligand binding properties of GQs, whereas the ^{19}F label enabled the assessment of ligand-induced changes in GQ dynamics. Studies also revealed that mutations in the hairpin domain affected GQ formation and stability, which was further functionally verified in polymerase stop assay. We anticipate that these findings and useful properties of the nucleoside probe could be utilized in designing and evaluating binders that jointly target both GQ and hairpin domains for enhanced selectivity and druggability.

Introduction

The epidermal growth factor receptor (EGFR) is a transmembrane protein belonging to the family of protein kinase receptors. The EGFR gene codes for a tyrosine kinase receptor, which is activated by physiological extracellular ligands initiating an important signal transduction pathway that is required for normal cell growth, differentiation and proliferation in mammalian cells.[1,2] Overexpression or mutations that elevate the activity of the EGFR signalling pathway is directly linked to the progression of several cancers including that of lung, breast and glioblastoma.[3,4] Presently available therapeutic strategies to counter the effects of upregulation of EGFR activity employ tyrosine kinase inhibitors or monoclonal antibodies that interfere with the binding of ligands to the extracellular receptor domain.[5–7] However, their efficiency is limited by intrinsic or acquired resistance.[7,8] Alternatively, we envision that targeting non-canonical nucleic acid structural elements that act as natural regulators could be a viable strategy to control the EGFR expression at replication and transcription levels.[9–11] One such class of structures is a G-quadruplex (GQ), which is formed by guanine rich sequences.[12,13] Four guanine bases in a sequence *via* Hoogsteen hydrogen bonding form a tetrad (G-tetrad) and two or more G-tetrads, stabilized by monovalent cations (K^+ or Na^+), stack one above the other to form GQ structures.[14] Compelling experimental data indicate that GQs present in DNA and RNA serve as gene regulatory elements.[15,16] Dysfunction of these elements in several genes is linked to tumour progression,[17,18] and hence, small molecule ligands that stabilize these structures and down regulate gene expression are perceived as alternative therapeutic tools to mitigate cancer.[19–25] Though many small molecule ligands developed so far show good selectivity between GQ and duplex structures, they seldom distinguish between different topologies of GQs as they have a similar tetrad

[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
[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, fluorescence, mass and NMR spectra and gel images. See DOI: https://doi.org/10.1039/d3sc00519d

skeleton.[26] Recently, a few ligands have been developed that bind to a specific GQ topology.[27–32] Nevertheless, targeting a specific GQ motif amongst others in the genome remains a major challenge. Alternatively, G-rich segments that harbour GQ-duplex or GQ-hairpin junctions are considered as druggable targets,[33,34] wherein ligand scaffolds capable of simultaneously binding to GQ and duplex regions increase the specific targeting of such motifs.[35–37] In the context of the EGFR gene, a 30 nucleotide G-rich sequence upstream of the transcription start site (−272 position) forms two unique GQ topologies (parallel and hybrid-type), which are stabilized by a short hairpin located at the terminal loop.[38] This domain represents a new point of intervention to potentially attenuate the disease-causing activity of the gene. Therefore, it is important to first understand (i) the structural polymorphism of EGFR GQs, (ii) how the GQs interact with small molecule ligands and (iii) the functional role of GQs in a cellular process (*e.g.*, DNA replication).

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

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

Recently, we developed a highly environment-sensitive nucleoside analog by conjugating a fluorobenzofuran moiety at the C5 position of 2′-deoxyuridine (FBF-dU).[58] The heterocycle modification imparts fluorescence and endows an ^{19}F NMR label, thereby allowing a two-channel spectroscopic analysis of the GQ structures formed by the human telomeric overhang. While this analog is highly useful, assay conditions require very high concentrations of the labeled oligonucleotides (ONs, ~200 μM) and is not suitable for detecting lowly populated species coexisting in a dynamic equilibrium. In order to amplify the signal without compromising the sensitivity of the original probe system, we designed a second generation probe by conjugating trifluoromethyl benzofuran at the C5 position of 2′-deoxyuridine (TFBF-dU **1**). This probe containing three equivalent ^{19}F atoms significantly increased the detection limit and was used in detecting i-motif (iM) structures formed by model C-rich ONs.[59] Given the ability of an EGFR promoter to adopt unique GQ architectures, whose structural polymorphism could be perturbed by environmental conditions (*e.g.*, ionic conditions) and ligand binding, we decided to harness the true potential of TFBF-dU in studying this biologically important system in detail to validate its therapeutic potential.

Here, we report the use of TFBF-dU in probing GQ structures adopted by the wild-type and mutated EGFR promoter region, their population equilibrium and the influence of a hairpin junction in driving GQ formation. TFBF-dU incorporated into EGFR G-rich sequences provides distinct spectral readouts for the individual GQ structures, thereby enabling the quantification of their relative population under different conditions. The probe helps in identifying the preferred GQ topology in the absence and presence of small molecule ligands under intracellular ionic conditions. The utility of the probe in determining the GQ structure adopted by the ON in a cell-like environment is also demonstrated by performing ^{19}F NMR in frog egg lysate and extract. Furthermore, using polymerase stop

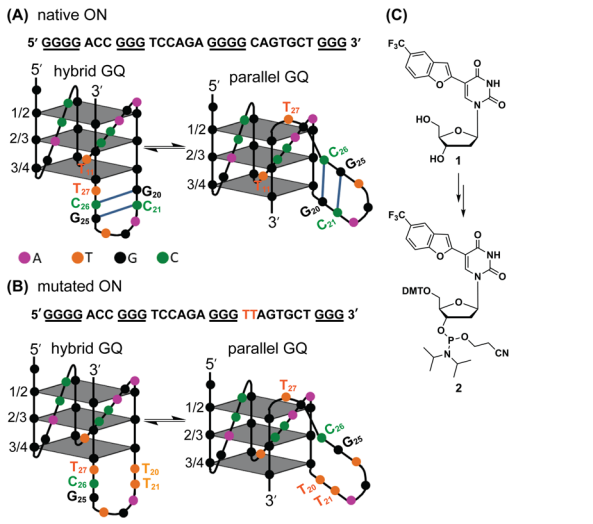


Fig. 1 (A) EGFR-272 G-rich sequence folds into a mixture of hybrid and parallel GQ topologies with a hairpin structure in their third loop.[38] (B) Mutations (G_{20} and C_{21} to T_{20} and T_{21}) in the EGFR G-rich ON abolishes the formation of the hairpin structure in both the GQ topologies. (C)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

assay we ascertained the implication of GQs and GQs bound to ligands in the DNA replication process.

Results and discussion

Synthesis of TFBF-dU-labeled ONs to detect GQs

To gain insights into the GQ folding dynamics and interaction of GQ structures with small molecule ligands, we decided to use TFBF-dU **1** for the following reasons. This fluorescent probe containing a ^{19}F label, when incorporated into ON sequences, is minimally perturbing and highly responsive to changes in the microenvironment as minute changes in the conformation and its interaction with neighbouring bases are manifested in the form of changes in fluorescence and ^{19}F chemical shifts.[59] 100% natural abundance, high sensitivity and absence of fluorine in biological systems make ^{19}F a good label for in-cell NMR analysis.[60,61] A notable number of fluorine-labeled probes have been developed to investigate different nucleic acid structures and their interactions with ligands or proteins.[62–68] The next important consideration we took into account is the placement of TFBF-dU in GQ-forming sequences. In comparison to G-tetrads, the loop residues that connect G-tetrads exhibit differences in conformation as well as interactions with adjacent bases in different GQ topologies (Fig. 1A). Hence, replacing a loop dT residue with TFBF-dU would have a very minimum impact on the GQ formation and would also provide distinct spectral signatures for different topologies by virtue of its ability to sense microenvironment changes (Fig. 1A). Before incorporation into the EGRF sequence, which forms multiple GQs, we first studied the use of the nucleoside analog by employing a hybrid-type GQ formed by a human telomeric DNA repeat ON sequence (Telo1, Table 1).[69,70] For this purpose, the modified ON Telo2 in which one of the loop dTs is replaced with the nucleoside analog was synthesized by the solid-phase ON synthesis method using phosphoramidite **2** (Fig. 1C and Table 1). Similarly, in EGFR ONs **3** and **4**, the nucleoside analog **1** was placed in the second loop (T_{11}) and third loop (T_{27}), respectively (Fig. 1A and Table 1). In addition, a mutated ON **5** (G_{20} and C_{21} to T) with a modification at T_{27} was synthesized to probe the role of a hairpin in the formation of GQs (Fig. 1B). All the ONs were purified by PAGE under denaturing conditions, and their purity and identity were confirmed by RP-HPLC and mass analyses, respectively (Fig. S1–S3 and Table S1†).

Table 1 TFBF-dU modified and respective control unmodified telomeric repeat and EGFR ONs

ON	5′–sequence—3′[a]
Telo1	TAGGGTTAGGGTTAGGGTTAGGGTT
Telo2	TAGGGTTAGGGT*TAGGGTTAGGGTT
TeloC	CCCTAACCCTAACCCTAACCCT
3	GGGGACCGGGT*CCAGAGGGGCAGTGCTGGG
4	GGGGACCGGGTCCAGAGGGGCAGTGCT*GGG
5	GGGGACCGGGTCCAGAGGGTTAGTGCT*GGG
6	GGGGACCGGGTCCAGAGGGGCAGTGCTGGG
7	GGGGACCGGGTCCAGAGGGTTAGTGCTGGG

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

Impact of TFBF-dU labeling on the formation of GQs

Native Telo1 and modified Telo2 ONs were annealed in 10 mM Tris·HCl buffer (pH 7.4) containing 100 mM KCl and the formation of the hybrid-type GQ structure and its stability were studied by CD and UV-thermal melting analysis. CD profiles of both control and modified ONs gave a positive band at ~285 nm, a prominent shoulder at ~269 nm and a negative band at ~240 nm characteristic of a hybrid-type GQ (Fig. S4A†).[71] The T_m values for the ONs were also found to be very similar (Fig. S4B and Table S2†). Furthermore, a thermal difference spectrum (TDS) obtained by subtracting the UV absorption spectrum of the folded state from that of the unfolded state gave a profile corresponding to a GQ structure (Fig. S5†).[72,73] Together, these results suggest that the modification has negligible impact on the structure and stability of a Telo GQ. Similarly, EGFR ONs **3–5** and control unmodified ONs **6** and **7** were annealed in 10 mM Tris·HCl buffer (pH 7.4), and the formation of GQ structures with an increase in K^+ ion concentration was monitored by CD spectroscopy. In the absence of KCl, EGFR ONs did not show a CD profile corresponding to a GQ structure (Fig. 2). In the presence of K^+ ions (100 mM), control ON **6** exhibited a positive peak at 265 nm and a negative peak at 240 nm indicating the formation of a parallel GQ (Fig. 2A and B, dashed lines). In addition, a prominent shoulder at 285 nm representing a hybrid-type GQ form was observed irrespective of the K^+ ion concentration (Fig. 2A, B and S6†).[38,74] Modified ON **3** gave significantly less intense peaks at +265 nm and −240 nm as compared to the native ON, suggesting that the modification at position T_{11} hampered GQ formation (Fig. 2A, red solid line). On the other hand, ON **4** displayed a CD profile (peak at +265 nm, −240 nm, and a shoulder at 285 nm) similar to that of control ON **6** (Fig. 2B, red lines). This confirms that nucleoside **1** placed at T_{27} in ON **4** has negligible impact on the GQ structure formation. Mutations in the stem of the hairpin affected GQ formation of ONs **5** and **7** with a more pronounced effect on the parallel topology (Fig. 2C). Furthermore, we observed that labeled ON **4** and control ON **6** exhibited a similar profile in all K^+ ion concentrations tested (50 mM to 200 mM, Fig. S6B†). Although the parallel GQ structure seemed to be higher in amounts as compared to the hybrid form, CD profiles suggested that these structures coexist with no apparent change in the structural equilibrium as a function of KCl concentration. UV-thermal melting analysis indicates that a TFBF-dU modification at T_{11} and T_{27} positions only marginally affects the stability of GQs as compared to that of unmodified control ONs (Fig. S7 and Table S2†). However, CD data indicate that a modification at T_{11} (ON **3**) affects the GQ equilibrium, and hence for further analysis native ON **4** and mutated ON **5** containing the modification at T_{27} were used.

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

TFBF-dU modified Telo2 ON predominantly forms a mixed parallel–antiparallel stranded hybrid-type 2 GQ structure like

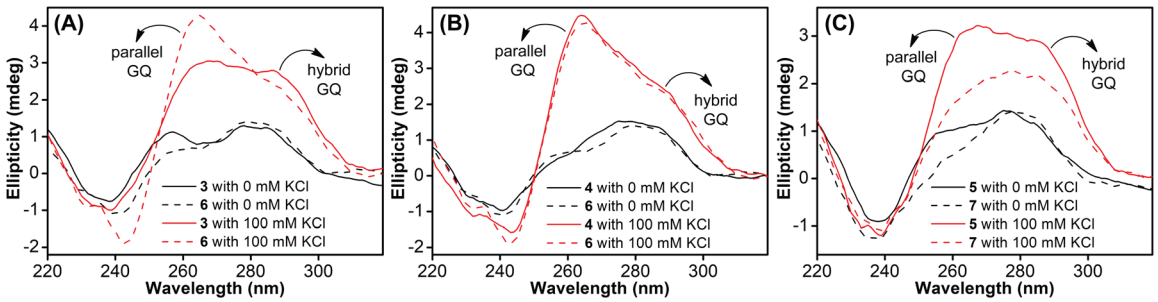


Fig. 2 (A, B and C) CD spectra (5 μM) of modified EGFR ONs **3**–**5** (solid lines) and their control ONs **6** and **7** (dashed lines) under different ionic conditions. See Fig. S6† for CD profiles at different KCl concentrations (50 mM to 200 mM).

that of control Telo1 ON (Fig. S8†).[69,70] Upon excitation at 330 nm, the GQ form of Telo2 exhibited a fluorescence band with an emission maximum at ~423 nm (Fig. 3A). When Telo2 was hybridized to its complementary C-rich strand (TeloC), it formed a duplex structure (Fig. S8†), which exhibited a very weak fluorescence band. The observed fluorescence outcome in these structures is due to the difference in the microenvironment around the probe. It is to be noted that several probes display significant quenching in fluorescence due to stacking interaction with neighbouring bases and when located near a guanine base.[75,76] TFBF-dU placed at the T_{12} position upon duplex formation would base pair and partially stack with adjacent bases namely, G_{11} and T_{13}. As a result of stacking interaction and closeness to the guanine base, the probe shows a quenched emission band. Based on the NMR structure of the hybrid 2 GQ form (PDB: 2JPZ),[70] the TFBF modification at the C5-position would be projected into the groove away from the G-tetrad formed by the G_{11} residue (Fig. S9†). Possibly due to reduced stacking interaction the GQ form shows higher fluorescence intensity compared to the duplex structure.[76] Much like the fluorescent component, the fluorine label was also found to be conformation sensitive. TFBF-dU gave a distinct ^{19}F NMR signal for GQ and duplex structures (Fig. 3B). The formation of these structures was also confirmed by an imino proton signal appearing in GQ and duplex regions (Fig. S10†). Collectively, these results indicate that our probe is structurally non-invasive and highly responsive to nucleic acid conformations, and hence, we embarked on using TFBF-dU in the analysis of EGFR GQs.

Probing EGFR GQs by fluorescence

Samples of EGFR ON **4** annealed in Tris·HCl buffer (pH 7.4) containing 100 mM LiCl or varying concentrations of K^+ ions were excited at 330 nm, and changes in steady-state emission were recorded. The unstructured ON in the presence of LiCl or in the absence of KCl (see CD profiles, Fig. 2B and S6B†) exhibited a very weak fluorescence band centered at around 419 nm (Fig. 4A). As the concentration of K^+ ions was increased (25–200 mM), a progressive increase in the fluorescence intensity with a minor shift in emission maximum was observed as a result of formation of GQs. The fluorescence intensity of ON **4** saturated at 150 mM of KCl (~8-fold increase as compared to the non-GQ form). In contrast, the CD profile was very similar at all K^+ ion concentrations tested suggesting that there is no apparent change in the structural equilibrium (Fig. S6B†). To further probe the structures formed by EGFR ON **4**, we performed time-resolved fluorescence analysis (Fig. 4B). In the absence of KCl, three decay components corresponding to lifetimes $\tau_1 = 1.07$ ns, $\tau_2 = 2.99$ ns, and $\tau_3 = 0.11$ ns were observed (Table 2). This is consistent with the literature report suggesting the possibility of three hairpin structures co-existing at zero concentration of KCl.[38] At 25 mM KCl, we observed three different components with lifetimes $\tau_1 = 1.44$ ns, $\tau_2 = 4.70$ ns, and $\tau_3 = 0.14$ ns. The individual lifetime components could correspond to random coil, parallel GQ and hybrid-type GQ structures in no particular order. The longest lifetime component ($\tau_2 = 4.70$ ns) was found to be significantly more

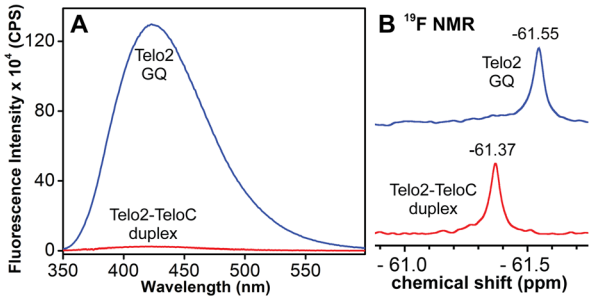


Fig. 3 (A) Fluorescence and (B) ^{19}F NMR spectra of telomeric GQ and duplex structures. See the ESI† for experimental conditions.

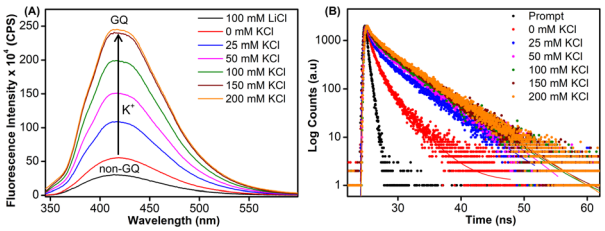


Fig. 4 (A) Steady-state and (B) time-resolved fluorescence spectra of modified ON **4** at different KCl concentrations. For steady-state fluorescence, samples (1 μM) were excited at 330 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively. The instrument response (prompt) is shown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

Table 2 Fluorescence lifetimes of TFBF-modified ON **4**

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

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

populated as compared to the other two fast decay components (τ_1 and τ_3). Interestingly, as we increased the concentration of KCl (25 mM to 200 mM), the population of the species corresponding to τ_2 increased from 63% to 80% with a concomitant decrease in the population of τ_1 and τ_3 (Table 2). Lifetime analysis also indicated a saturation point at 150 mM KCl similar to steady-state analysis. It is to be noted that the fluorescence of the free nucleoside analog was not significantly affected by changes in the K^+ ion concentration, indicating that the observed fluorescence is due to differences in the microenvironment of the nucleoside analog in different conformations adopted by the ON (Fig. S11A†).

Based on the fluorescence and CD data, and by comparing the structural models predicted for EGFR GQs in an earlier report,[38] we infer the following. The CD profile of ON **4** in the presence of KCl (200 mM) shows signatures for parallel and hybrid-type GQs with a ratio of 70 : 30.[38] This trend was consistent with the fluorescence data, and hence, we assigned the longest lifetime component (τ_2 = ~4.70 ns) to the parallel GQ topology and τ_1 and τ_3 to either a random coil or hybrid-type GQ structure. The loop residues in a juxtaposed quadruplex and hairpin motif can orient coaxially or orthogonally depending on the GQ topology adopted by the sequence. While a coaxial orientation between the hybrid-type GQ and stem-loop results in continuous base stacking across the two structures, orthogonal interaction between the parallel GQ and hairpin does not involve base stacking between the two components.[77,78] To gain further insights, models of hybrid and parallel GQ forms were built by computational analysis. The hairpin structure was found to be coaxially oriented with the hybrid-type GQ core (Fig. S12A†). In this topology, T_{27} was sandwiched between the hairpin domain and a G-tetrad (Fig. 5A and S12A†). Hence, TFBF-dU placed at this position may experience a similar environment, wherein it would be stacked between C_{26} of the hairpin and G_4 of the bottom tetrad. As a result of stacking interaction, the hybrid GQ structure of ON **4** exhibits a lower fluorescence intensity and a shorter lifetime. In the case of parallel topology, the hairpin structure was found to be orthogonal to the GQ core (Fig. S12B†). In this orientation, T_{27} was found to partially stack with G_{17} that formed the top G-tetrad (Fig. 5B and S12B†). TFBF-dU placed at T_{27} should experience reduced stacking interaction with neighbouring bases, and hence, the parallel form of ON **4** displays higher fluorescence and a longer lifetime. For these reasons, the fluorescence intensity and contribution of τ_2 increases as a function of K^+ ion concentration due to a progressive shift in the equilibrium towards the formation of the more emissive parallel GQ form.

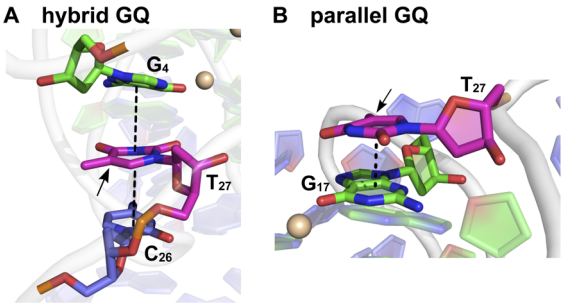


Fig. 5 Representative structure of the major cluster of (A) hybrid and (B) parallel GQ topologies. A zoomed-in image representing the stacking interaction of T_{27} (in magenta) with adjacent bases is shown here. Dashed lines represent the stacking interaction. The C5 position of T_{27} is shown by an arrow where TFBF heterocycle modification is placed in ON **4**. Carbon atoms of G-tetrads are represented in green and all other nucleotides are represented in purple. Nitrogen atoms are represented in blue, oxygen in red, and phosphate in orange. Hydrogen atoms are omitted for clarity. See the ESI† for details.

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

^{1}H NMR spectra of modified ON **4** and control ON **6** revealed broad peaks for imino protons in the GQ region (10 to 12 ppm) suggesting the presence of multiple GQ topologies in equilibrium. In addition, a peak at 12.8 ppm indicated the presence of a Watson–Crick paired hairpin structure (Fig. 6B and S13†). However, due to only minor changes in the signals, the effect of ionic conditions on the conformational equilibrium could not be deduced from ^{1}H NMR analysis. Rewardingly, the incorporated nucleoside analog **1** gave distinct and resolved ^{19}F signals for different structures. A systematic analysis using ^{19}F signatures obtained under different ionic conditions allowed us to gain a deeper understanding of the GQ structural equilibrium. Notably, upon increasing the concentration of KCl, the peak at −60.55 ppm increased progressively at the expense of other peaks (Fig. 6A). At 200 mM KCl, this peak was found to be predominant. CD and fluorescence data obtained at high concentrations of K^+ ions

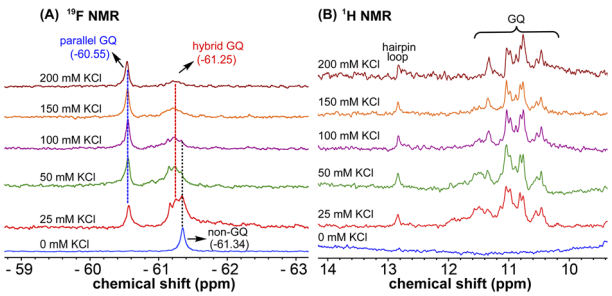


Fig. 6 (A) ^{19}F NMR and (B) ^{1}H NMR spectra of TFBF-modified ON **4** at different KCl concentrations.

indicate the formation of a parallel GQ as the major form and a hybrid-type GQ as the minor form. So, the peak at −60.55 ppm was judiciously assigned to a parallel GQ topology and the broad peak (minor, −61.25 ppm) to hybrid GQ topology (Fig. 6A). Furthermore, integrating the ^{19}F NMR peaks at different KCl concentrations provided valuable information on the population dynamics of parallel and hybrid GQ topologies. At 25 mM of KCl, the relative population of the parallel GQ was found to be 23%, and the hybrid topology and non-GQ structures accounted for the remaining population. Population of the parallel GQ significantly increased from 23% to 66% with increase in KCl concentration, while the population of hybrid GQ and non-GQ structures decreased to 34%, a result similar to the fluorescence data. A control experiment with nucleoside probe **1** in the presence of different concentrations of KCl did not affect the ^{19}F NMR chemical shift similar to fluorescence (Fig. S11B†). Collectively, these results confirm that ^{19}F signatures exhibited by modified ON **4** is due to differences in the microenvironment of the probe in different structures.

Juxtaposed hairpin structure influences the GQ structural equilibrium

To study the role of the hairpin junction, G_{20} and C_{21} nucleotides, involved in base paring, were mutated with Ts (ON **5**, Table 1). Mutated ON **5** exhibited contrasting fluorescence compared to native ON **4**. The random coil form of ON **5** in the absence of KCl displayed a weak emission band centered at 419 nm (Fig. 7A). Addition of KCl (25 mM) resulted in quenching in fluorescence with a slight blue shift in the emission maximum as a result of formation of GQ structures (CD profiles also suggest the formation of GQs, Fig. 2C). Furthermore, addition of KCl did not affect the fluorescence profile of ON **5**. The ON exhibited three different decay components with lifetimes τ_1 = 0.95 ns, τ_2 = 0.11 ns and τ_3 = 2.67 ns in the presence of K^+ ions (Fig. 7B and Table 3). Much like steady-state fluorescence, the population of these three components was more or less even and did not change with an increase in the KCl concentration (Table 3). The highest lifetime component (τ_3 = 2.67 ns) was assigned to the parallel GQ topology as before, whereas shorter lifetimes could correspond to hybrid-type GQ and random coil structures (Table 3). Collectively, fluorescence at different K^+ ion concentrations indicates that mutations in the hairpin domain significantly reduced the population of the more emissive parallel GQ topology and increased the amount of less emissive hybrid GQ topology.

Table 3 Fluorescence lifetimes of TFBF-modified mutated ON **5**

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

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

^{1}H NMR spectra of mutated ON **5** and its corresponding unmodified ON **7** in the presence of K^+ ions revealed broad imino proton peaks between 10 and 12 ppm indicating the formation of more than one GQ structure (Fig. 8B and S14†). Furthermore, the absence of a peak at 12.8 ppm confirmed that the 3rd loop did not adopt a hairpin structure. ^{19}F NMR of **5** in the absence of KCl showed a single peak at −61.44 ppm corresponding to a non-GQ structure (Fig. 8A). In the presence of KCl, ON **5** folded into a mixture of parallel (−60.83 ppm) and hybrid GQ (−61.25 ppm) topologies. Even at high concentrations of K^+ ions, the ON formed a random coil structure. The CD profile revealed the presence of a mixture of GQs along with a random coil form (Fig. 2C). As compared to native ON **4**, which majorly formed a parallel GQ conformation, mutated ON **5** formed different structures that were more or less evenly populated (*e.g.*, see peaks corresponding to parallel and hybrid-type at 100 mM KCl, Fig. S15†). These observations are also consistent with fluorescence intensity and lifetime data, alluding that both the labels behave similarly in reporting different conformations adopted by the ONs. Furthermore, the existence of a random coil structure along with GQs clearly indicates that mutations in the hairpin domain reduced the

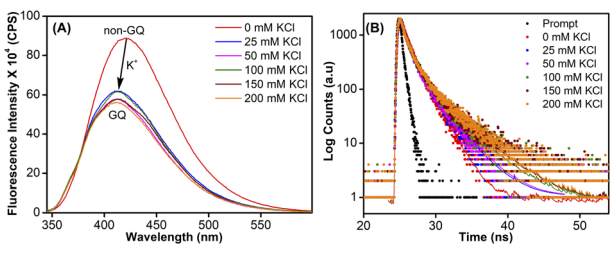


Fig. 7 (A) Steady-state and (B) time-resolved fluorescence spectra of modified ON **5** at different KCl concentrations. For steady state fluorescence, samples (1 μM) were excited at 330 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively. The instrument response (prompt) is shown in grey dots and decay curve fits are shown in solid lines.

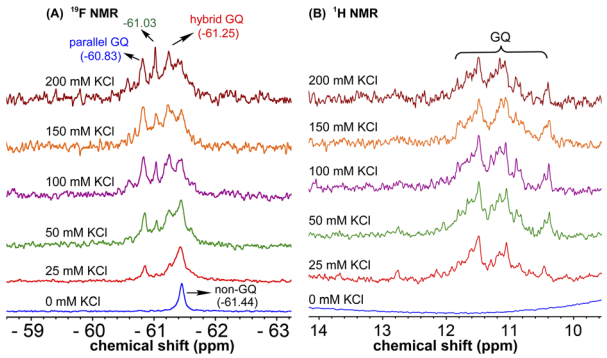


Fig. 8 (A) ^{19}F NMR and (B) ^{1}H NMR spectra of mutated modified ON **5** at different KCl concentrations. Peak at −61.03 could not be assigned.

overall stability of GQ structures. A ΔT_m of 4 °C between ON **4** and ON **5** also supports the above results obtained from fluorescence and NMR studies (Table S2†). Collectively, these results highlight the usefulness of our nucleoside probe in elucidating the role of a juxtaposed hairpin motif on the formation and stability of different GQs of the EGRF promoter region.

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

Using the dual-labeled analog, we devised assays to estimate the ligand binding to GQ structures formed by the EGFR G-rich segment. While the fluorescent label enabled the quantification of ligand binding to an ensemble of GQs, the ^{19}F label provided information on the efficiency of ligand interaction with different co-existing GQ structures. Titration of modified ON **4** with increasing concentrations of commercially available GQ binders (TMPyP4, PDS, and BRACO-19, Fig. 9A) resulted in a dose-dependent quenching in the fluorescence intensity with no apparent change in the emission maximum (Fig. 9B and S16†). At low concentrations of ligands, we observed a slight increase in fluorescence intensity. However, the quenching effect was significant as the ligand concertation was increased further, which gave reliable K_d values. The observed fluorescence quenching is likely due to the proximity of the probe to the bound polyaromatic ligands favouring a nonradiative decay pathway.[79] TMPyP4 and PDS displayed a similar affinity for an EGFR GQ with K_d values of 0.35 ± 0.02 μM and 0.37 ± 0.04 μM, respectively (Fig. 9C). However, BRACO-19 displayed a slightly lower affinity with a K_d value of 0.51 ± 0.05 μM. Furthermore, the tested ligands showed a cooperative binding with a hill coefficient of nearly 2. Since TMPyP4 binding to GQ structures exhibit a red-shifted absorption band,[80] the binding of this ligand to EGFR ON **4** was also evaluated by UV-vis absorption spectroscopy. Notably, UV titration experiments gave a comparable apparent K_d value (0.14 ± 0.01 μM) for TMPyP4 binding to GQs of ON **4** (Fig. S17†).

It is not straight forward to determine topology-specific binding of ligands when a sequence forms multiple structures. Ligands could bind either to a specific topology or topologies and stabilize the structures or could transform the structure, thereby redistributing the population.[81] The ability of our probe to display distinct ^{19}F signals for different GQs was put to use in detecting topology-specific binding of ligands. Modified ON **4**, which folds into a mixture of parallel (−60.56 ppm) and hybrid GQ (−61.20 ppm) topologies at 100 mM KCl was titrated with TMPyP4, PDS, and BRACO-19 (Fig. 9D and S18†). Upon increasing the TMPyP4 concentration, peak intensity corresponding to the parallel GQ (−60.56 ppm) decreased, and furthermore, it disappeared at a higher concentration of TMPyP4 (ON : ligand = 1 : 2, Fig. 9D). A broad peak at around −61.20 ppm increased with increasing concentrations of the ligand. A CD spectrum revealed the presence of multiple conformations (mainly parallel and hybrid forms, Fig. S19A†). Hence, it is speculated that the signal at −61.20 ppm could be associated with different GQ–ligand complexes. In contrast, PDS and BRACO-19 showed affinity for both parallel and hybrid GQ topologies. We observed that with an increase in PDS and BRACO-19 concentration, the peaks corresponding to parallel (−60.55 ppm) and hybrid (−61.21 ppm) GQs reduced, and new peaks corresponding to different GQ–ligand complexes appeared (Fig. S18A and B†), which was confirmed by the presence of CD signatures for both parallel and hybrid GQs (Fig. S19B and C†). Furthermore, ligands are known to interact

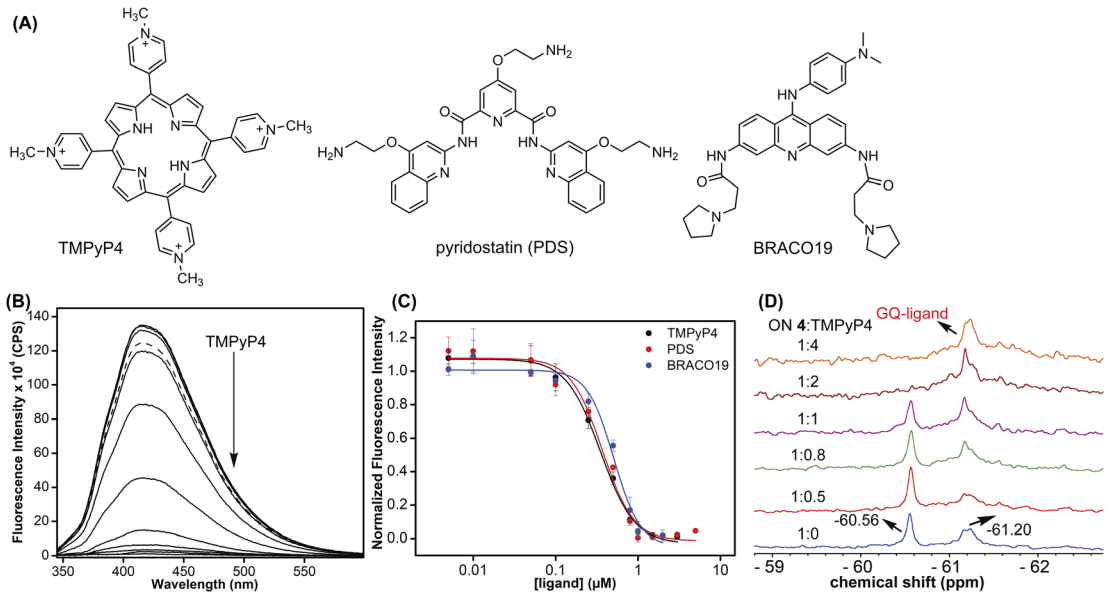


Fig. 9 (A) Structures of ligands used in this study. (B) Fluorescence spectra of ON **4** (0.5 μM) in 10 mM Tris·HCl buffer (pH 7.4) containing 100 mM KCl with increasing concentrations of TMPyP4 (solid lines). The fluorescence spectrum of ON **4** in the absence of a ligand (dashed line). Samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively. (C) Curve fits for the binding of ligands (TMPyP4, PDS, and BRACO-19) to ON **4**. (D) ^{19}F NMR of modified ON **4** (15 μM) as a function of increasing TMPyP4 concentration.

with GQs *via* different binding modes (external stacking, intercalation, and groove binding).[79] We speculate that multiple peaks observed upon PDS and BRACO-19 binding might be due to different modes of interaction of these ligands with GQs.[82,83] From our studies, we find that although ligands (TMPyP4, PDS, and BRACO-19) interact with EGFR GQs with similar K_d values (~0.35–0.5 μM), their specificity, mode of interaction and influence on structural equilibrium are different.

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

Next, we studied the utility of the ^{19}F label in detecting the formation of GQ structures in a model mimicking a cellular environment. Lysate and egg extract from frog oocytes serve as good *ex vivo* models to conduct NMR experiments to study the structure of biopolymers.[84,85] To gain a progressive understanding of GQ structures formed by EGFR ON **4** in cell-free and cellular environments, ^{19}F NMR experiments were performed under intraoocyte ionic conditions, oocyte clear lysate and egg extract. The clear lysate was prepared by mechanically crushing the oocytes and the suspension thus obtained was incubated at 95 °C for 10 min to denature the proteins. The lysate was obtained by centrifugation. To obtain an egg extract, the oocytes were crushed and without any further manipulations the mixture was centrifuged. The inter-phase egg extract is considered to provide a molecular crowded environment and also known to maintain most of the biological activity of a cell.[84] ^{19}F NMR of ON **4** recorded in a buffer mimicking intraoocyte ionic conditions (25 mM HEPES pH = 7.5, 110 mM KCl, 10.5 mM NaCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA) gave a spectral profile reminiscent of GQ structures formed in Tris buffer containing 100 mM KCl (Fig. S20†). However, under intraoocyte ionic conditions the population of a hybrid GQ (−61.23 ppm) was found to be discernibly higher compared to that of the parallel form (−60.55 ppm), which was the major conformer in Tris buffer containing 100 mM KCl (compare with Fig. 6A). Interestingly, in both lysate and egg extract the signal corresponding to the hybrid GQ population dominated as the signal for a parallel GQ structure almost disappeared. Based on these data it appears that a hybrid topology of the EGFR ON is supported in a cellular environment. The observed signal broadening in the egg extract is not surprising as inhomogeneity and crowding can reduce the rate of tumbling and augment the relaxation process.[84–86] To ascertain if the ^{19}F signals obtained in lysate are associated with the GQ structures formed by the intact ON, the lysate sample was subjected to HPLC and mass analyses. A comparison of HPLC chromatograms of lysate, lysate plus ON **4** (after recording NMR), ON **4** and modified nucleoside analog clearly indicates that ON **4** is stable in the lysate and did not degrade during the NMR acquisition process (Fig. S21†). Furthermore, mass analysis of ON **4** isolated from the lysate sample confirmed the identity of the ON (Fig. S22†). It is important to note that the ^{1}H NMR spectrum of ON **4** gives basic information on GQ formation till the lysate level, but it fails to provide information on the egg extract due to severe line broadening. This is a commonly observed effect and a limitation of ^{1}H NMR use in the cell analysis.

GQ formation stalls the processivity of DNA polymerase

Noncanonical DNA structures can cause genomic instability by stalling the DNA replication process.[87] To study the effect of GQs, we designed three templates (T_1–T_3) composed of a wild-type EGFR G-rich sequence, a mutated EGFR sequence or a random non-GQ forming sequence downstream of the polymerization start site (Table 4). A 5′-FAM-labeled primer P_1 was annealed with templates T_1–T_3 in Tris·HCl buffer containing 100 mM KCl, and the replication reaction was performed using *Taq* DNA polymerase at 37 °C. At different time intervals, individual reaction samples were quenched by adding denaturing gel loading buffer and flashed cooled on a dry ice bath. Reaction products were resolved by PAGE under denaturing conditions and imaged using a fluorescence scanner. Reactions with T_3 majorly produced the full-length product in 1 min as a result of the absence of a GQ-forming motif in the template (Fig. 10A, blue bars and Fig. S23C,† gel image). On the other hand, T_1 containing the wild-type GQ-forming sequence stalled the polymerase activity and produced largely truncated primer-extended products just above the primer band along with minor amounts of the full-length product (Fig. 10A black bars and Fig. S23A,† gel image). At a reaction time of 5 min, it produced nearly 40% of the full-length product as compared to 60% product in the presence of a control template T_3. A mutated EGFR template T_2 also yielded stalled products similar to template T_1 (Fig. 10A, red bars and Fig. S23B,† gel image). While these results indicate that GQ formation hampers the processivity of *Taq* polymerase, mutations in the loop hairpin of the GQ has only a minor impact on the polymerization process. This is because mutations in the hairpin *per se* only destabilized the GQs by ~4 °C (Table S2†) and do not fully hamper the formation of

Table 4 DNA ON sequences used in Taq polymerase stop assay

ON	5′–sequence—3′
P_1	FAM-GGAGCTCAGCCTTCACTGC
T_1	GGGGACCGGGTCCAGAGGGGCAGTGCTGGGCGGCGCAGTGAAGGCTGAGCTCC
T_2[a]	GGGGACCGGGTCCAGAGGGTTAGTGCTGGGCGGCGCAGTGAAGGCTGAGCTCC
T_3	TCCTAACCCTAACTCTAACTCTAACGGCGCAGTGAAGGCTGAGCTCC

[a] T represents mutation points.

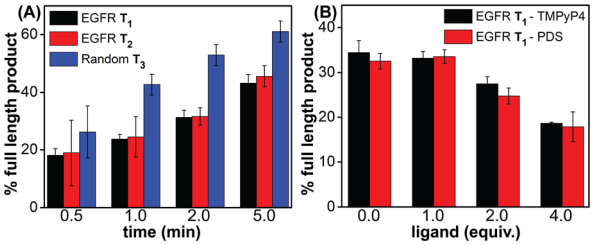


Fig. 10 Percentage of the full-length product obtained from *Taq* polymerase reactions. (A) Reactions performed using templates T_1–T_3 at different time points. (B) Reactions performed using T_1 with increasing concentrations of ligands (TMPyP4 and PDS) at 2 min. For gel images, see Fig. S23B and S24.†

respective GQs of the EGFR as proved by our CD, fluorescence and ^{19}F NMR studies (*vide supra*).

Next, to evaluate the effect of GQ–ligand interaction on the polymerase activity, stop assay was performed in the presence of ligands TMPyP4 and PDS. Since the primer extension with a wild-type EGFR template produced reasonable amounts of the full-length product in 2 min (~35%), reactions with increasing concentration of ligands were performed for 2 min at 37 °C (Fig. S24†). Upon increasing the concentration of both the ligands, the intensity of the full-length product reduced significantly with a concomitant increase in the stalled products. At 4 equivalents of the ligand, a nearly two-fold decrease in the efficiency of formation of the full-length product was observed (Fig. 10B and S24†). The observed extent of inhibition of the polymerase activity as a result of ligand binding to the GQs is reasonable in the case of the EGFR because regulated EGRF expression is very important for normal cell growth, differentiation and proliferation. Hence, instead of totally abrogating the expression of the EGRF gene, it is more useful if the overexpression of this gene is suppressed so that its normal cellular functions are not affected. In addition, EGRF expression can be more specifically controlled by designing ligands that target hybrid GQ-hairpin conformations adopted by the promoter region.

Conclusions

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

Data availability

The datasets supporting this article are available in the ESI.†

Author contributions

S. G. S. designed and supervised this project. S. Y. K. designed and performed the experiments. S. S. and P. I. P. built computational models of GQ structures. S. M. and J. K. isolated the frog eggs and helped in performing NMR experiments in egg lysate and extract. S. Y. K. and S. G. S. analysed the results and wrote the manuscript in consultation with all the authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

S. Y. K. acknowledges IISER Pune and India Alliance for a graduate research fellowship. S. S. acknowledges the Prime Minister's Research Fellows for the PhD fellowship. We thank SpaceTime HPC, IIT Bombay for computing facilities. This work was supported by India Alliance (IA/S/16/1/502360) and SERB (CRG/2022/000284) grants to S. G. S.

References

1 A. Ullrich, L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield and P. H. Seeburg, *Nature*, 1984, **309**, 418–425.
2 P. Blume-Jensen and T. Hunter, *Nature*, 2001, **411**, 355–365.
3 S. Sigismund, D. Avanzato and L. Lanzetti, *Mol. Oncol.*, 2018, **12**, 3–20.
4 N. Normanno, A. D. Luca, C. Bianco, L. Strizzi, M. Mancino, M. R. Maiello, A. Carotenuto, G. D. Feo, F. Caponigro and D. S. Salomon, *Gene*, 2006, **366**, 2–16.

5 J.-Y. Han, K. Park, S.-W. Kim, D. H. Lee, H. Y. Kim, H. T. Kim, M. J. Ahn, T. Yun, J. S. Ahn, C. Suh, J.-S. Lee, S. J. Yoon, J. H. Han, J. W. Lee, S. J. Jo and J. S. Lee, *J. Clin. Oncol.*, 2012, **30**, 1122–1128.
6 J. R. Goffin and K. Zbuk, *Clin. Ther.*, 2013, **35**, 1282–1303.
7 A. J. Cooper, L. V. Sequist and J. J. Lin, *Nat. Rev. Clin. Oncol.*, 2022, **19**, 499–514.
8 A. N. Hata, M. J. Niederst, H. L. Archibald, M. Gomez-Caraballo, F. M. Siddiqui, H. E. Mulvey, Y. E. Maruvka, F. Ji, H. E. C. Bhang, V. K. Radhakrishna, G. Siravegna, H. Hu, S. Raoof, E. Lockerman, A. Kalsy, D. Lee, C. L. Keating, D. A. Ruddy, L. J. Damon, A. S. Crystal, C. Costa, Z. Piotrowska, A. Bardelli, A. J. Iafrate, R. I. Sadreyev, F. Stegmeier, G. Getz, L. V. Sequist, A. C. Faber and J. A. Engelman, *Nat. Med.*, 2016, **22**, 262–269.
9 J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893–5909.
10 A. Jain, G. Wang and K. M. Vasquez, *Biochimie*, 2008, **90**, 1117–1130.
11 H. A. Assi, M. Garavís, C. González and M. J. Damha, *Nucleic Acids Res.*, 2018, **46**, 8038–8056.
12 G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867–5892.
13 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770–780.
14 S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 275–283.
15 D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627–8637.
16 D. Varshney, J. Spiegel, K. Zyner, D. Tannahill and S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 459–474.
17 H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2021, **49**, 7839–7855.
18 C. L. Grand, T. J. Powell, R. B. Nagle, D. J. Bearss, D. Tye, M. Gleason-Guzman and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A*, 2004, **101**, 6140–6145.
19 A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A*, 2002, **99**, 11593–11598.
20 J.-H. Tan, T.-M. Ou, J.-Q. Hou, Y.-J. Lu, S.-L. Huang, H.-B. Luo, J.-Y. Wu, Z.-S. Huang, K.-Y. Wong and L.-Q. Gu, *J. Med. Chem.*, 2009, **52**, 2825–2835.
21 K. I. E. McLuckie, Z. A. E. Waller, D. A. Sanders, D. Alves, R. Rodriguez, J. Dash, G. J. McKenzie, A. R. Venkitaraman and S. Balasubramanian, *J. Am. Chem. Soc.*, 2011, **133**, 2658–2663.
22 M. Micco, G. W. Collie, A. G. Dale, S. A. Ohnmacht, I. Pazitna, M. Gunaratnam, A. P. Reszka and S. Neidle, *J. Med. Chem.*, 2013, **56**, 2959–2974.
23 M. Debnath, S. Ghosh, D. Panda, I. Bessi, H. Schwalbe, K. Bhattacharyya and J. Dash, *Chem. Sci.*, 2016, **7**, 3279–3285.
24 V. Dhamodharan and P. I. Pradeepkumar, *ACS Chem. Biol.*, 2019, **14**, 2102–2114.
25 S. Asamitsu, S. Obata, Z. Yu, T. Bando and H. Sugiyama, *Molecules*, 2019, **24**, 429–457.
26 B. R. Vummidi, J. Alzeer and N. W. Luedtke, *ChemBioChem*, 2013, **14**, 540–558.
27 S. Wu, L. Wang, N. Zhang, Y. Liu, W. Zheng, A. Chang, F. Wang, S. Li and D. Shangguan, *Chem.–Eur. J.*, 2016, **22**, 6037–6047.
28 V. Grande, C.-A. Shen, M. Deiana, M. Dudek, J. Olesiak-Banska, K. Matczyszyn and F. Würthner, *Chem. Sci.*, 2018, **9**, 8375–8382.
29 Y. Ma, K. Iida, S. Sasaki, T. Hirokawa, B. Heddi, A. T. Phan and K. Nagasawa, *Molecules*, 2019, **24**, 263–277.
30 J.-H. Yuan, W. Shao, S.-B. Chen, Z.-S. Huang and J.-H. Tan, *Biochem. Biophys. Res. Commun.*, 2020, **531**, 18–24.
31 T. Vo, S. Oxenford, R. Angell, C. Marchetti, S. A. Ohnmacht, W. D. Wilson and S. Neidle, *ACS Med. Chem. Lett.*, 2020, **11**, 991–999.
32 S. Kumar, S. P. P. Pany, S. Sudhakar, S. B. Singh, C. S. Todankar and P. I. Pradeepkumar, *Biochemistry*, 2022, **61**, 2546–2559.
33 K. W. Lim, P. Jenjaroenpun, Z. J. Low, Z. J. Khong, Y. S. Ng, V. A. Kuznetsov and A. T. Phan, *Nucleic Acids Res.*, 2015, **43**, 5630–5646.
34 Y. M. Vianney and K. Weisz, *Nucleic Acids Res.*, 2022, **50**, 11948–11964.
35 T. Q. N. Nguyen, K. W. Lim and A. T. Phan, *Sci. Rep.*, 2017, **7**, 11969–11975.
36 S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando and H. Sugiyama, *Chem.–Eur. J.*, 2018, **24**, 4428–4435.
37 M. Yang, S. Carter, S. Parmar, D. D. Bume, D. R. Calabrese, X. Liang, K. Yazdani, M. Xu, Z. Liu, C. J. Thiele and J. S. Schneekloth Jr, *Nucleic Acids Res.*, 2021, **49**, 7856–7869.
38 M. L. Greco, A. Kotar, R. Rigo, C. Cristofari, J. Plavec and C. Sissi, *Nucleic Acids Res.*, 2017, **45**, 10132–10142.
39 J. T. Grün and H. Schwalbe, *Biopolymers*, 2022, **113**, e23477–e23491.
40 P. A. Rachwal, T. Brown and K. R. Fox, *Biochemistry*, 2007, **46**, 3036–3044.
41 H. You, J. Wu, F. Shao and J. Yan, *J. Am. Chem. Soc.*, 2015, **137**, 2424–2427.
42 J. Marquevielle, C. Robert, O. Lagrabette, M. Wahid, A. Bourdoncle, L. E. Xodo, J.-L. Mergny and G. F. Salgado, *Nucleic Acids Res.*, 2020, **48**, 9336–9345.
43 R. C. Monsen, L. DeLeeuw, W. L. Dean, R. D. Gray, T. M. Sabo, S. Chakravarthy, J. B. Chaires and J. O. Trent, *Nucleic Acids Res.*, 2020, **48**, 5720–5734.
44 M. Cheng, Y. Cheng, J. Hao, G. Jia, J. Zhou, J.-L. Mergny and C. Li, *Nucleic Acids Res.*, 2018, **46**, 9264–9275.
45 L. Lacroix, A. Séosse and J.-L. Mergny, *Nucleic Acids Res.*, 2011, **39**, e21–e31.
46 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963–9970.
47 G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
48 F. Doria, M. Nadai, M. Zuffo, R. Perrone, M. Freccero and S. N. Richter, *Chem. Commun.*, 2017, **53**, 2268–2271.
49 Y. V. Suseela, N. Narayanaswamy, S. Pratihar and T. Govindaraju, *Chem. Soc. Rev.*, 2018, **47**, 1098–1131.
50 G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182–186.

51 A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh Jr, D. Sen and P. M. Lansdrop, *Nucleic Acids Res.*, 2014, **42**, 860–869.
52 A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J.-M. Perrier-Cornet, J. M. Y. Wong and D. Monchaud, *J. Am. Chem. Soc.*, 2015, **137**, 8521–8525.
53 A. Shivalingam, M. A. Izquierdo, A. L. Marois, A. Vyšniauskas, K. Suhling, M. K. Kuimova and R. Vilar, *Nat. Commun.*, 2015, **6**, 8178–8187.
54 Z. Yu, J. D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu and H. Mao, *J. Am. Chem. Soc.*, 2009, **131**, 1876–1882.
55 H. You, X. Zeng, Y. Xu, C. J. Lim, A. K. Efremov, A. T. Phan and J. Yan, *Nucleic Acids Res.*, 2014, **42**, 8789–8795.
56 L. Ying, J. J. Green, H. Li, D. Klenerman and S. Balasubramanian, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14629–14634.
57 S. L. Noer, S. Preus, D. Gudnason, M. Aznauryan, J.-L. Mergny and V. Birkedal, *Nucleic Acids Res.*, 2016, **44**, 464–471.
58 S. Manna, D. Sarkar and S. G. Srivatsan, *J. Am. Chem. Soc.*, 2018, **140**, 12622–12633.
59 S. Y. Khatik and S. G. Srivatsan, *Bioconjugate Chem.*, 2022, **33**, 1515–1526.
60 H. Chen, S. Viel, F. Ziarelli and L. Peng, *Chem. Soc. Rev.*, 2013, **42**, 7971–7982.
61 L. B. T. Pham, A. Costantino, L. Barbieri, V. Calderone, E. Luchinat and L. Banci, *J. Am. Chem. Soc.*, 2023, **145**, 1389–1399.
62 M. Himmelstoß, K. Erharter, E. Renard, E. Ennifar, C. Kreutz and R. Micura, *Chem. Sci.*, 2020, **11**, 11322–11330.
63 Q. Li, J. Chen, M. Trajkovski, Y. Zhou, C. Fan, K. Lu, P. Tang, X. Su, J. Plavec, Z. Xi and C. Zhou, *J. Am. Chem. Soc.*, 2020, **142**, 4739–4748.
64 H.-L. Bao, H.-S. Liu and Y. Xu, *Nucleic Acids Res.*, 2019, **47**, 4940–4947.
65 M. R. Baranowski, M. Warminski, J. Jemielity and J. Kowalska, *Nucleic Acids Res.*, 2020, **48**, 8209–8224.
66 J. Riedl, R. Pohl, L. Rulíšek and M. Hocek, *J. Org. Chem.*, 2012, **77**, 1026–1044.
67 T. Sakamoto, D. Hasegawa and K. Fujimoto, *Chem. Commun.*, 2015, **51**, 8749–8752.
68 N. B. Barhate, R. N. Barhate, P. Cekan, G. Drobny and S. Th. Sigurdsson, *Org. Lett.*, 2008, **10**, 2745–2747.
69 A. T. Phan, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2006, **34**, 5715–5719.
70 J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.
71 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 2723–2735.
72 J.-L. Mergny, J. Li, L. Lacroix, S. Amrane and J. B. Chaires, *Nucleic Acids Res.*, 2005, **33**, e138–e144.
73 L. Evans, A. Kotar, M. Valentini, A. Filloux, S. Jamshidi, J. Plavec, K. M. Rahman and R. Vilar, *RSC Chem. Biol.*, 2023, **4**, 94–100.
74 J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713–1725.
75 S. O. Kelley and J. K. Barton, *Science*, 1999, **283**, 375–381.
76 S. Doose, H. Neuweiler and M. Sauer, *ChemPhysChem*, 2009, **10**, 1389–1398.
77 K. W. Lim and A. T. Phan, *Angew. Chem., Int. Ed.*, 2013, **52**, 8566–8569.
78 K. W. Lim, Z. J. Khong and A. T. Phan, *Biochemistry*, 2014, **53**, 247–257.
79 D. D. Le, M. D. Antonio, L. K. M. Chan and S. Balasubramanian, *Chem. Commun.*, 2015, **51**, 8048–8050.
80 C. Wei, G. Jia, J. Yuan, Z. Feng and C. Li, *Biochemistry*, 2006, **45**, 6681–6691.
81 Y. Ma, K. Iida and K. Nagasawa, *Biochem. Biophys. Res. Commun.*, 2020, **531**, 3–17.
82 B. Machireddy, H.-J. Sullivan and C. Wu, *Molecules*, 2019, **24**, 1010–1035.
83 L.-Y. Liu, T.-Z. Ma, Y.-L. Zeng, W. Liu and Z.-W. Mao, *J. Am. Chem. Soc.*, 2022, **144**, 11878–11887.
84 R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res.*, 2011, **39**, 5768–5775.
85 H.-L. Bao, T. Ishizuka, T. Sakamoto, K. Fujimoto, T. Uechi, N. Kenmochi and Y. Xu, *Nucleic Acids Res.*, 2017, **45**, 5501–5511.
86 Y. Ye, X. Liu, Z. Zhang, Q. Wu, B. Jiang, L. Jiang, X. Zhang, M. Liu, G. J. Pielak and C. Li, *Chem.–Eur. J.*, 2013, **19**, 12705–12710.
87 L. K. Lerner and J. E. Sale, *Genes*, 2019, **10**, 95–119.

Electronic Supplementary Information (ESI)

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

Saddam Y. Khatik,[a] Sruthi Sudhakar,[b] Satyajit Mishra,[c] Jeet Kalia,[c,d] P.I. Pradeepkumar[b] and Seergazhi G. Srivatsan*[a]

[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.

[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.

1. Materials. 5-Trifluoromethyl benzofuran-modified nucleoside analog **1** and corresponding phosphoramidite substrate **2** for solid-phase oligonucleotide (ON) synthesis were synthesized as per our previously reported procedure.[S1] *N*-acetyl-protected dC, *N*-Benzoyl-protected dA, *N,N*-dimethylformamide-protected dG, and dT phosphoramidite substrates needed for solid-phase DNA synthesis were procured from ChemGenes. Solid supports required for DNA synthesis were purchased from Glen Research. All other reagents required for solid phase ON synthesis were obtained from Sigma-Aldrich. Control DNA ONs Telo1, TeloC, **6** and **7** were purchased from Integrated DNA Technology, and purified using 18% denaturing polyacrylamide gel electrophoresis (PAGE). All other reagents (BioUltra grade) for the preparation of buffer solutions were purchased from Sigma-Aldrich. Autoclaved water was used for preparing all buffer solutions, and in all biophysical studies.

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

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

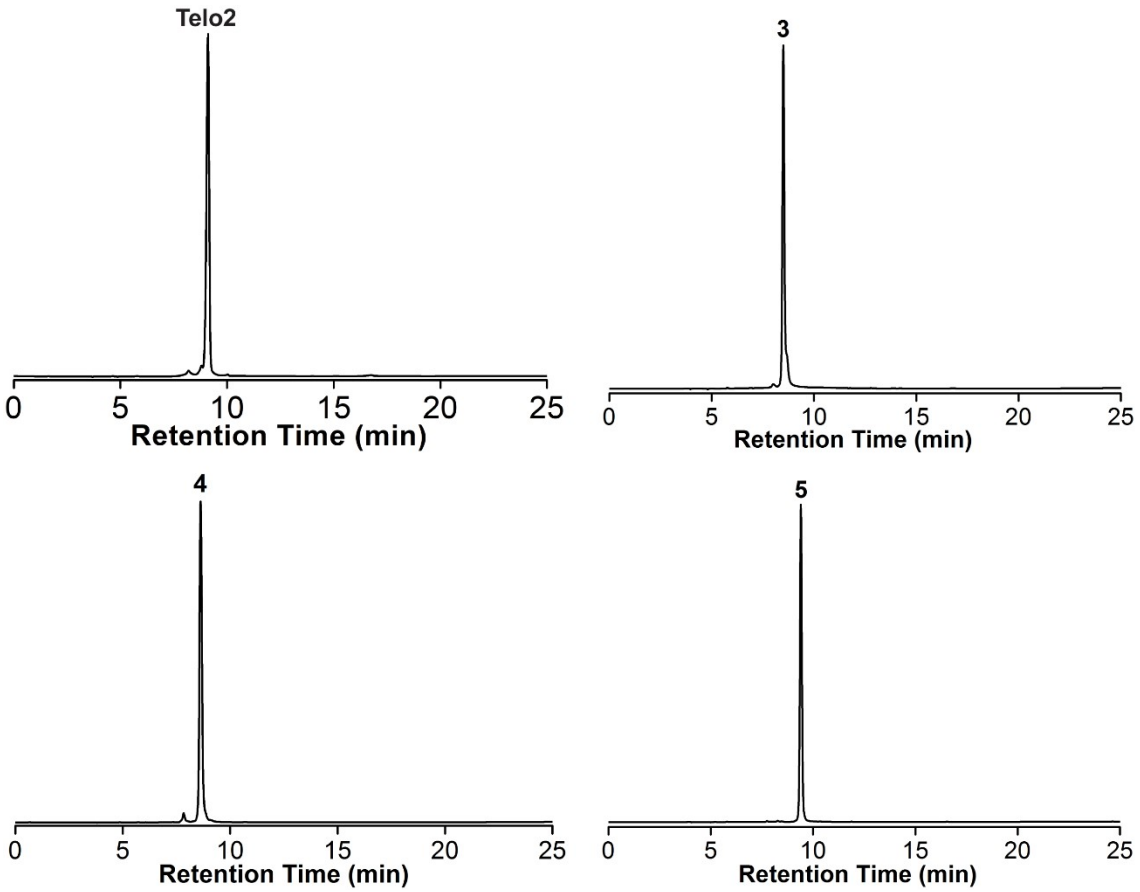


Fig. S1. RP-HPLC chromatograms of TFBF-modified Telo2 and EGFR ONs **3**–**5** analyzed at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).

4. Mass analysis of modified ON.

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

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

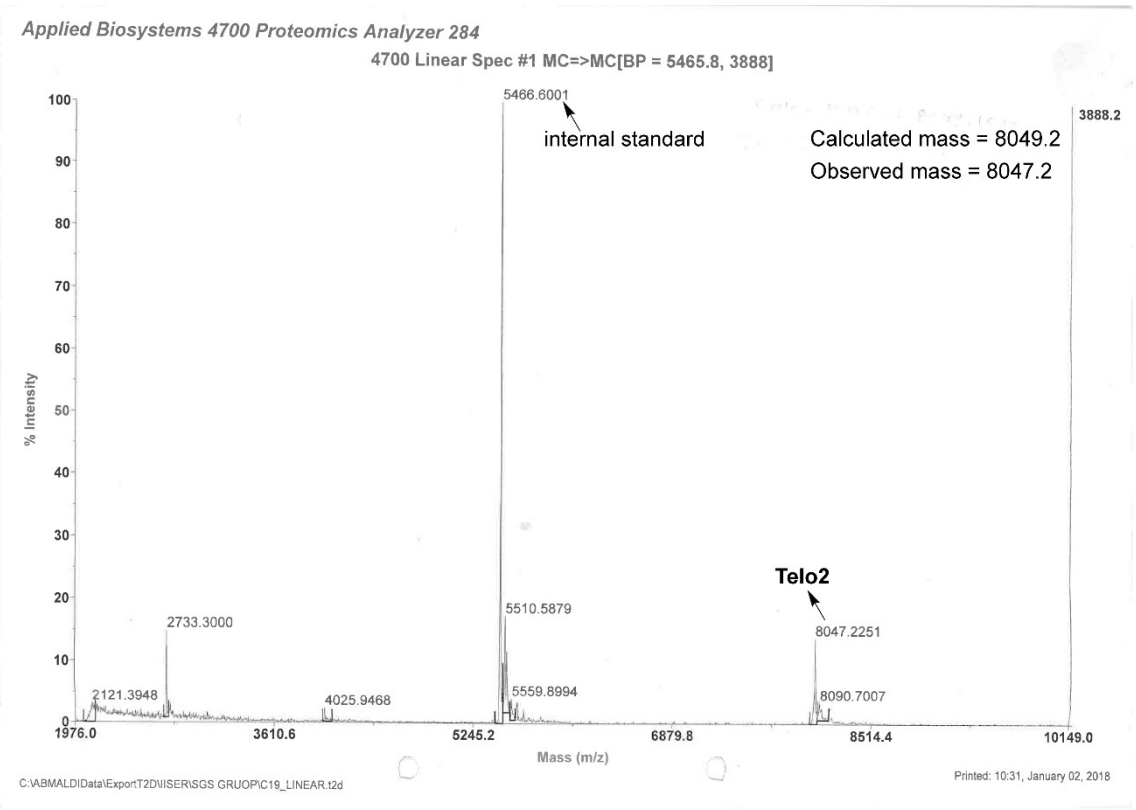
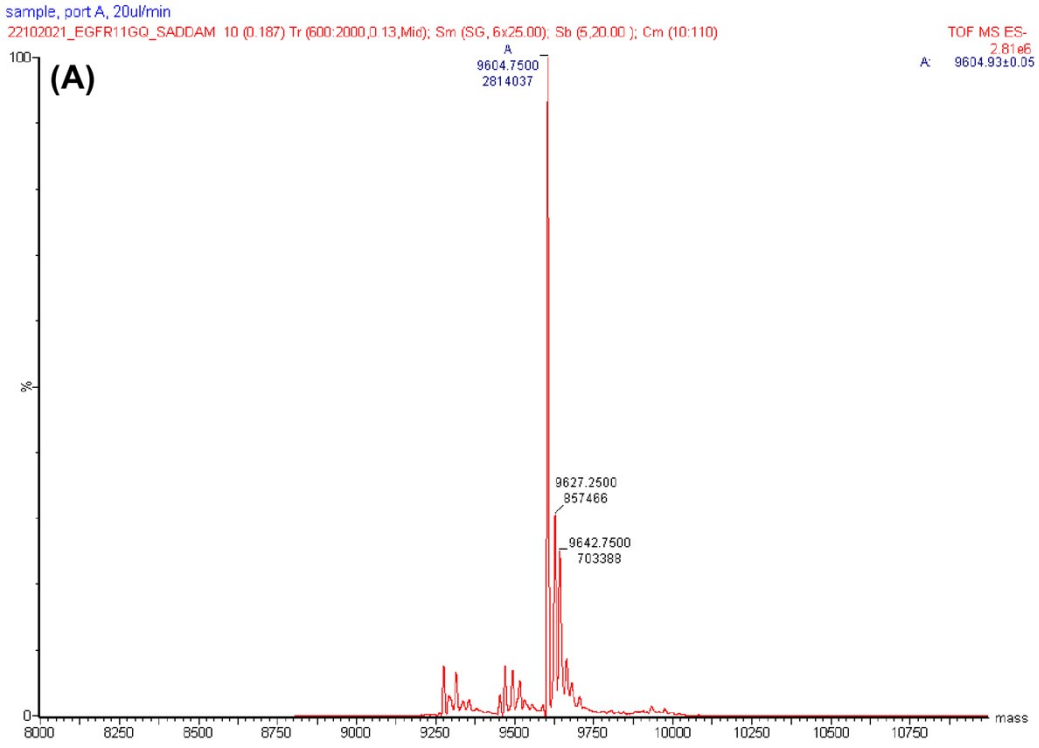
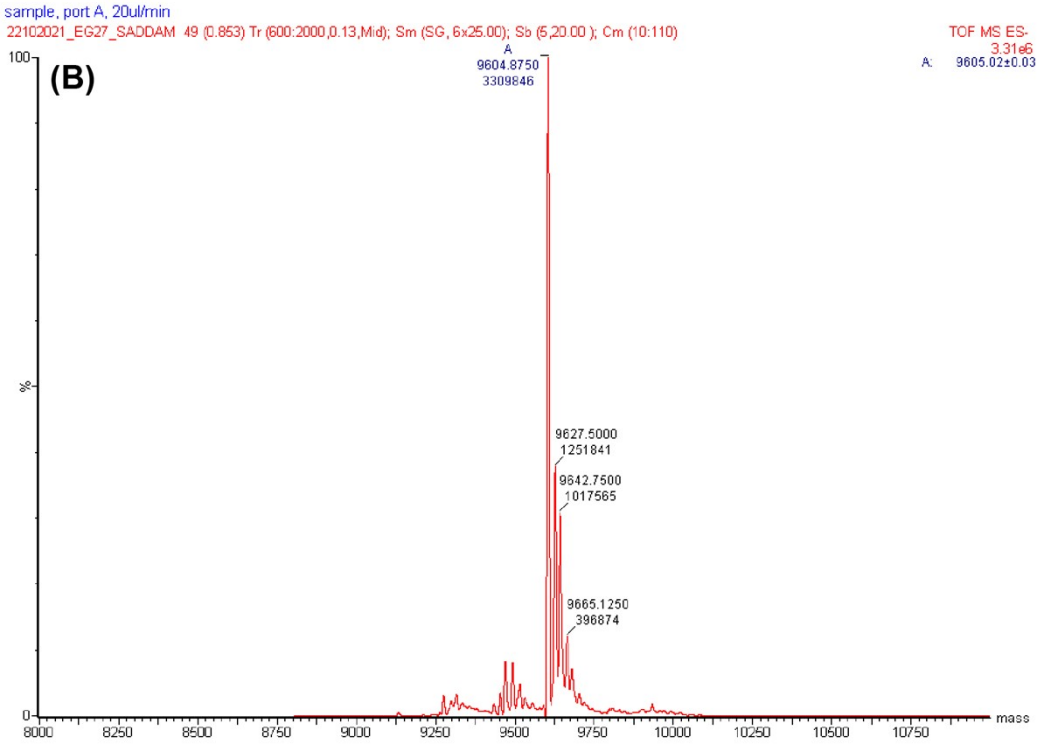


Fig. S2. MALDI-TOF MS spectrum of TFBF-modified Telo2. Internal DNA ON standard m/z of +1 and +2 ions are 5466.6 and 2733.3. See Section 4 for details.


sample, port A, 20ul/min
22102021_EGFR11GQ_SADDAM 10 (0.187) Tr (600:2000,0.13,Mid); Sm (SG, 6x25.00); Sb (5,20.00); Cm (10:110)
TOF MS ES-
2.81e6
A: 9604.93±0.05
(A)
A
9604.7500
2814037
9627.2500
857466
9642.7500
703388
%
mass
100
0
8000
8250
8500
8750
9000
9250
9500
9750
10000
10250
10500
10750


sample, port A, 20ul/min
22102021_EG27_SADDAM 49 (0.853) Tr (600:2000,0.13,Mid); Sm (SG, 6x25.00); Sb (5,20.00); Cm (10:110)
TOF MS ES-
3.31e6
A: 9605.02±0.03
(B)
A
9604.8750
3309846
9627.5000
1251841
9642.7500
1017565
9665.1250
396874
%
mass
100
0
8000
8250
8500
8750
9000
9250
9500
9750
10000
10250
10500
10750

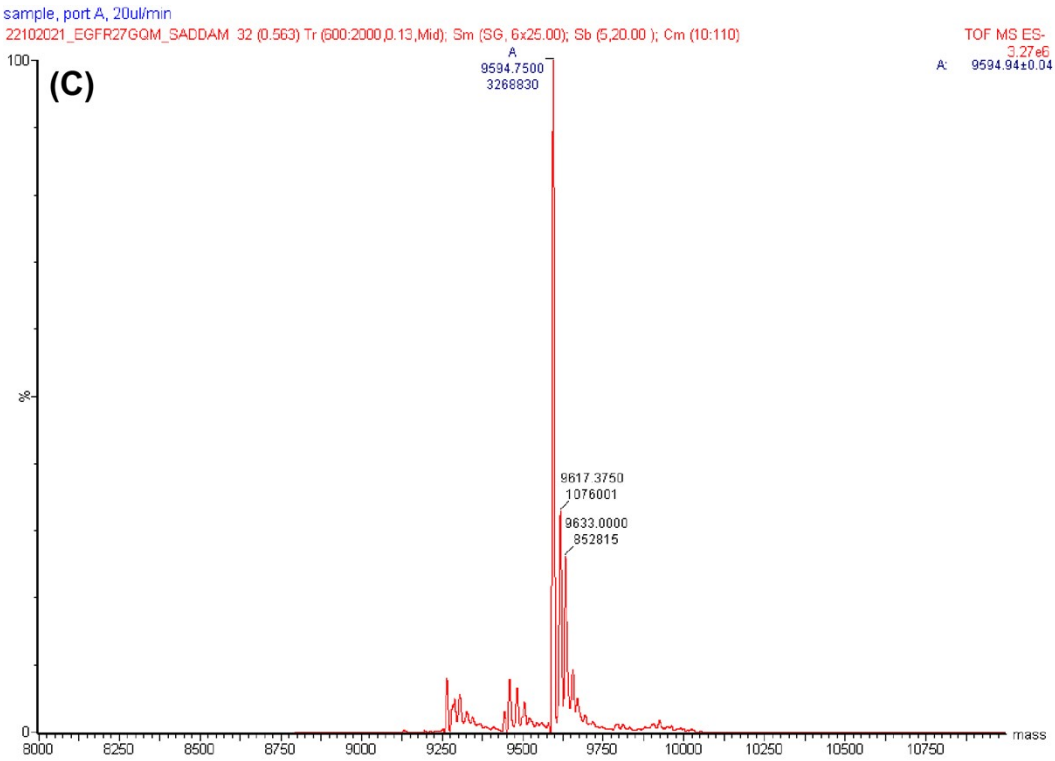


Fig. S3. ESI-MS spectra of TFBF-modified EGFR ONs (**A**) **3**, (**B**) **4**, (**C**) **5**.

Table S1. Molar absorptivity and mass of modified DNA ONs.

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

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

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

duplicate with an average of three scans for each sample. The spectrum of buffer in absence of ON was subtracted from all ON sample spectra.

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

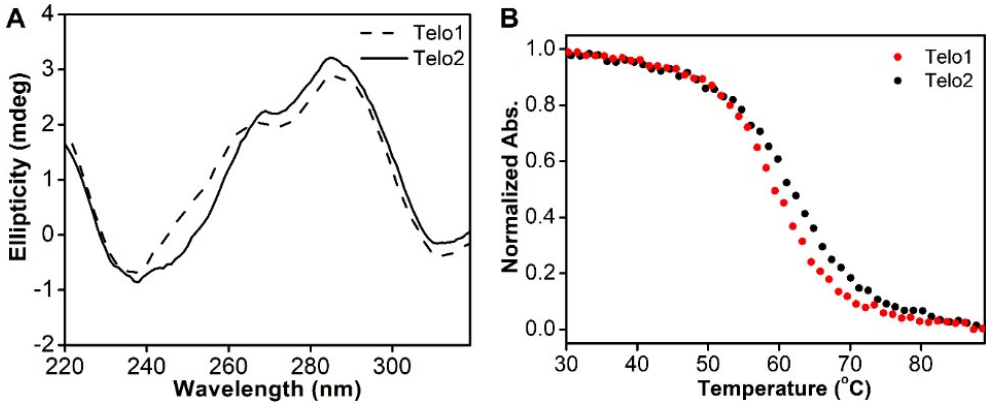


Fig. S4. (**A**) CD spectra of control Telo1 and modified Telo2 ONs (5 µM). (**B**) UV-thermal melting profile (at 295 nm) of Telo1 and Telo2 ONs (1 µM). See Section 5 and 6 for details.

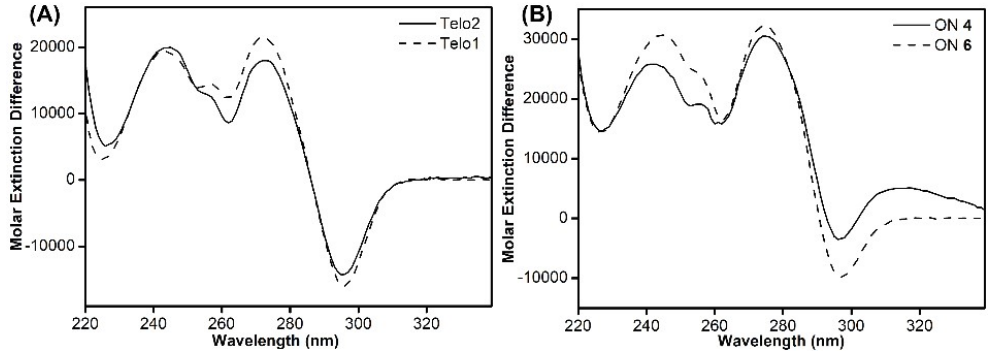


Fig. S5. Thermal difference spectrum (TDS) for Telo2 (**A**) and EGFR ON **4** (**B**), and corresponding unmodified control ON sequences. ONs samples (5 µM) were annealed in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and UV spectrum of ONs was recorded at 25 °C and 90 °C. The TDS was obtained by subtracting the UV absorption profile of the folded form from the unfolded form.

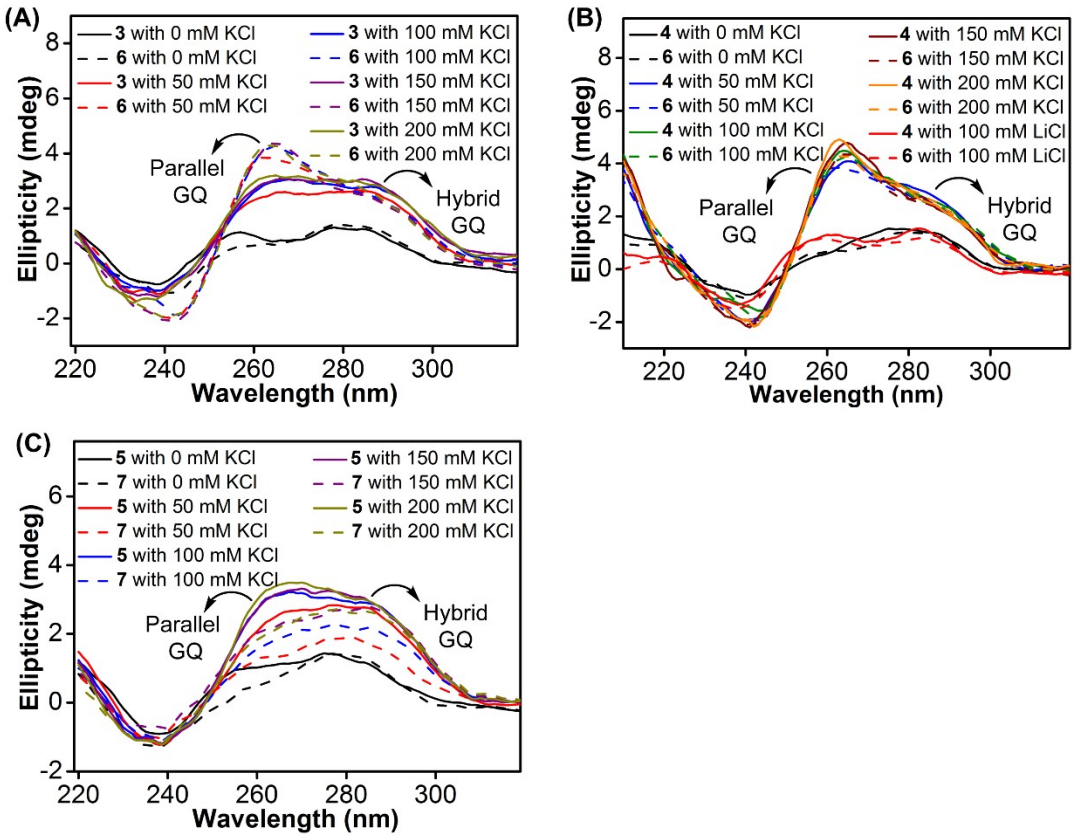


Fig. S6. (**A**, **B** and **C**) CD spectra (5 µM) of modified EGFR ONs **3**–**5** (solid lines) and their control ONs **6** and **7** (dashed lines) in different ionic conditions.

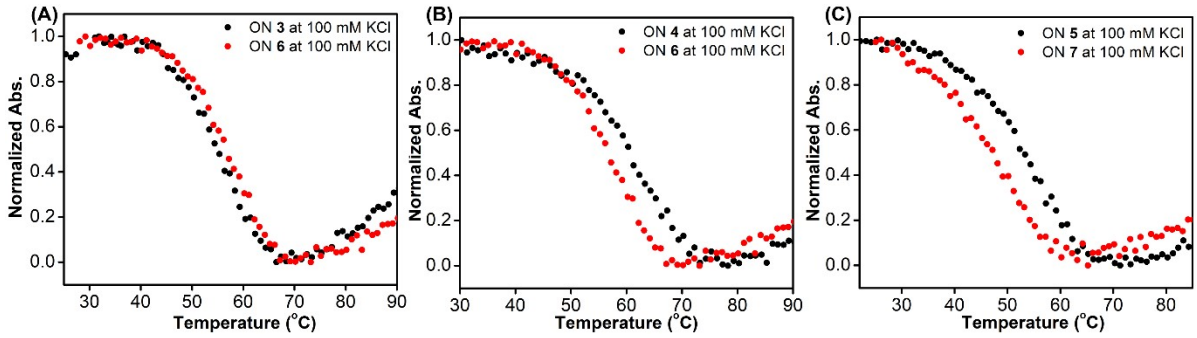


Fig. S7. UV-thermal melting profile (at 295 nm) of modified EGFR ONs **3**–**5** and corresponding control unmodified ONs **6** and **7** (1 µM).

Table S2. T_m values of modified Telo2 and EGFR ONs **3**–**5** and control unmodified Telo1 and EGFR ONs **6** and **7.**

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

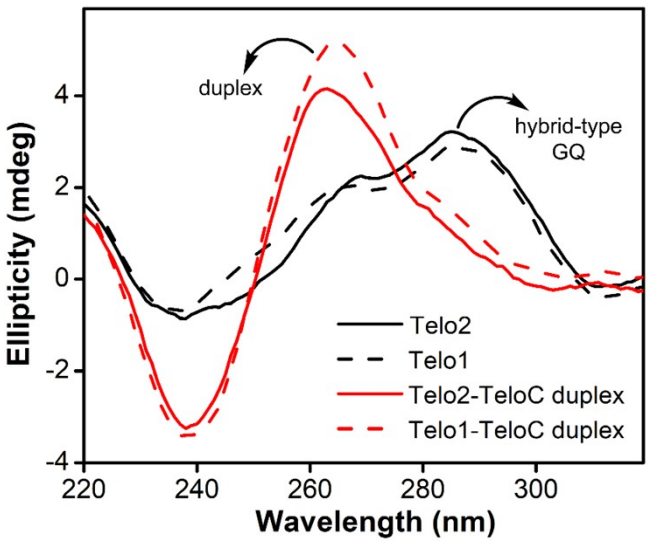


Fig. S8. A comparison of CD spectra of duplex and GQ structures formed by the telomeric repeat ONs (5 µM). See section 5 for experimental details.

7. Fluorescence and NMR studies of modified Telo2 ON.

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

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

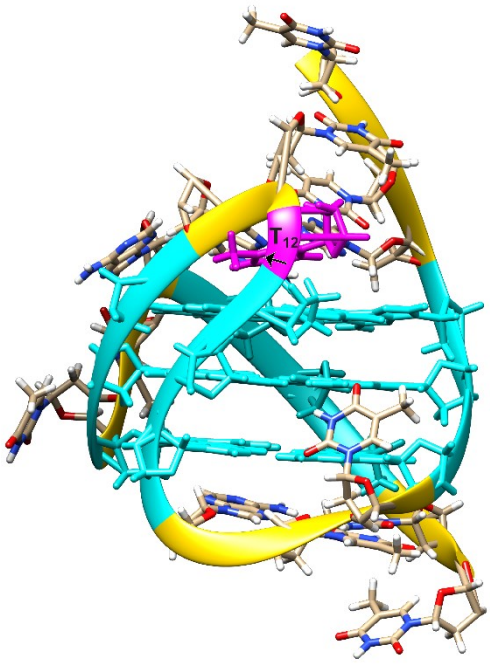


Fig. S9. NMR structure of a hybrid GQ topology of the human telomeric repeat (PDB: 2JPZ).[S2] Guanosines participating in the tetrad formation are shown in cyan. T_{12} residue is shown in magenta. Arrow shows the C5-postion of T_{12} where the TFBF heterocycle modification is placed in the modified Telo2 ON. TFBF modification does not affect the GQ structure. Hence, TFBF modification at C5-position is likely to be projected into the groove away from the G-tetrad. As a result of reduced stacking interaction, the GQ form shows higher fluorescence intensity. The figure was generated using UCSF Chimera version 1.15.

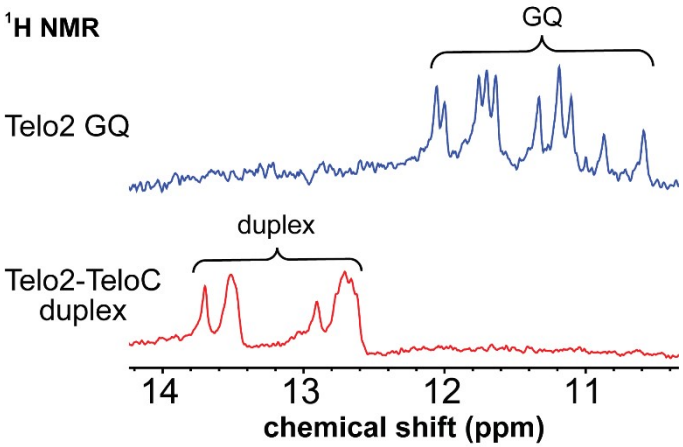


Fig. S10. ^{1}H NMR spectra of the modified telomeric ON forming GQ and duplex structures. See section 7 for details.

8. Fluorescence study of modified EGFR ONs.

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

Time-resolved fluorescence analysis of modified ONs. Samples of TFBF-dU modified ONs **4** and **5** (4 µM) in 10 mM Tris.HCl (pH 7.4) buffer were prepared and used for excited-state decay kinetics measurements. ONs samples were excited using a 340 nm LED source and the fluorescence decay was collected at respective emission maximum with an increase in KCl concentrations. The concentration of the KCl was increased by adding different aliquots of 3 M KCl into the ON sample and incubated for one hour before measurements. Fluorescence decay of the ON samples were deconvoluted using EZ software and fitted by an exponential decay with χ^2 value close to unity.

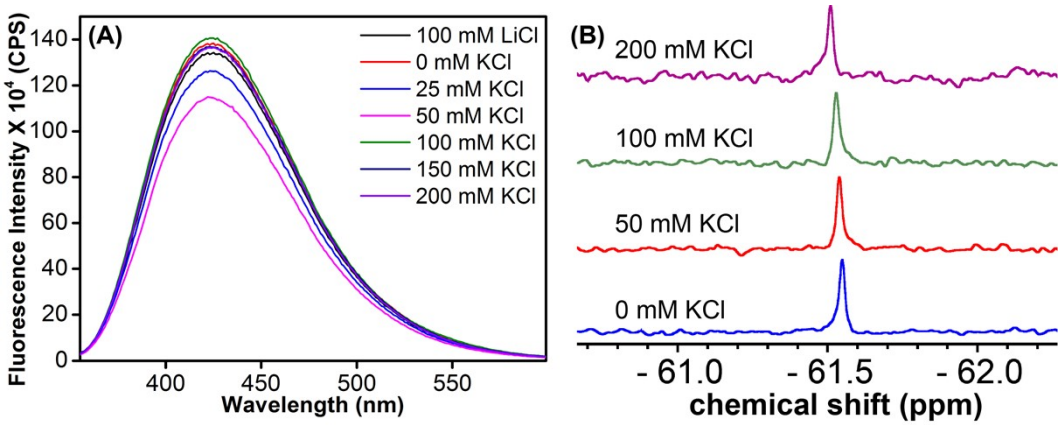


Fig. S11. (**A**) Fluorescence spectra of TFBF-dU (2 µM), and (**B**) ^{19}F-NMR spectra of TFBF-dU (10 µM) at different KCl concentrations. In the fluorescence study, samples were excited at 320 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively.

9. Computational analysis. In order to generate the model for the two architectures of the EGFR G-rich sequences, a combination of 3D-NuS[S3] webserver and manual editing has been employed. For generating the hybrid architecture, the loop sequences were initially added to the 3D-NuS web server, where the hybrid GQ structure of class Q17 was generated. The 5' terminal dG was excluded in both architectures. The web server-generated structure was extracted in PDB format. The hairpin domain of the structure lacked any base pairing interaction. Therefore, the sequence of the hairpin domain was separately modeled with the help of the DNA folding form of the Mfold web server,[S4] RNAalifold web server, and 3DNA[S5]. The hairpin domain from the 3D-NuS PDB structure was removed, and the newly generated hairpin was placed with the help of PyMOL software. Further, the coordinates of the GQ and the newly generated hairpin domain were combined manually to generate a new PDB file. Similarly, for the parallel GQ, the primary

architecture was generated from the 3D-NuS web server, after which the hairpin domain was placed separately. The hairpin domain, as well as dT26 (dT27 of ON **4**), were manually placed with the help of PyMOL, and residue numbers were corrected and then saved in the PDB format.

Both the PDB structures generated were added into the Tleap module of AmberTools 21, and K^+ counter ions were added to generate the coordinate and topology files. The central K^+ ions were also added in both cases. The OL15[S6] force field was used for DNA and TIP3P water model for water and counter ions. The structures were subjected to a 100000-step minimization using the implicit solvent model in AMBER 18.[S7] The final structures after the minimization was again loaded to the tleap module. These structures were enclosed in a rectangular water box of 10 Å, and the coordinate and topology files were generated. The systems were then subjected to 10000 steps of minimization by the steepest descent method with a restraint of 2.0 kcal/mol $Å^2$ on the DNA and central K^+ ions. The minimization followed 100 ps of heating and 100 ps of density equilibration with restraints of 50 kcal/mol $Å^2$ and 2.0 kcal/mol $Å^2$. The systems were then equilibrated for 800 ps in NPT ensemble, and unrestrained MD simulation was performed using NPT ensemble for 400 ns in GPU accelerated version of PMEMD[S8-S10] in AMBER 18. SHAKE algorithm was applied to subject the hydrogen atoms to bond length constraints. Langevin[S11] thermostat with a collision frequency of 2 ps^{-1} was used to maintain the temperature at 300 K, and Berendsen[S12] barostat with a relaxation time of 2 ps was used to maintain pressure. The trajectories were visualized using UCSF Chimera[S13], and images were rendered using PyMOL. The analysis was carried out using the CPPTRAJ[S14] module of AmberTools.

The 400 ns trajectories were clustered into 5 ensembles using the hierarchical agglomerative clustering algorithm. The hybrid GQ has one major cluster which existed for ~60 % of the simulation time (Fig. S12A†), and the parallel GQ has one major cluster which existed for ~50 % (Fig. S12B†) of the simulation time. Since the 5′ terminal dG was excluded in both architectures, the numbering of the nucleotides in the ON sequence has been renumbered (Fig. 5). In the hybrid GQ, the hairpin domain maintains the two base pairs (dG19:dC25 and dC20:dG24) intact during the entire simulation, while the same is lost towards the end of the simulation in the parallel GQ. dT26 (dT27 in ON **4** and **6**) stacks with both dC25 and dG3 maintaining an average distance of 3.80 ± 0.24 Å and 4.10 ± 0.32 Å respectively in the hybrid GQ. In the case of parallel GQ, dT26 partially stacks over dG16, maintaining an average distance of 4.49 ± 1.17 Å. The center of mass of the heavy atoms of the bases (excluding sugar) was considered for the distance calculation.

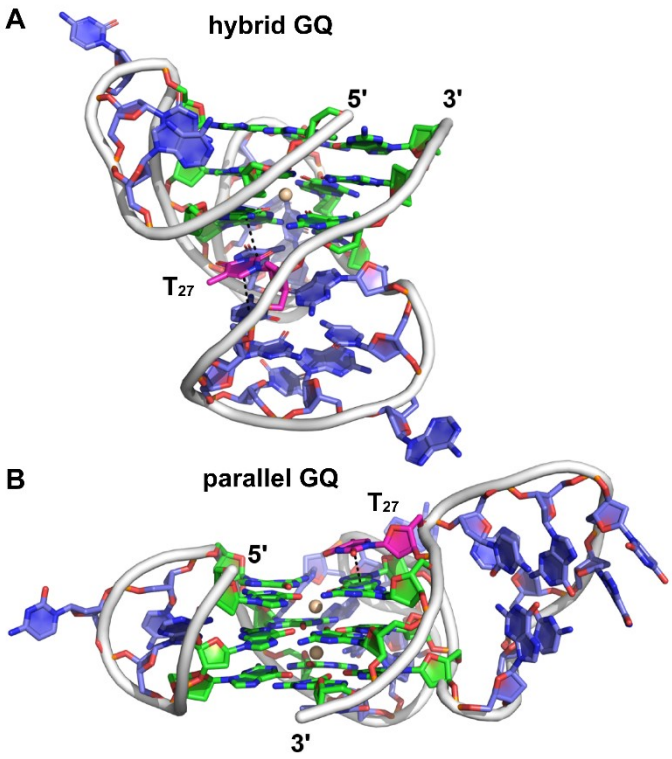


Fig. S12. Representative structure of the major cluster of (**A**) hybrid and (**B**) parallel GQ topologies. Carbon atoms of G-tetrads are represented in green, T_{27} is shown in magenta, and all other nucleotides are represented in purple. Nitrogen atoms are represented in blue, oxygen in red, and phosphate in orange. Hydrogen atoms are omitted for clarity. (**A**) In the hybrid form, T_{27} is stacked between the hairpin domain and a G-tetrad. (**B**) In the parallel form, T_{27} is partially stacked over the top G-tetrad. See Section 9 for details.

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

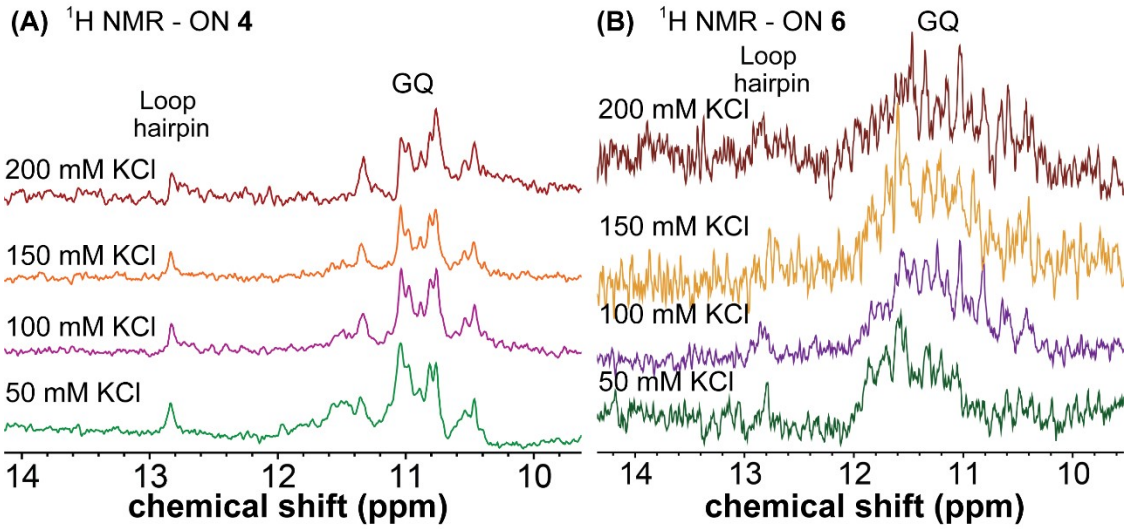


Fig. S13. ^{1}H NMR of TFBF-dU modified ON **4** and unmodified ON **6** of the native sequence.

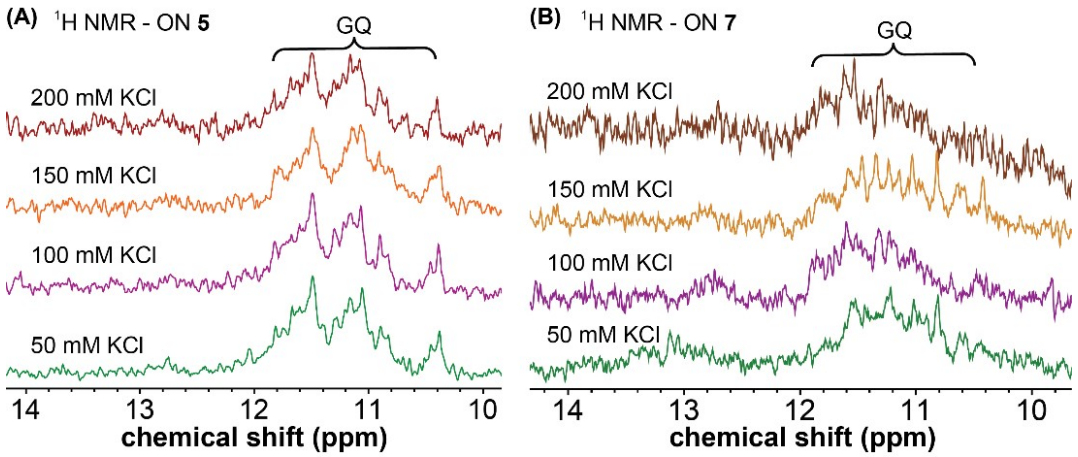


Fig. S14. ^{1}H NMR of TFBF-dU modified ON **5** and unmodified ON **7** of the mutated sequence.

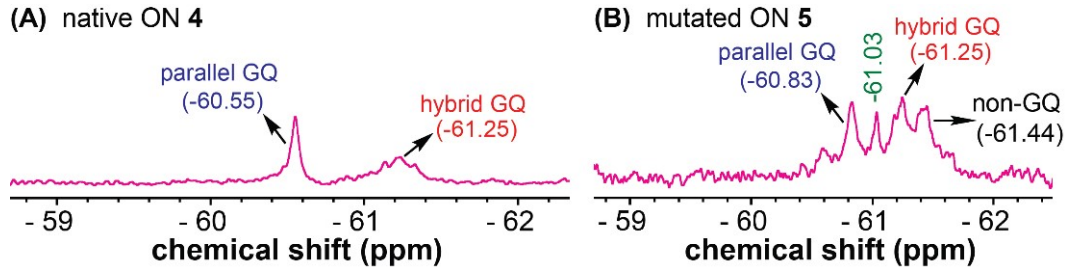


Fig. S15. ^{19}F NMR spectra of the native ON **4** and mutated ON **5** at 100 mM KCl.

11. GQ-ligand interaction by fluorescence. A series of samples of modified ON **4** (0.5 µM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ligands (5 nM to 5 µM) were prepared. Samples were incubated at 4 °C for 1 h. Fluorescence emission spectra were recorded by exciting the samples at 330 nM with excitation and emission wavelength slit widths of 6 nm and 7 nm, respectively. The fluorescence of samples was recorded in triplicate reading at 25 °C. Further, the fluorescence of the blank sample containing control ON **6** and respective ligand concentration was subtracted from the individual sample reading. The apparent K_d values of ligands were determined by plotting the normalized fluorescence intensity (F_N) *vs* ligand concentration. The graph was prepared using OriginPro 8.5 software.[S15]

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

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

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

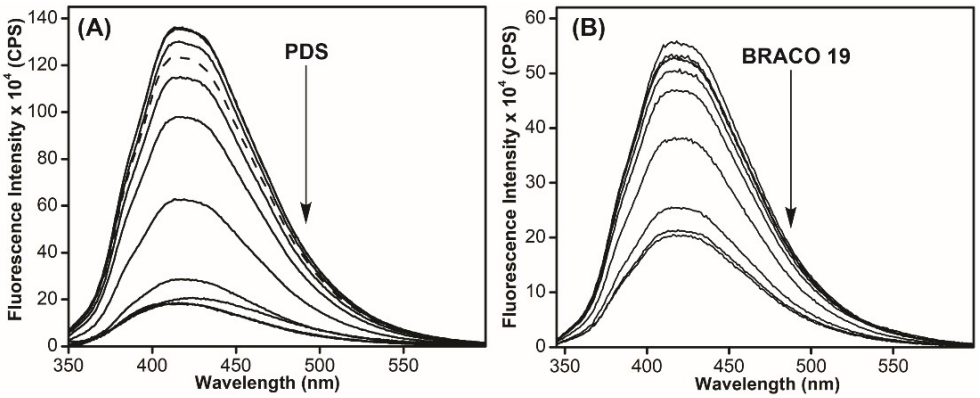


Fig. S16. Fluorescence titration of ON **4** (0.5 µM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of (**A**) PDS, and (**B**) BRACO-19. Fluorescence spectra of the modified ON **4** in the absence of ligands represented by a dashed line. Samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively.

12. GQ-ligand interaction by UV absorption.

The samples of TMPyP4 (2 µM) were prepared in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of pre-annealed ON **4** (5 nM to 2 µM). Samples were incubated at 4 °C for 1 h and UV absorption spectra were acquired at 25 °C. Titration was performed until the wavelength and intensity of the absorption band of TMPyP4 remained unchanged.[S16,S17] UV experiment was performed in duplicate. Binding constant (K_d) was obtained from the plot of normalized absorbance (A_N) at 422 nm vs concentration of ON **4**. The graph was prepared using OriginPro 8.5 software.

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

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

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

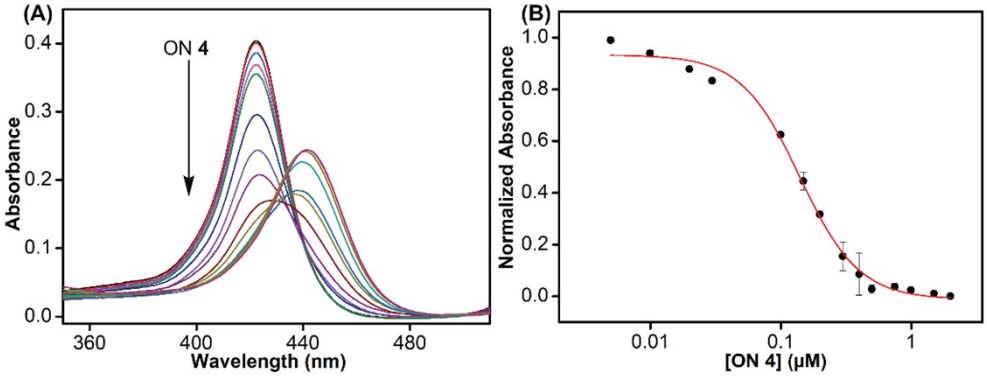


Fig. S17. (**A**) UV-vis absorption spectra. Titration of TMPyP4 (2 μM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ON **4** (5 nM to 2 μM). (**B**) Curve fit plotted for normalized absorbance of TMPyP4 at 422 nm with increasing concentrations of ON **4**.

13. GQ-ligand interaction by ^{19}F NMR. Modified ON **4** (15 μM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and 20% D_2O was prepared and ^{19}F NMR spectra (ns = 4000) of the sample was recorded with increasing concentration of ligand (0–60 μM). Spectral parameters are the same as mentioned in Section 7 and the data was processed with an exponential window function using lb = 20 Hz.

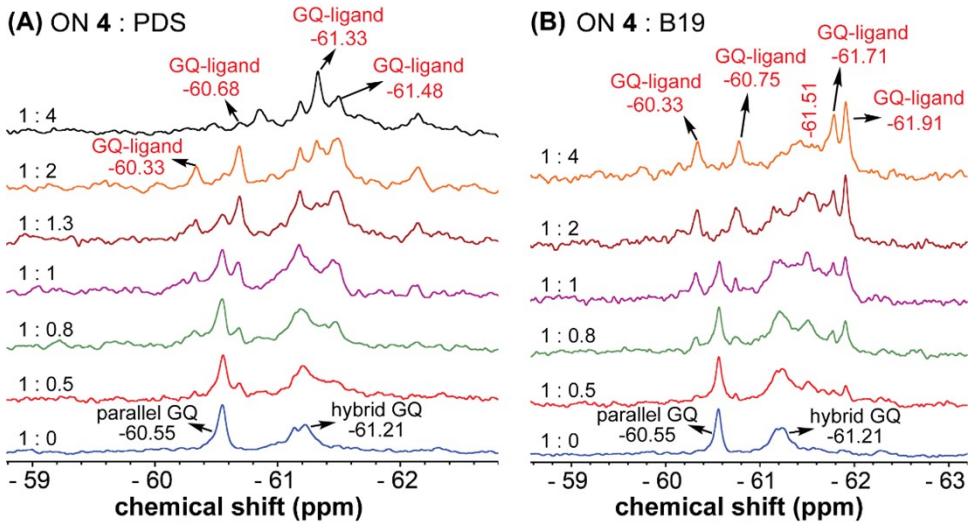


Fig. S18. ^{19}F NMR of modified ON **4** (15 µM) in Tris.HCl buffer (pH 7.4) containing 100 mM KCl with increasing concentrations of (**A**) PDS, and (**B**) BRACO-19 (B19).

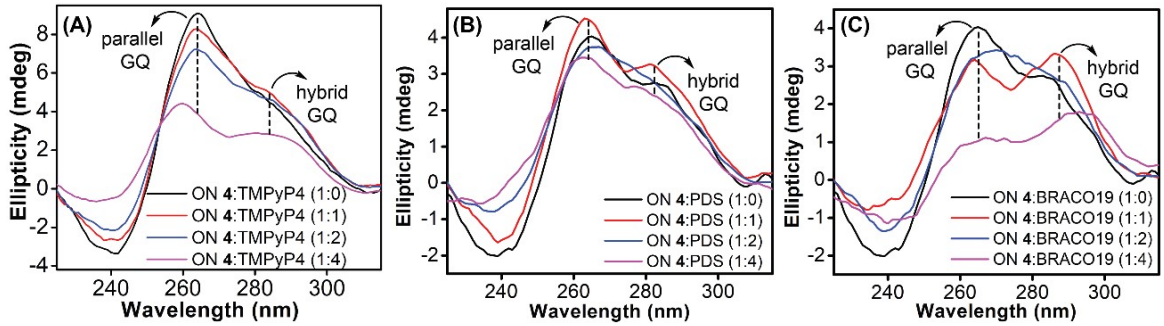


Fig. S19. (**A**) CD spectra of modified ON **4** (10 µM) upon addition of increasing concentrations of TMPyP4. CD of modified ON **4** (5 µM) at different concentrations of (**B**) PDS, and (**C**) BRACO-19.

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

Intraoocyte buffer. Modified ON **4** (50 µM) was annealed in an intraoocyte buffer (25 mM HEPES pH = 7.5, 110 mM KCl, 10.5 mM NaCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA)[S18] containing 20% D_2O at 90 °C for 3 min. The sample was cooled slowly to RT and incubated for 1 h at RT. The ^{19}F (number of scans = 1000) and ^{1}H NMR (number of scans = 1500) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned in section 7 and the ^{19}F NMR data was processed with an exponential window function using lb = 20 Hz.

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

Clear lysate.[S18] Healthy *Xenopus laevis* stage V/VI oocytes (~275) were selected and suspended in a Petri dish containing Ori-Ca^{2+} buffer (5 mM HEPES pH = 7.5, 110 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, and 1 mM $MgCl_2$). The Petri dish was kept on an ice bath for 15 min. The oocytes were washed with ice-cold intraoocyte buffer (3 x 10 mL) and resuspended in the same buffer. Oocytes were transferred in an Eppendorf tube, allowed them to settle down and the supernatant was removed carefully without disturbing the settled oocytes. Oocytes were rinsed with intraoocyte buffer (200 µL) containing 20% D_2O and then removed (this step was performed twice). Finally, 200 µL of intraoocyte buffer containing 20% D_2O was added to the Eppendorf tube containing oocytes, which were then mechanically crushed. The suspension was centrifuged at 20000g for 20 min at 4 °C, the interphase layer was carefully transferred into another Eppendorf and heated at 95 °C for 10 min. The solution was centrifuged at 20000g for 10 min at 4 °C and the clear lysate (285 µL) was transferred to another Eppendorf tube. 1 mM of preannealed ON **4** (15 µL) in an intraoocyte buffer supplemented with 20% D_2O was added to the clear lysate (final ON concentration = 50 µM). The sample was incubated at 4 °C for 30 min and transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F (number of scans = 2000)

and ^{1}H NMR (number of scans = 3072) spectra were acquired at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR plot was processed with an exponential window function using lb = 20 Hz.

Egg extract. [S19] Oocytes (~900) were kept in a Petri dish containing Ori Ca^{2+} buffer (pH 7.5) for 15 min on an ice bath. The oocytes were washed with ice-cold intraoocyte buffer (3 x 20 mL) and transferred to an Eppendorf tube. The buffer just above the oocytes was removed carefully and oocytes were washed with intraoocyte buffer (400 µL) containing 20% D_2O (two times). The oocytes were centrifuged at 400g for 1 min at 4 °C and the supernatant buffer was removed. The oocytes were resuspended in the intraoocyte buffer (100 µL) containing 30% D_2O and centrifuged at 12000g for 5 min at 4 °C. The eggs were mechanically crushed and the suspension was centrifuged at 12000g for 30 min at 4 °C to obtain the interphase layer. This crude egg extract thus obtained was directly used in the NMR analysis. 1 mM of the preannealed ON **4** (15 µL) in an intraoocyte buffer containing 20% D_2O was added to the above crude egg extract (285 µL) and incubated for 30 min at 4 °C. The ^{19}F (number of scans = 3500) and ^{1}H NMR (number of scans = 4000) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR spectrum was processed with an exponential window function using lb = 20 Hz.

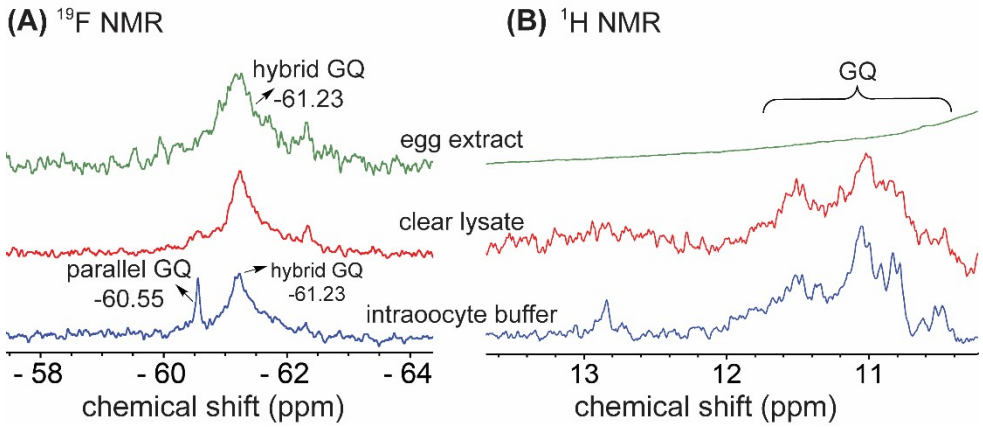


Fig. S20. ^{19}F and ^{1}H NMR spectra of ON **4** (50 µM) in intraoocyte buffer (blue), lysate (red) and egg extract (green). See section for 13 details.

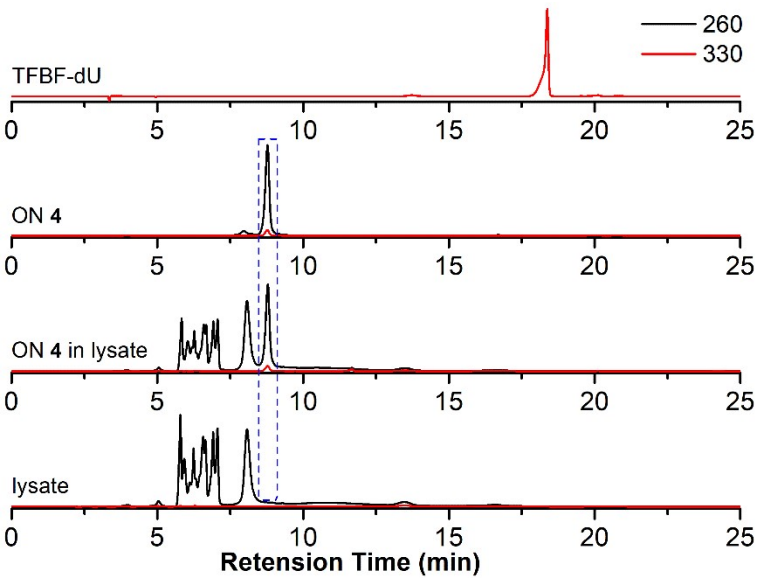


Fig. S21. Comparison of HPLC chromatograms of lysate, lysate containing ON **4** (after recording NMR), ON **4** and modified nucleoside analog (TFBF-dU). There was no detectable degradation of ON **4** in lysate.

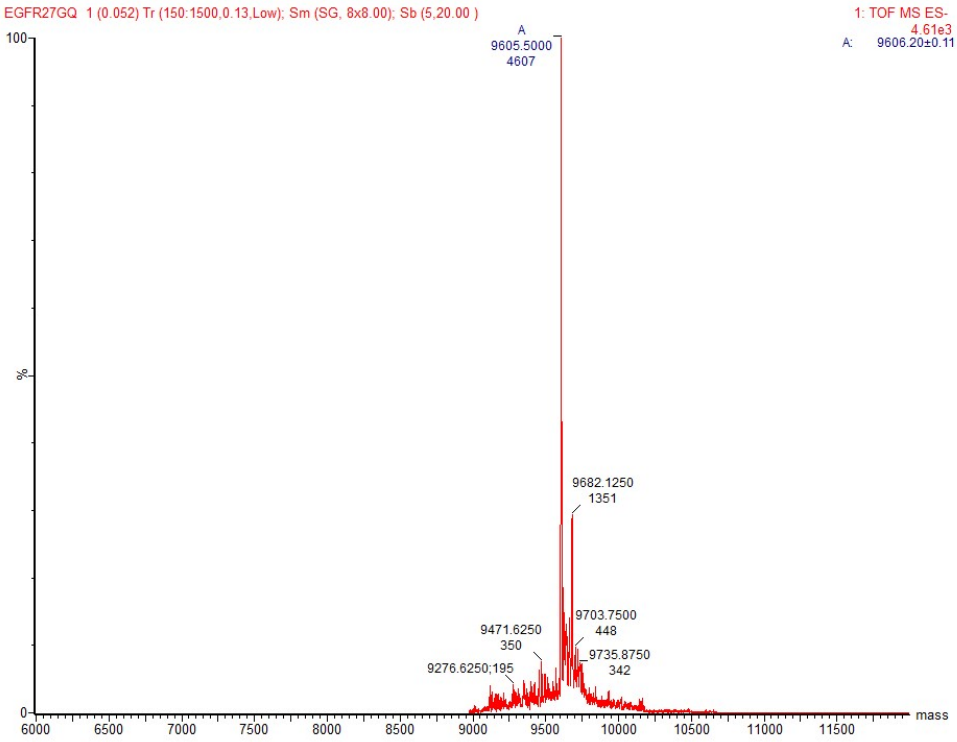


Fig. S22. ESI-MS spectra of modified ON **4** extracted from lysate sample after NMR analysis. (calculated mass = 9605.1, observed mass = 9605.5).

15. *Taq* polymerase stop assay. 5'-FAM-labeled primer (5 µM) and template DNA (**T_1–T_3**) (6 µM) were annealed in 10 mM Tris.HCl buffer containing 100 mM KCl by heating at 90 °C for 5 min. Samples were cooled slowly to RT and incubated on an ice batch for 1 h. The primer-DNA duplexes were diluted to a final concentration of 1 µM in 10 mM Tris.HCl buffer containing 100 mM KCl. Replication reaction was performed on a 20 µL reaction volume containing duplex (50 nM), dNTPs (500 µM), KCl (100 mM), 1X DNA polymerase buffer.[S20] The reaction mixture was incubated at 37 °C for 20 min and the reaction was initiated by adding 0.5 µL of Taq DNA polymerase (5 U/µL, *New England Biolabs*, Catlog. M0273S). The reaction was quenched at different time points by adding 10 µL of gel loading buffer (80% formamide by volume, 10 mM NaOH, 0.005% bromophenol blue (w/v)) and flashed cooled on a dry ice bath. The reaction mixture was concentrated on a speed-vac concentrator and analysed on a 15 % denaturing polyacrylamide gel. The gel was imaged by using a Typhoon gel scanner at FAM wavelength.

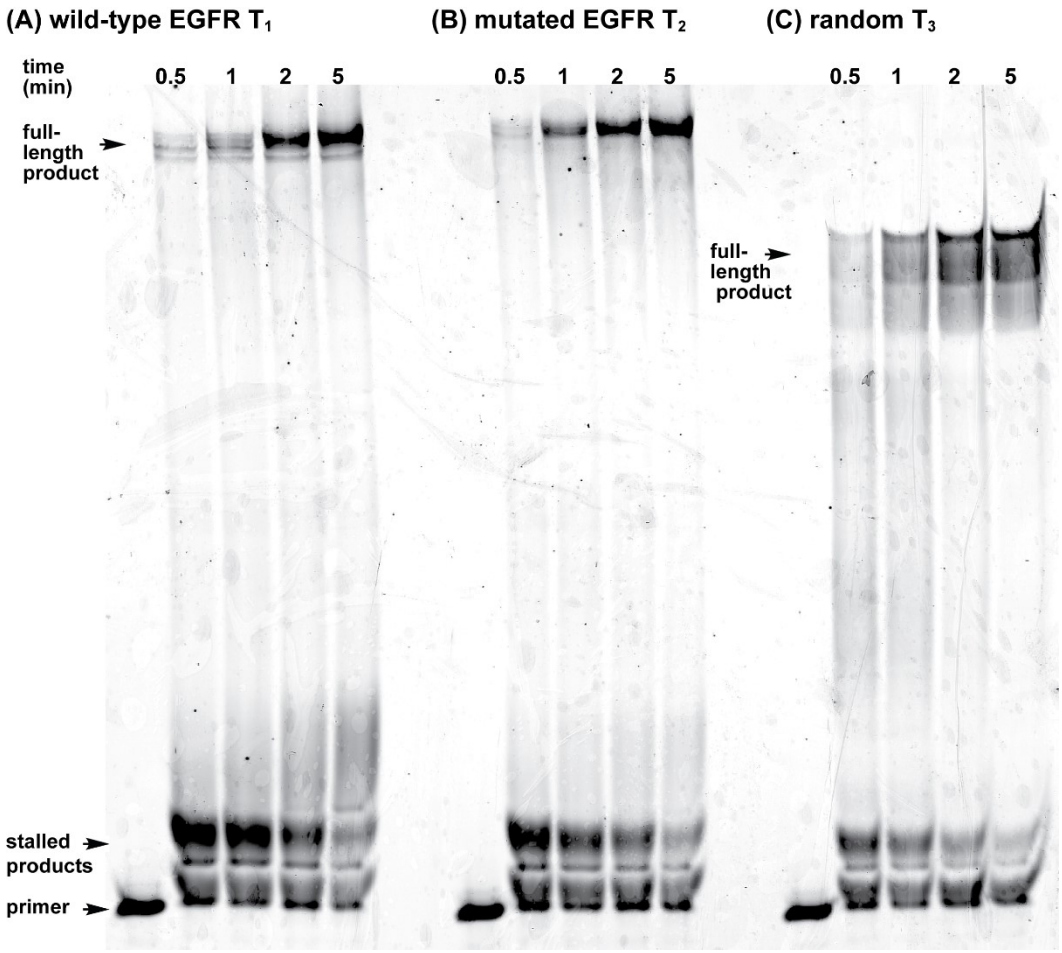


Fig. S23. PAGE analysis of the replication reactions using (**A**) a wild-type EGFR G-rich template $\mathbf{T_1}$, (**B**) a mutated EGFR G-rich template $\mathbf{T_2}$, and (**C**) a random non-GQ forming template $\mathbf{T_3}$.

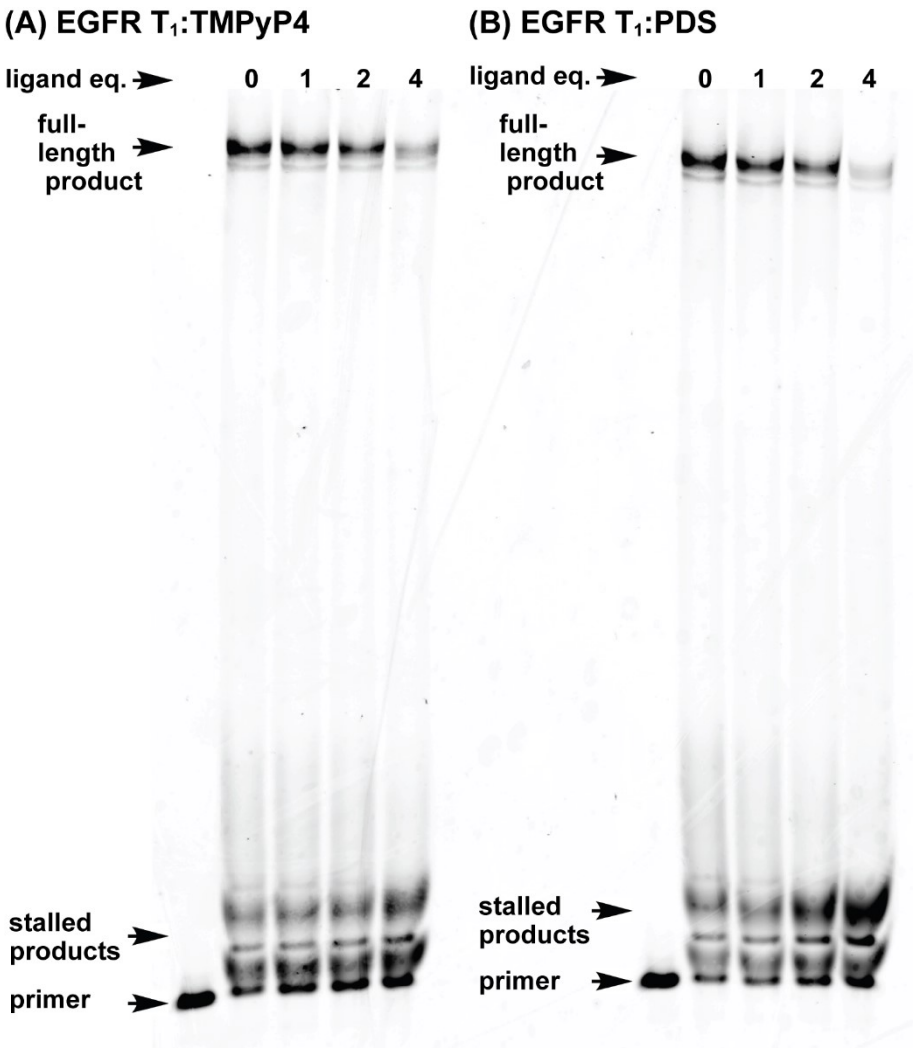


Fig. S24. PAGE analysis of the replication reactions using a wild-type EGFR $\mathbf{T_1}$ with increasing concentrations of the ligands (**A**) TMPyP4, (**B**) PDS.

16. References

S1. S. Y. Khatik and S. G. Srivatsan, *Bioconjugate Chem.*, 2022, **33**, 1515–1526.

S2. J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.

S3. L. P. P. Patro, A. Kumar, N. Kolimi and T. Rathinavelan, *J. Mol. Biol.*, 2017, **429**, 2438–2448.

S4. M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.

S5. X.-J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–1227.

S6. R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

S7. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi and M. K. Gilson, AMBER 2018; 2018. *Univ. California, San Fr.* **2018**.

S8. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory. Comput.*, 2013, **9**, 3878–3888.
S9. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
S10. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
S11. P. Turq and F. Lantelme, *J. Chem. Phys.*, 1977, **66**, 3039–3044.
S12. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
S13. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
S14. D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
S15. S. Manna, D. Sarkar and S. G. Srivatsan, *J. Am. Chem. Soc*., 2018, **140**, 12622–12633.
S16. N. V. Anantha, M. Azam and R. D. Sheardy, *Biochemistry*, 1998, **37**, 2709–2714.
S17. C. Wei, G. Jia, J. Yuan, Z. Feng and C. Li, *Biochemistry*, 2006, **45**, 6681-6691.
S18. R**.** Hänsel, S. Foldynová-Trantírková, F. Löhr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Dötsch and L. Trantírek, *J. Am. Chem. Soc.*, 2009, **131**, 15761–15768.
S19. R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res.*, 2011, **39**, 5768–5775.
S20. G. Wu and H. Han, *Methods Mol. Biol.*, 2019, **2035**, 223–231.

Electronic Supplementary Information (ESI)

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

Saddam Y. Khatik,[a] Sruthi Sudhakar,[b] Satyajit Mishra,[c] Jeet Kalia,[c,d] P.I. Pradeepkumar[b] and Seergazhi G. Srivatsan*[a]

[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.

[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.

1. Materials. 5-Trifluoromethyl benzofuran-modified nucleoside analog **1** and corresponding phosphoramidite substrate **2** for solid-phase oligonucleotide (ON) synthesis were synthesized as per our previously reported procedure.[S1] *N*-acetyl-protected dC, *N*-Benzoyl-protected dA, *N,N*-dimethylformamide-protected dG, and dT phosphoramidite substrates needed for solid-phase DNA synthesis were procured from ChemGenes. Solid supports required for DNA synthesis were purchased from Glen Research. All other reagents required for solid phase ON synthesis were obtained from Sigma-Aldrich. Control DNA ONs Telo1, TeloC, **6** and **7** were purchased from Integrated DNA Technology, and purified using 18% denaturing polyacrylamide gel electrophoresis (PAGE). All other reagents (BioUltra grade) for the preparation of buffer solutions were purchased from Sigma-Aldrich. Autoclaved water was used for preparing all buffer solutions, and in all biophysical studies.

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

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

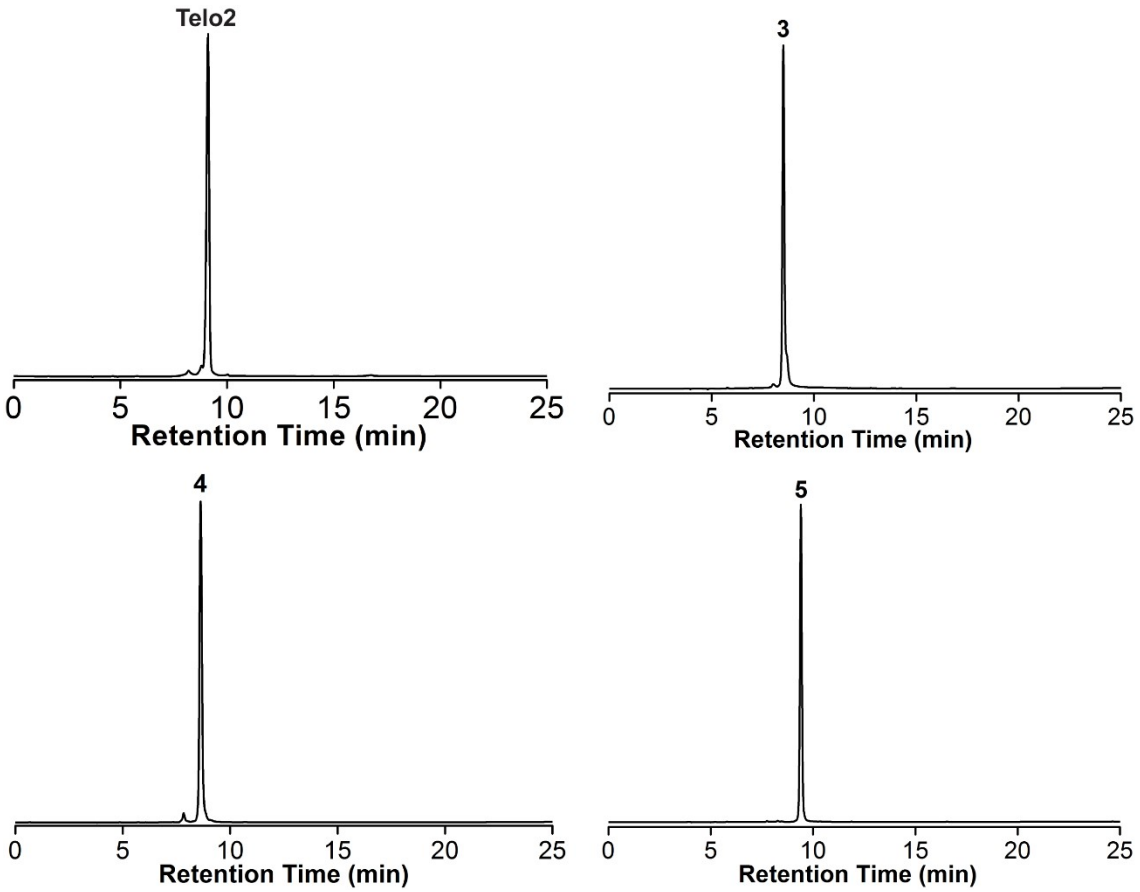


Fig. S1. RP-HPLC chromatograms of TFBF-modified Telo2 and EGFR ONs **3**–**5** analyzed at 260 nm. Mobile phase A = 50 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0-100 % B in 30 min. HPLC analysis was performed using a Luna C18 column (250 x 4.6 mm, 5 micron).

4. Mass analysis of modified ON.

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

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

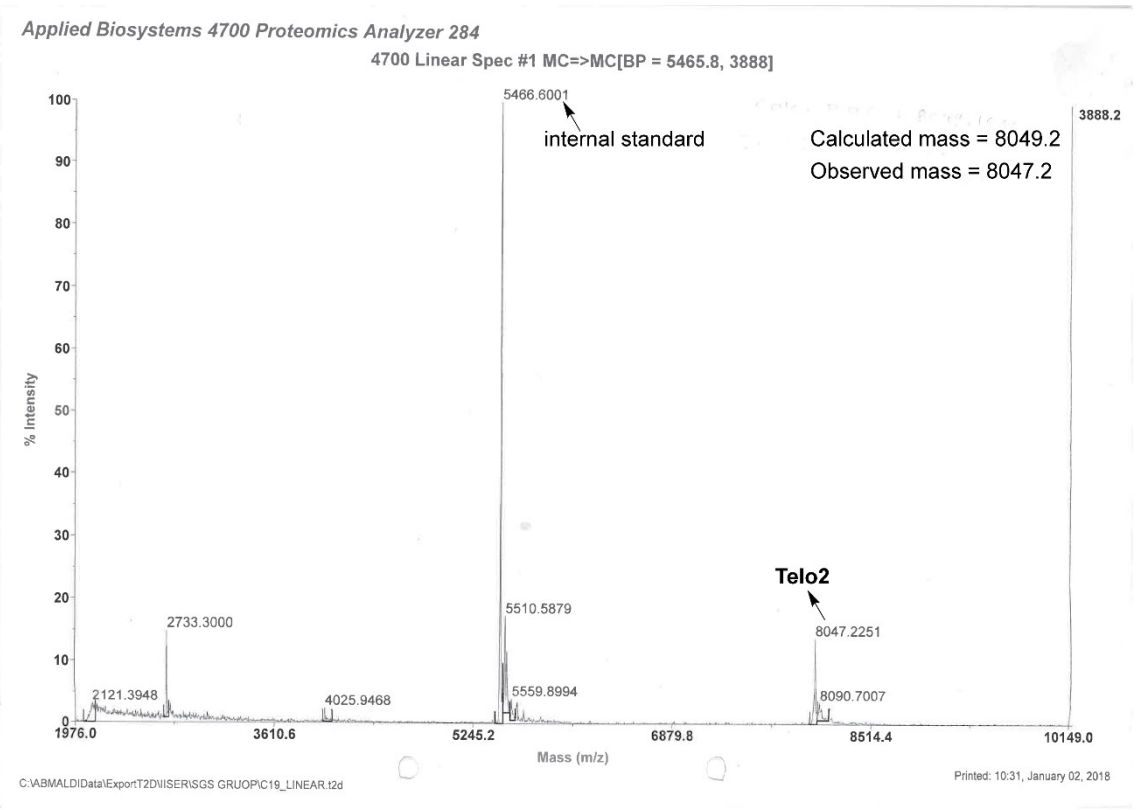
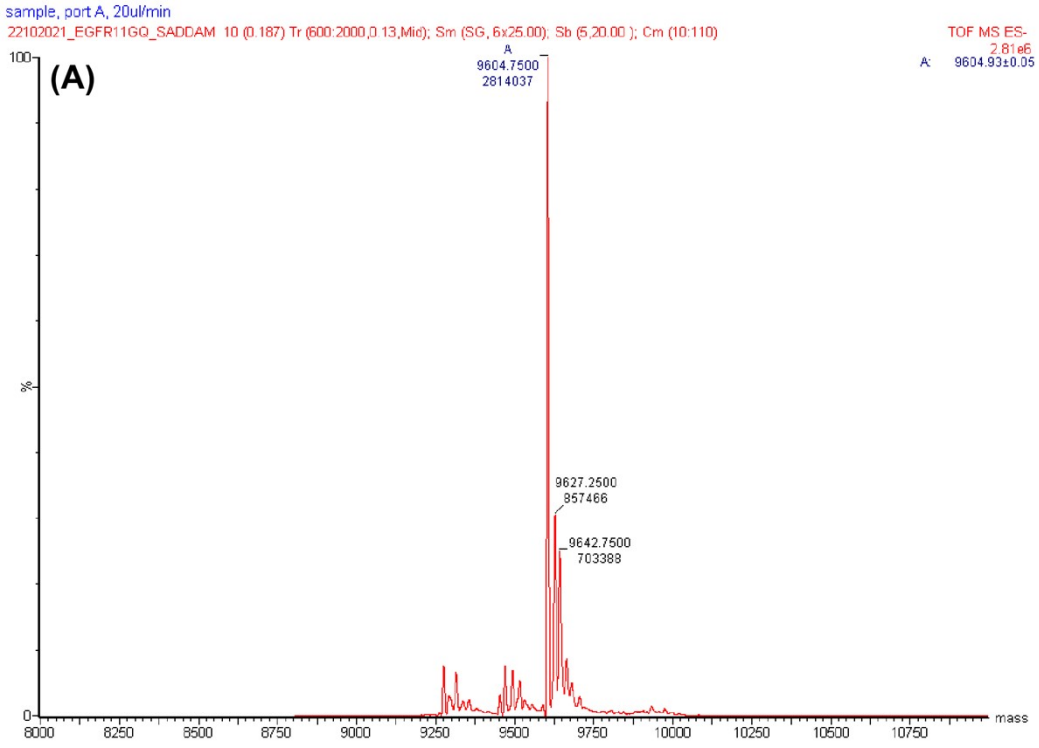
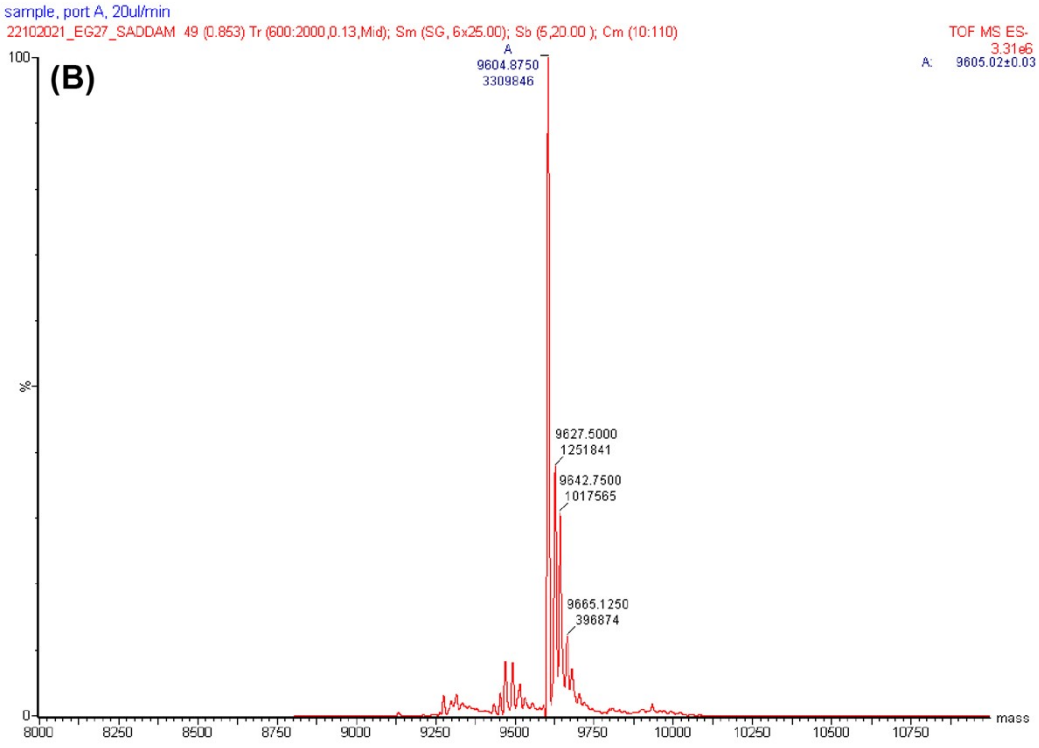


Fig. S2. MALDI-TOF MS spectrum of TFBF-modified Telo2. Internal DNA ON standard m/z of +1 and +2 ions are 5466.6 and 2733.3. See Section 4 for details.


sample, port A, 20ul/min
22102021_EGFR11GQ_SADDAM 10 (0.187) Tr (600:2000,0.13,Mid); Sm (SG, 6x25.00); Sb (5,20.00); Cm (10:110)
TOF MS ES-
2.81e6
A: 9604.93±0.05
(A)
A
9604.7500
2814037
9627.2500
857466
9642.7500
703388
%
mass
100
0
8000
8250
8500
8750
9000
9250
9500
9750
10000
10250
10500
10750


sample, port A, 20ul/min
22102021_EG27_SADDAM 49 (0.853) Tr (600:2000,0.13,Mid); Sm (SG, 6x25.00); Sb (5,20.00); Cm (10:110)
TOF MS ES-
3.31e6
A: 9605.02±0.03
(B)
A
9604.8750
3309846
9627.5000
1251841
9642.7500
1017565
9665.1250
396874
%
mass
100
0
8000
8250
8500
8750
9000
9250
9500
9750
10000
10250
10500
10750

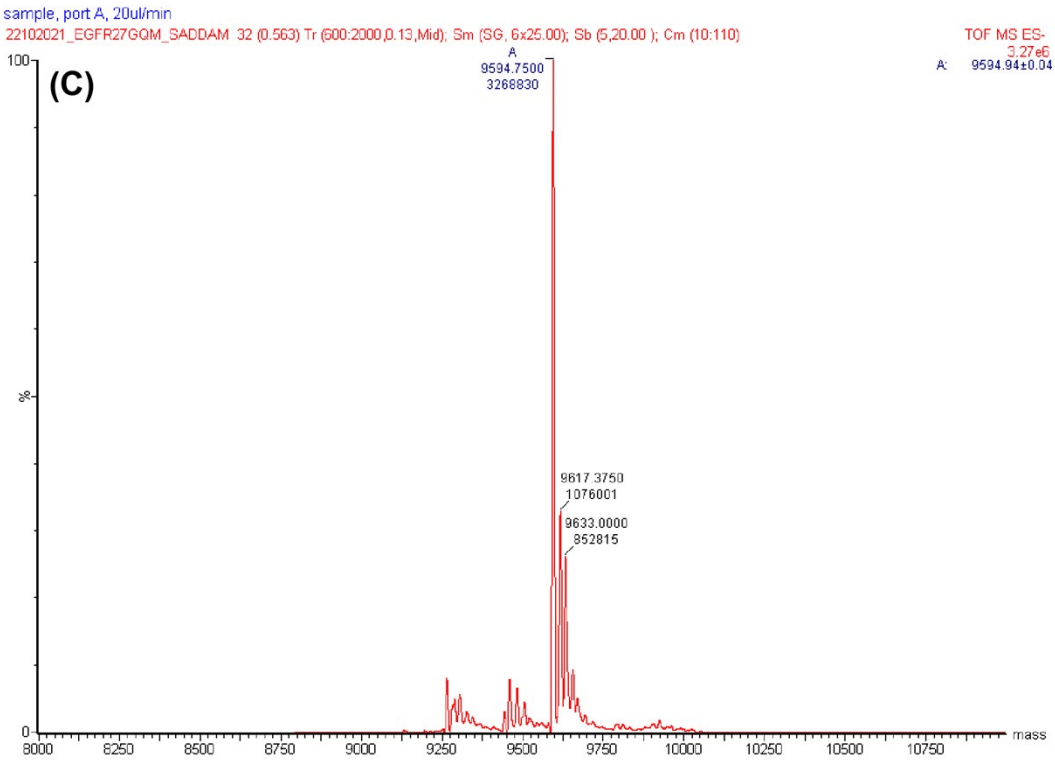


Fig. S3. ESI-MS spectra of TFBF-modified EGFR ONs (**A**) **3**, (**B**) **4**, (**C**) **5**.

Table S1. Molar absorptivity and mass of modified DNA ONs.

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

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

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

duplicate with an average of three scans for each sample. The spectrum of buffer in absence of ON was subtracted from all ON sample spectra.

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

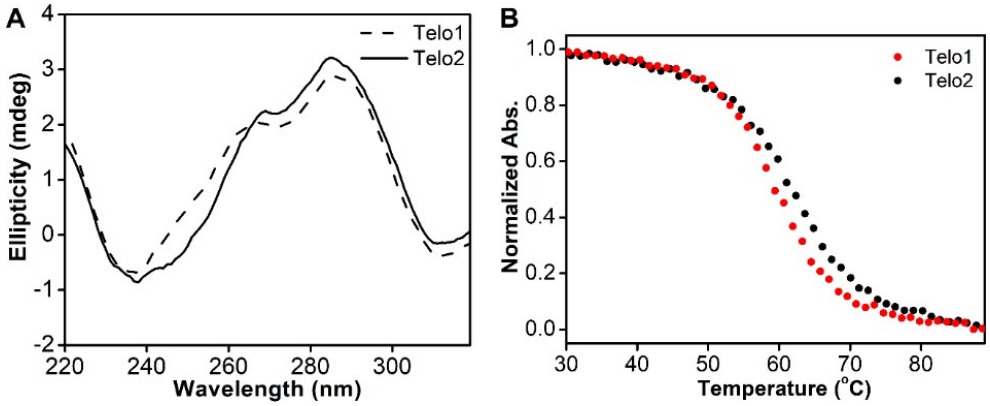


Fig. S4. (**A**) CD spectra of control Telo1 and modified Telo2 ONs (5 µM). (**B**) UV-thermal melting profile (at 295 nm) of Telo1 and Telo2 ONs (1 µM). See Section 5 and 6 for details.

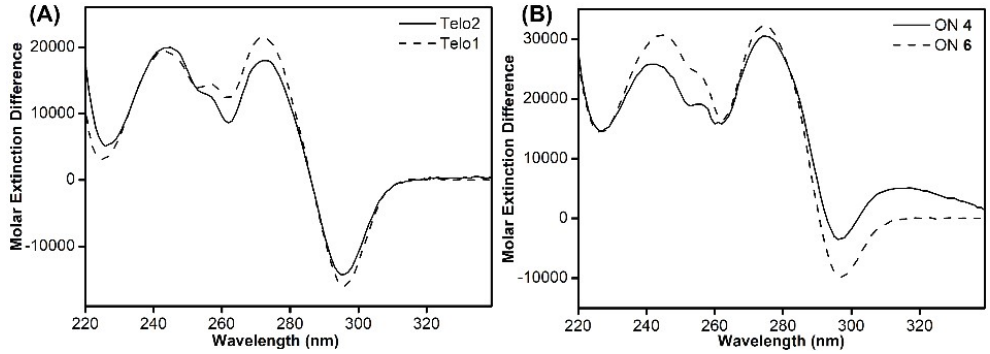


Fig. S5. Thermal difference spectrum (TDS) for Telo2 (**A**) and EGFR ON **4** (**B**), and corresponding unmodified control ON sequences. ONs samples (5 µM) were annealed in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and UV spectrum of ONs was recorded at 25 °C and 90 °C. The TDS was obtained by subtracting the UV absorption profile of the folded form from the unfolded form.

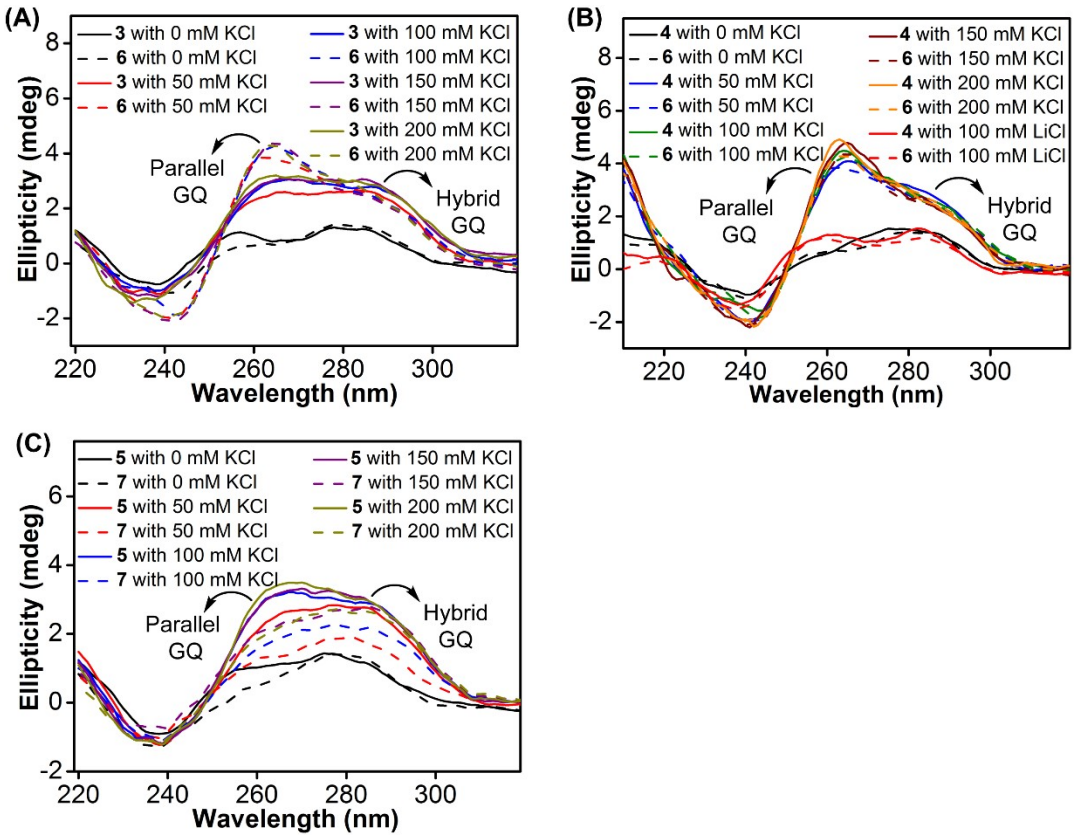


Fig. S6. (**A**, **B** and **C**) CD spectra (5 µM) of modified EGFR ONs **3**–**5** (solid lines) and their control ONs **6** and **7** (dashed lines) in different ionic conditions.

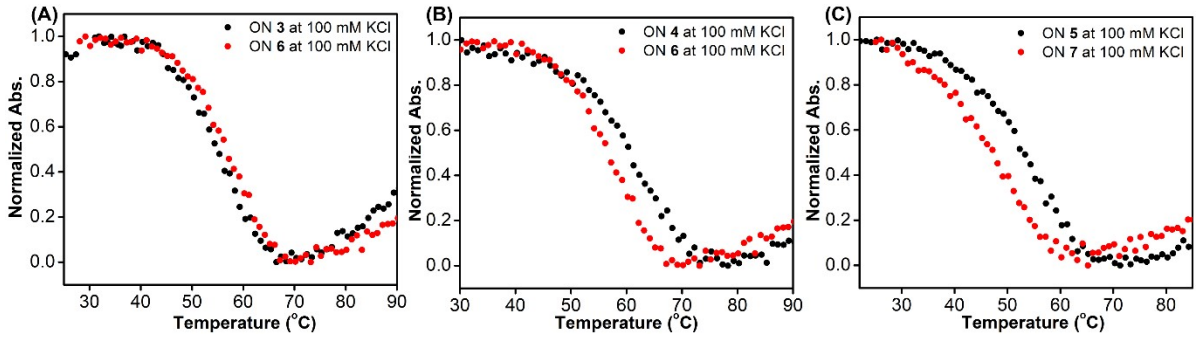


Fig. S7. UV-thermal melting profile (at 295 nm) of modified EGFR ONs **3**–**5** and corresponding control unmodified ONs **6** and **7** (1 µM).

Table S2. T_m values of modified Telo2 and EGFR ONs **3**–**5** and control unmodified Telo1 and EGFR ONs **6** and **7.**

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

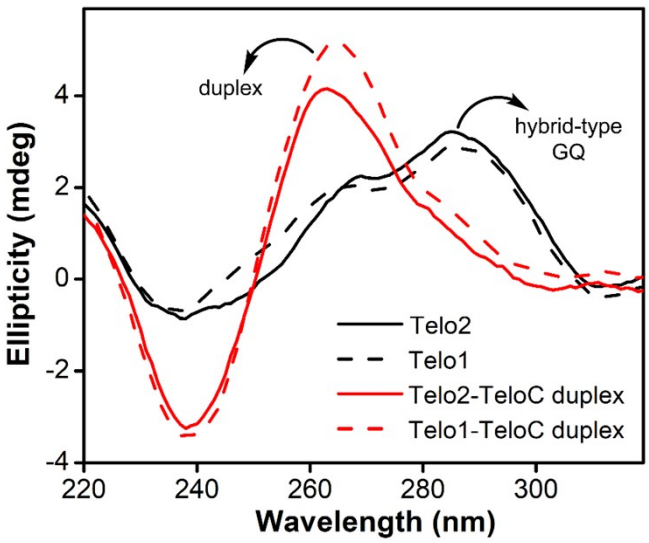


Fig. S8. A comparison of CD spectra of duplex and GQ structures formed by the telomeric repeat ONs (5 µM). See section 5 for experimental details.

7. Fluorescence and NMR studies of modified Telo2 ON.

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

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

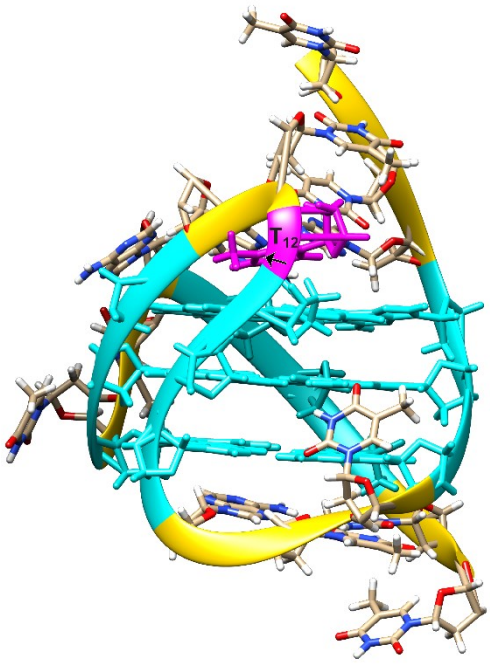


Fig. S9. NMR structure of a hybrid GQ topology of the human telomeric repeat (PDB: 2JPZ).[S2] Guanosines participating in the tetrad formation are shown in cyan. T_{12} residue is shown in magenta. Arrow shows the C5-postion of T_{12} where the TFBF heterocycle modification is placed in the modified Telo2 ON. TFBF modification does not affect the GQ structure. Hence, TFBF modification at C5-position is likely to be projected into the groove away from the G-tetrad. As a result of reduced stacking interaction, the GQ form shows higher fluorescence intensity. The figure was generated using UCSF Chimera version 1.15.

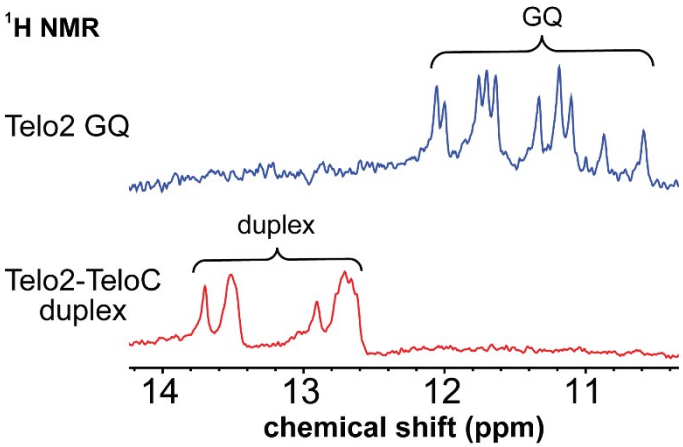


Fig. S10. ^{1}H NMR spectra of the modified telomeric ON forming GQ and duplex structures. See section 7 for details.

8. Fluorescence study of modified EGFR ONs.

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

Time-resolved fluorescence analysis of modified ONs. Samples of TFBF-dU modified ONs **4** and **5** (4 µM) in 10 mM Tris.HCl (pH 7.4) buffer were prepared and used for excited-state decay kinetics measurements. ONs samples were excited using a 340 nm LED source and the fluorescence decay was collected at respective emission maximum with an increase in KCl concentrations. The concentration of the KCl was increased by adding different aliquots of 3 M KCl into the ON sample and incubated for one hour before measurements. Fluorescence decay of the ON samples were deconvoluted using EZ software and fitted by an exponential decay with χ^2 value close to unity.

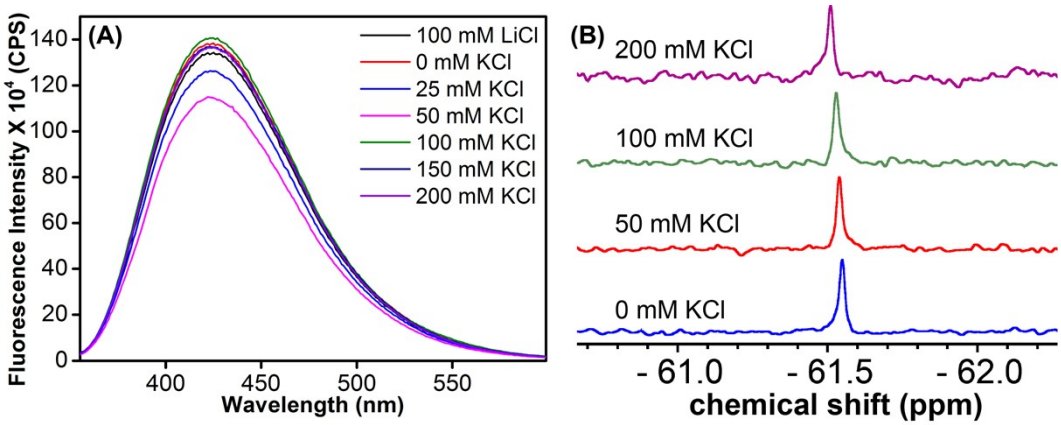


Fig. S11. (**A**) Fluorescence spectra of TFBF-dU (2 µM), and (**B**) ^{19}F-NMR spectra of TFBF-dU (10 µM) at different KCl concentrations. In the fluorescence study, samples were excited at 320 nm with excitation and emission slit widths of 5 nm and 6 nm, respectively.

9. Computational analysis. In order to generate the model for the two architectures of the EGFR G-rich sequences, a combination of 3D-NuS[S3] webserver and manual editing has been employed. For generating the hybrid architecture, the loop sequences were initially added to the 3D-NuS web server, where the hybrid GQ structure of class Q17 was generated. The 5' terminal dG was excluded in both architectures. The web server-generated structure was extracted in PDB format. The hairpin domain of the structure lacked any base pairing interaction. Therefore, the sequence of the hairpin domain was separately modeled with the help of the DNA folding form of the Mfold web server,[S4] RNAalifold web server, and 3DNA[S5]. The hairpin domain from the 3D-NuS PDB structure was removed, and the newly generated hairpin was placed with the help of PyMOL software. Further, the coordinates of the GQ and the newly generated hairpin domain were combined manually to generate a new PDB file. Similarly, for the parallel GQ, the primary

architecture was generated from the 3D-NuS web server, after which the hairpin domain was placed separately. The hairpin domain, as well as dT26 (dT27 of ON **4**), were manually placed with the help of PyMOL, and residue numbers were corrected and then saved in the PDB format.

Both the PDB structures generated were added into the Tleap module of AmberTools 21, and K^+ counter ions were added to generate the coordinate and topology files. The central K^+ ions were also added in both cases. The OL15[S6] force field was used for DNA and TIP3P water model for water and counter ions. The structures were subjected to a 100000-step minimization using the implicit solvent model in AMBER 18.[S7] The final structures after the minimization was again loaded to the tleap module. These structures were enclosed in a rectangular water box of 10 Å, and the coordinate and topology files were generated. The systems were then subjected to 10000 steps of minimization by the steepest descent method with a restraint of 2.0 kcal/mol $Å^2$ on the DNA and central K^+ ions. The minimization followed 100 ps of heating and 100 ps of density equilibration with restraints of 50 kcal/mol $Å^2$ and 2.0 kcal/mol $Å^2$. The systems were then equilibrated for 800 ps in NPT ensemble, and unrestrained MD simulation was performed using NPT ensemble for 400 ns in GPU accelerated version of PMEMD[S8-S10] in AMBER 18. SHAKE algorithm was applied to subject the hydrogen atoms to bond length constraints. Langevin[S11] thermostat with a collision frequency of 2 ps^{-1} was used to maintain the temperature at 300 K, and Berendsen[S12] barostat with a relaxation time of 2 ps was used to maintain pressure. The trajectories were visualized using UCSF Chimera[S13], and images were rendered using PyMOL. The analysis was carried out using the CPPTRAJ[S14] module of AmberTools.

The 400 ns trajectories were clustered into 5 ensembles using the hierarchical agglomerative clustering algorithm. The hybrid GQ has one major cluster which existed for ~60 % of the simulation time (Fig. S12A†), and the parallel GQ has one major cluster which existed for ~50 % (Fig. S12B†) of the simulation time. Since the 5′ terminal dG was excluded in both architectures, the numbering of the nucleotides in the ON sequence has been renumbered (Fig. 5). In the hybrid GQ, the hairpin domain maintains the two base pairs (dG19:dC25 and dC20:dG24) intact during the entire simulation, while the same is lost towards the end of the simulation in the parallel GQ. dT26 (dT27 in ON **4** and **6**) stacks with both dC25 and dG3 maintaining an average distance of 3.80 ± 0.24 Å and 4.10 ± 0.32 Å respectively in the hybrid GQ. In the case of parallel GQ, dT26 partially stacks over dG16, maintaining an average distance of 4.49 ± 1.17 Å. The center of mass of the heavy atoms of the bases (excluding sugar) was considered for the distance calculation.

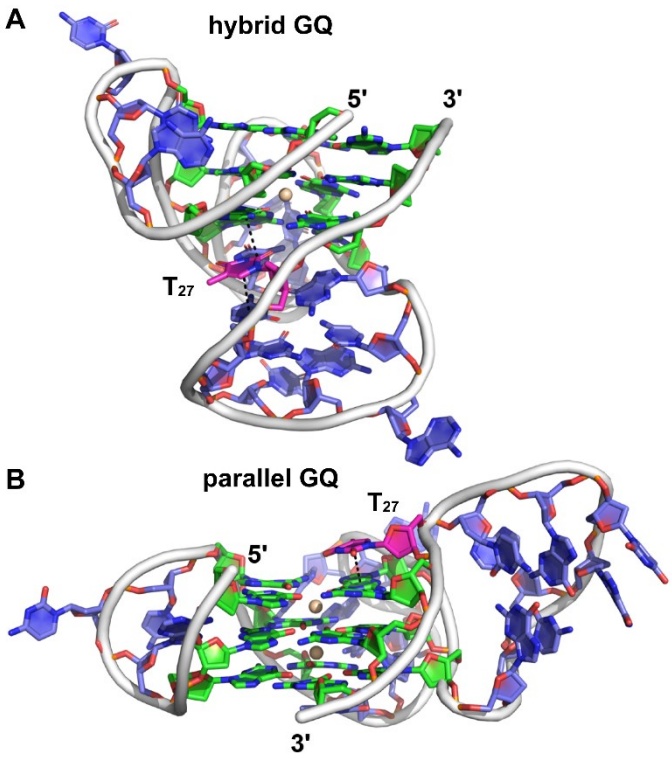


Fig. S12. Representative structure of the major cluster of (**A**) hybrid and (**B**) parallel GQ topologies. Carbon atoms of G-tetrads are represented in green, T_{27} is shown in magenta, and all other nucleotides are represented in purple. Nitrogen atoms are represented in blue, oxygen in red, and phosphate in orange. Hydrogen atoms are omitted for clarity. (**A**) In the hybrid form, T_{27} is stacked between the hairpin domain and a G-tetrad. (**B**) In the parallel form, T_{27} is partially stacked over the top G-tetrad. See Section 9 for details.

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

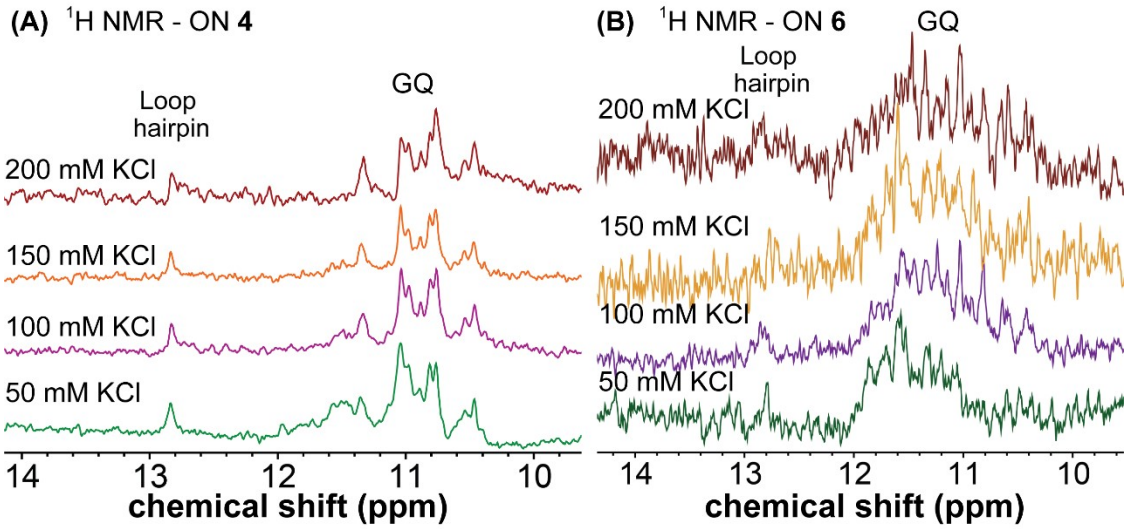


Fig. S13. ^{1}H NMR of TFBF-dU modified ON **4** and unmodified ON **6** of the native sequence.

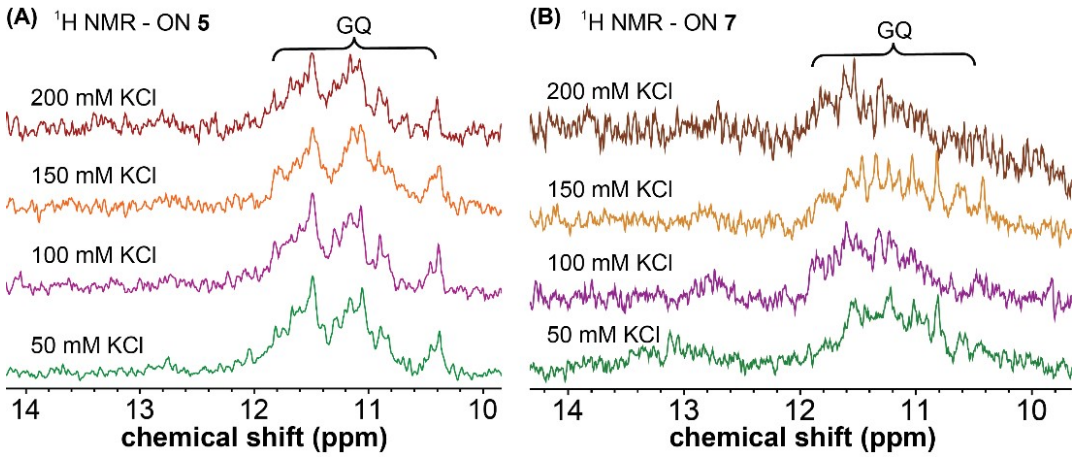


Fig. S14. ^{1}H NMR of TFBF-dU modified ON **5** and unmodified ON **7** of the mutated sequence.

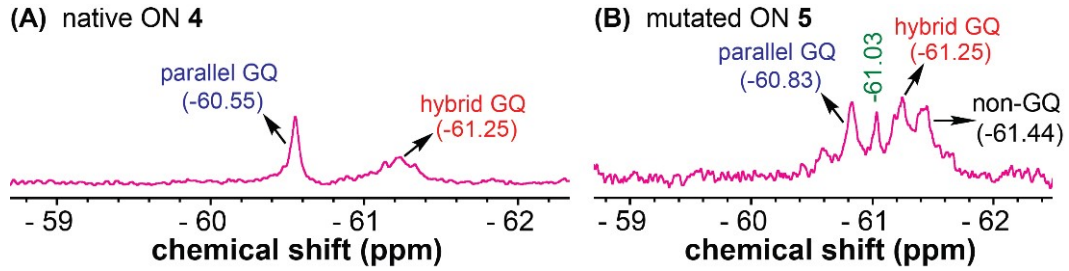


Fig. S15. ^{19}F NMR spectra of the native ON **4** and mutated ON **5** at 100 mM KCl.

11. GQ-ligand interaction by fluorescence. A series of samples of modified ON **4** (0.5 µM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ligands (5 nM to 5 µM) were prepared. Samples were incubated at 4 °C for 1 h. Fluorescence emission spectra were recorded by exciting the samples at 330 nM with excitation and emission wavelength slit widths of 6 nm and 7 nm, respectively. The fluorescence of samples was recorded in triplicate reading at 25 °C. Further, the fluorescence of the blank sample containing control ON **6** and respective ligand concentration was subtracted from the individual sample reading. The apparent K_d values of ligands were determined by plotting the normalized fluorescence intensity (F_N) *vs* ligand concentration. The graph was prepared using OriginPro 8.5 software.[S15]

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

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

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

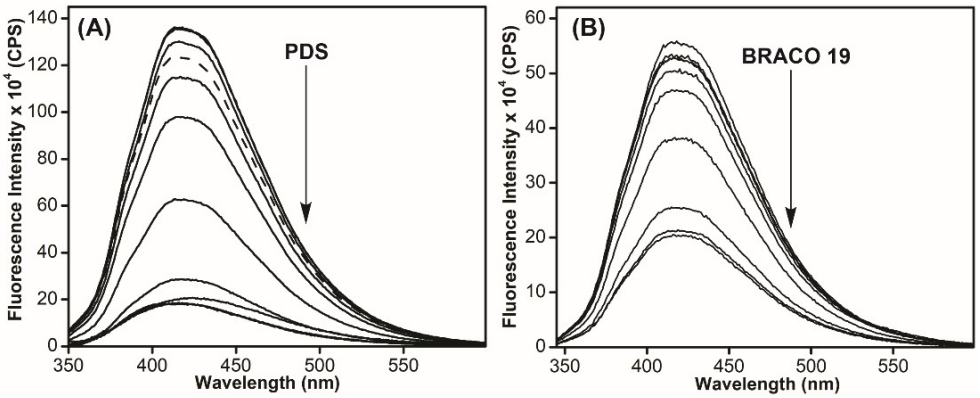


Fig. S16. Fluorescence titration of ON **4** (0.5 µM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of (**A**) PDS, and (**B**) BRACO-19. Fluorescence spectra of the modified ON **4** in the absence of ligands represented by a dashed line. Samples were excited at 330 nm with excitation and emission slit widths of 6 nm and 7 nm, respectively.

12. GQ-ligand interaction by UV absorption.

The samples of TMPyP4 (2 µM) were prepared in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of pre-annealed ON **4** (5 nM to 2 µM). Samples were incubated at 4 °C for 1 h and UV absorption spectra were acquired at 25 °C. Titration was performed until the wavelength and intensity of the absorption band of TMPyP4 remained unchanged.[S16,S17] UV experiment was performed in duplicate. Binding constant (K_d) was obtained from the plot of normalized absorbance (A_N) at 422 nm vs concentration of ON **4**. The graph was prepared using OriginPro 8.5 software.

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

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

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

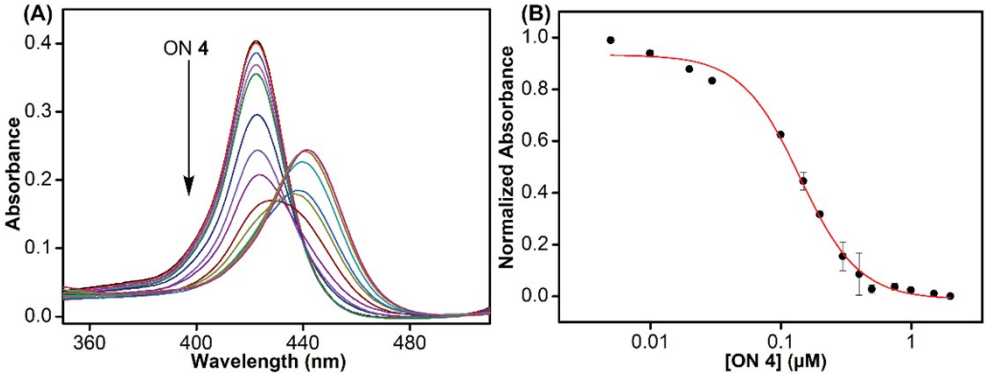


Fig. S17. (**A**) UV-vis absorption spectra. Titration of TMPyP4 (2 μM) in 10 mM Tris.HCl (pH 7.4) containing 100 mM KCl with increasing concentration of ON **4** (5 nM to 2 μM). (**B**) Curve fit plotted for normalized absorbance of TMPyP4 at 422 nm with increasing concentrations of ON **4**.

13. GQ-ligand interaction by ^{19}F NMR. Modified ON **4** (15 μM) in 10 mM Tris.HCl buffer (pH 7.4) containing 100 mM KCl and 20% D_2O was prepared and ^{19}F NMR spectra (ns = 4000) of the sample was recorded with increasing concentration of ligand (0–60 μM). Spectral parameters are the same as mentioned in Section 7 and the data was processed with an exponential window function using lb = 20 Hz.

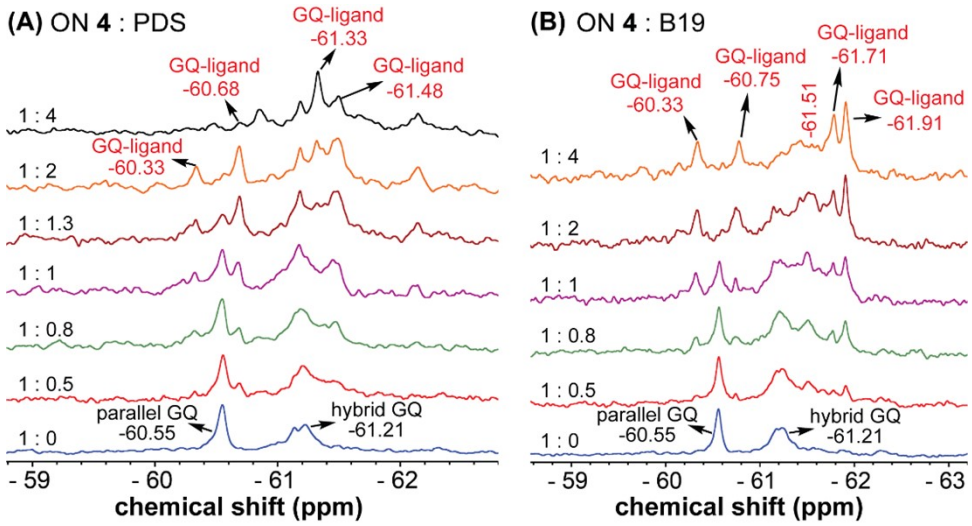


Fig. S18. ^{19}F NMR of modified ON **4** (15 µM) in Tris.HCl buffer (pH 7.4) containing 100 mM KCl with increasing concentrations of (**A**) PDS, and (**B**) BRACO-19 (B19).

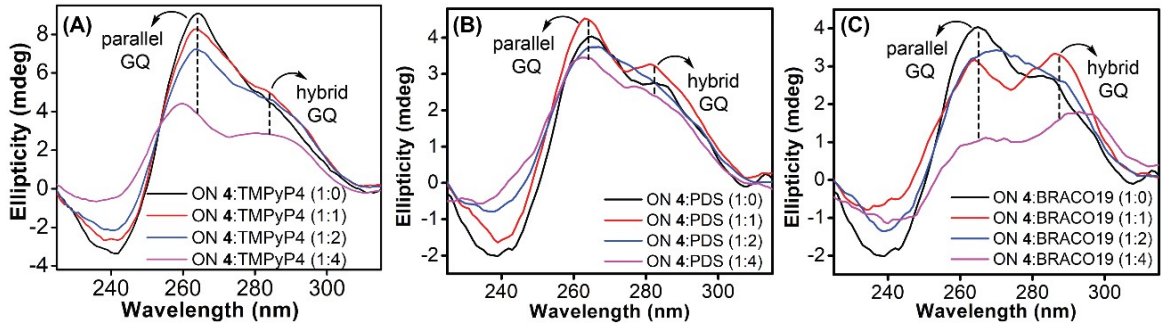


Fig. S19. (**A**) CD spectra of modified ON **4** (10 µM) upon addition of increasing concentrations of TMPyP4. CD of modified ON **4** (5 µM) at different concentrations of (**B**) PDS, and (**C**) BRACO-19.

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

Intraoocyte buffer. Modified ON **4** (50 µM) was annealed in an intraoocyte buffer (25 mM HEPES pH = 7.5, 110 mM KCl, 10.5 mM NaCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$, 0.1 mM EDTA)[S18] containing 20% D_2O at 90 °C for 3 min. The sample was cooled slowly to RT and incubated for 1 h at RT. The ^{19}F (number of scans = 1000) and ^{1}H NMR (number of scans = 1500) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned in section 7 and the ^{19}F NMR data was processed with an exponential window function using lb = 20 Hz.

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

Clear lysate.[S18] Healthy *Xenopus laevis* stage V/VI oocytes (~275) were selected and suspended in a Petri dish containing Ori-Ca^{2+} buffer (5 mM HEPES pH = 7.5, 110 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, and 1 mM $MgCl_2$). The Petri dish was kept on an ice bath for 15 min. The oocytes were washed with ice-cold intraoocyte buffer (3 x 10 mL) and resuspended in the same buffer. Oocytes were transferred in an Eppendorf tube, allowed them to settle down and the supernatant was removed carefully without disturbing the settled oocytes. Oocytes were rinsed with intraoocyte buffer (200 µL) containing 20% D_2O and then removed (this step was performed twice). Finally, 200 µL of intraoocyte buffer containing 20% D_2O was added to the Eppendorf tube containing oocytes, which were then mechanically crushed. The suspension was centrifuged at 20000g for 20 min at 4 °C, the interphase layer was carefully transferred into another Eppendorf and heated at 95 °C for 10 min. The solution was centrifuged at 20000g for 10 min at 4 °C and the clear lysate (285 µL) was transferred to another Eppendorf tube. 1 mM of preannealed ON **4** (15 µL) in an intraoocyte buffer supplemented with 20% D_2O was added to the clear lysate (final ON concentration = 50 µM). The sample was incubated at 4 °C for 30 min and transferred to a Shigemi tube (5 mm advance NMR micro-tube) for NMR analysis. ^{19}F (number of scans = 2000)

and ^{1}H NMR (number of scans = 3072) spectra were acquired at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR plot was processed with an exponential window function using lb = 20 Hz.

Egg extract. [S19] Oocytes (~900) were kept in a Petri dish containing Ori Ca^{2+} buffer (pH 7.5) for 15 min on an ice bath. The oocytes were washed with ice-cold intraoocyte buffer (3 x 20 mL) and transferred to an Eppendorf tube. The buffer just above the oocytes was removed carefully and oocytes were washed with intraoocyte buffer (400 µL) containing 20% D_2O (two times). The oocytes were centrifuged at 400g for 1 min at 4 °C and the supernatant buffer was removed. The oocytes were resuspended in the intraoocyte buffer (100 µL) containing 30% D_2O and centrifuged at 12000g for 5 min at 4 °C. The eggs were mechanically crushed and the suspension was centrifuged at 12000g for 30 min at 4 °C to obtain the interphase layer. This crude egg extract thus obtained was directly used in the NMR analysis. 1 mM of the preannealed ON **4** (15 µL) in an intraoocyte buffer containing 20% D_2O was added to the above crude egg extract (285 µL) and incubated for 30 min at 4 °C. The ^{19}F (number of scans = 3500) and ^{1}H NMR (number of scans = 4000) spectra were recorded at a frequency of 564.9 MHz and 600 MHz at 25 °C, respectively. Spectral parameters are the same as mentioned above. The ^{19}F NMR spectrum was processed with an exponential window function using lb = 20 Hz.

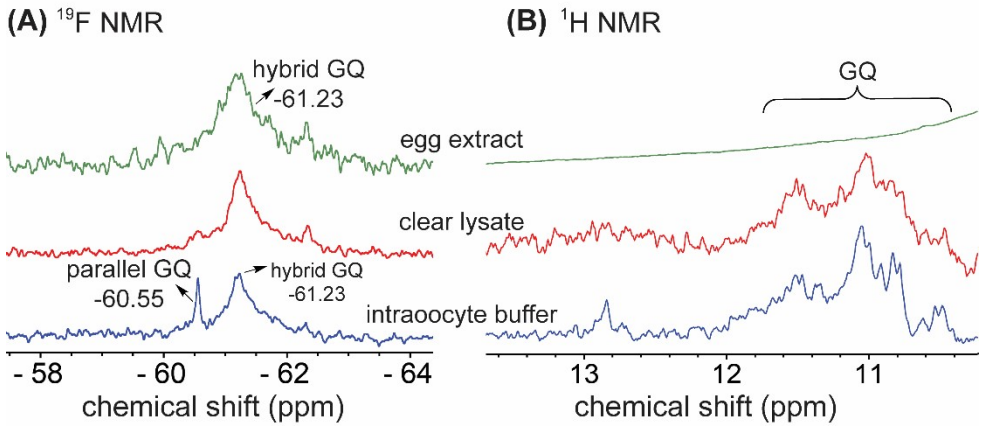


Fig. S20. ^{19}F and ^{1}H NMR spectra of ON **4** (50 µM) in intraoocyte buffer (blue), lysate (red) and egg extract (green). See section for 13 details.

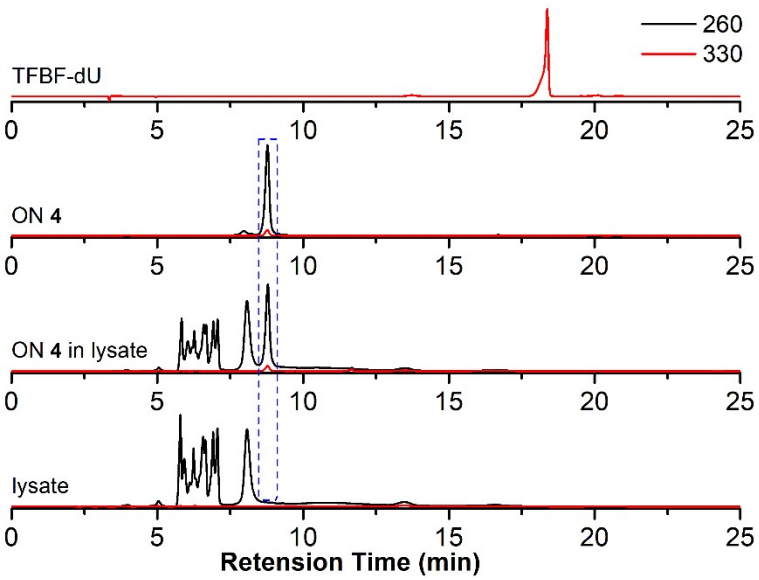


Fig. S21. Comparison of HPLC chromatograms of lysate, lysate containing ON **4** (after recording NMR), ON **4** and modified nucleoside analog (TFBF-dU). There was no detectable degradation of ON **4** in lysate.

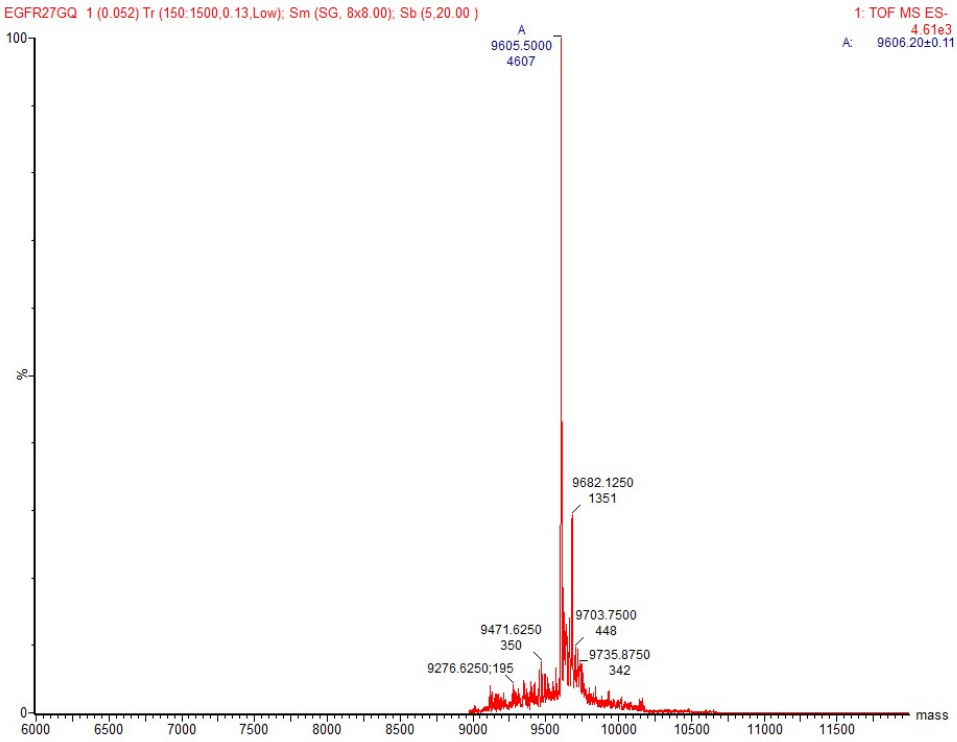


Fig. S22. ESI-MS spectra of modified ON **4** extracted from lysate sample after NMR analysis. (calculated mass = 9605.1, observed mass = 9605.5).

15. *Taq* polymerase stop assay. 5'-FAM-labeled primer (5 µM) and template DNA (**T_1–T_3**) (6 µM) were annealed in 10 mM Tris.HCl buffer containing 100 mM KCl by heating at 90 °C for 5 min. Samples were cooled slowly to RT and incubated on an ice batch for 1 h. The primer-DNA duplexes were diluted to a final concentration of 1 µM in 10 mM Tris.HCl buffer containing 100 mM KCl. Replication reaction was performed on a 20 µL reaction volume containing duplex (50 nM), dNTPs (500 µM), KCl (100 mM), 1X DNA polymerase buffer.[S20] The reaction mixture was incubated at 37 °C for 20 min and the reaction was initiated by adding 0.5 µL of Taq DNA polymerase (5 U/µL, *New England Biolabs*, Catlog. M0273S). The reaction was quenched at different time points by adding 10 µL of gel loading buffer (80% formamide by volume, 10 mM NaOH, 0.005% bromophenol blue (w/v)) and flashed cooled on a dry ice bath. The reaction mixture was concentrated on a speed-vac concentrator and analysed on a 15 % denaturing polyacrylamide gel. The gel was imaged by using a Typhoon gel scanner at FAM wavelength.

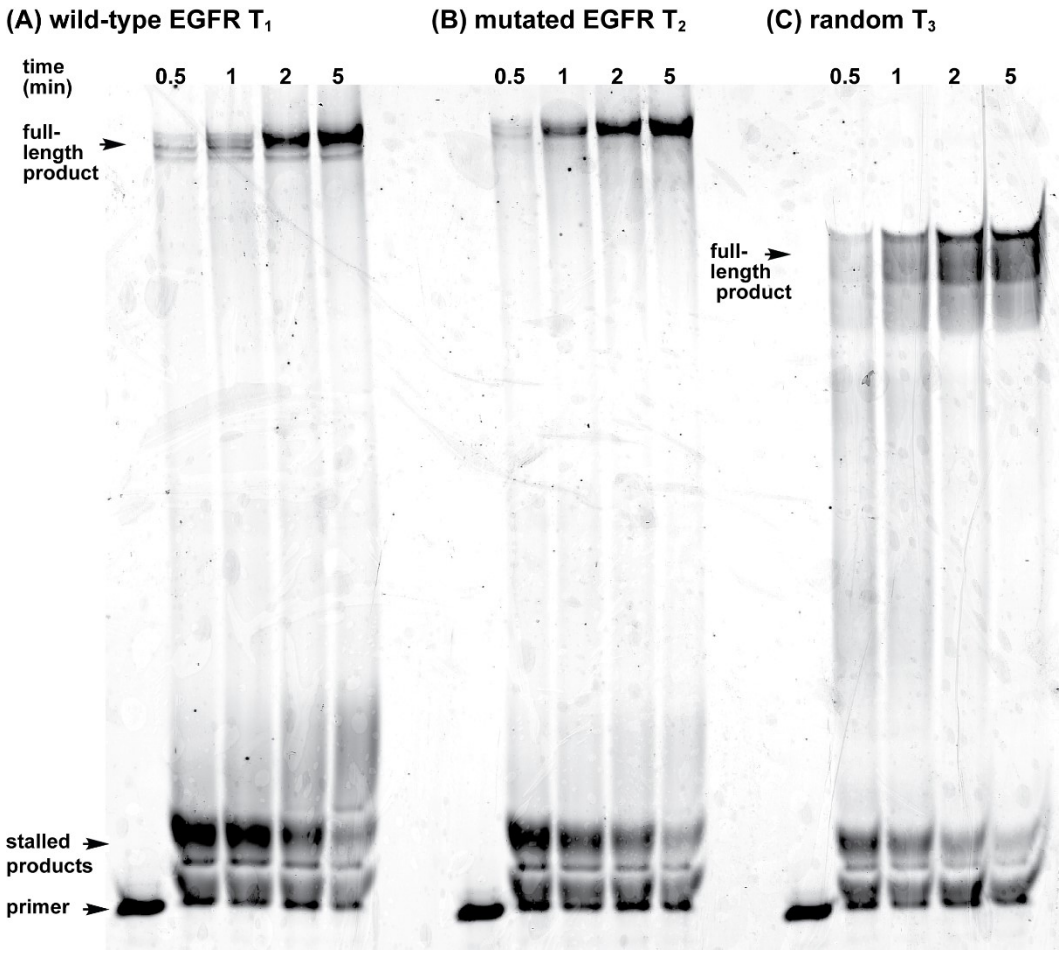


Fig. S23. PAGE analysis of the replication reactions using (**A**) a wild-type EGFR G-rich template $\mathbf{T_1}$, (**B**) a mutated EGFR G-rich template $\mathbf{T_2}$, and (**C**) a random non-GQ forming template $\mathbf{T_3}$.

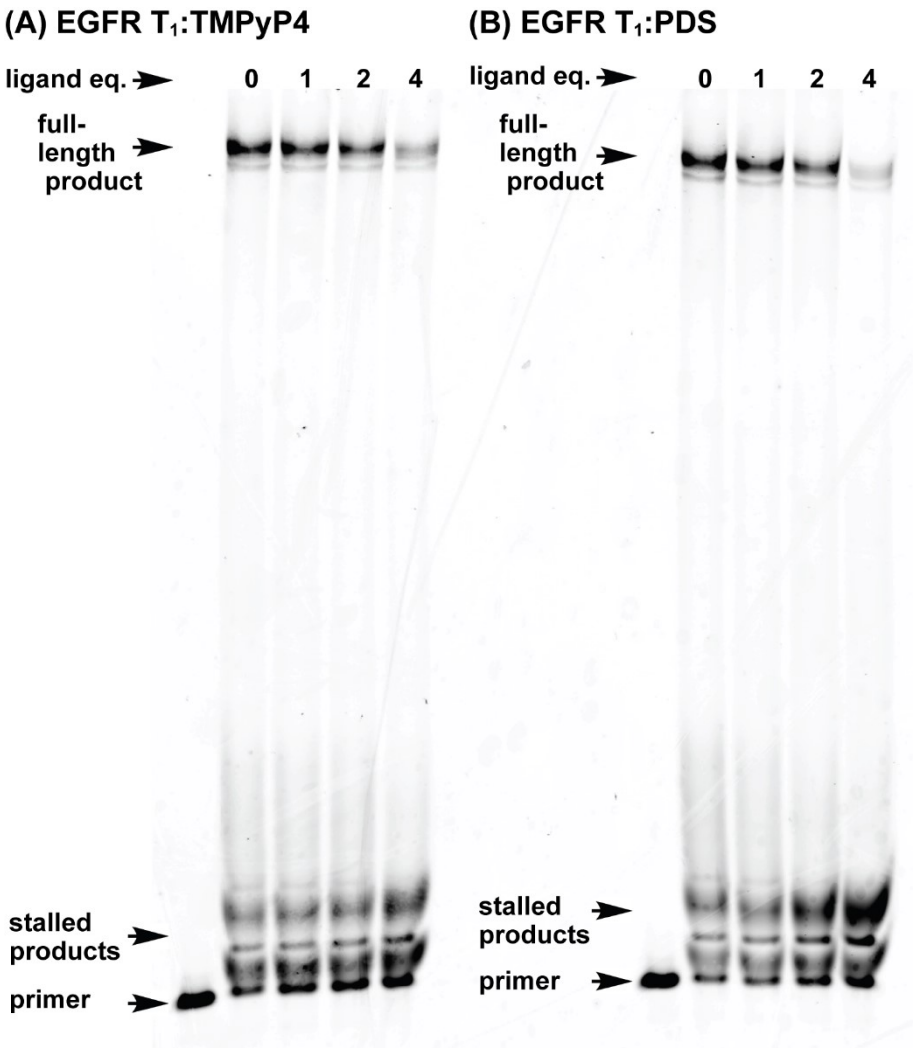


Fig. S24. PAGE analysis of the replication reactions using a wild-type EGFR $\mathbf{T_1}$ with increasing concentrations of the ligands (**A**) TMPyP4, (**B**) PDS.

16. References

S1. S. Y. Khatik and S. G. Srivatsan, *Bioconjugate Chem.*, 2022, **33**, 1515–1526.

S2. J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.

S3. L. P. P. Patro, A. Kumar, N. Kolimi and T. Rathinavelan, *J. Mol. Biol.*, 2017, **429**, 2438–2448.

S4. M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.

S5. X.-J. Lu and W. K. Olson, *Nat. Protoc.*, 2008, **3**, 1213–1227.

S6. R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka and T. E. Cheatham, *J. Chem. Theory Comput.*, 2016, **12**, 4114–4127.

S7. D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi and M. K. Gilson, AMBER 2018; 2018. *Univ. California, San Fr.* **2018**.

S8. R. Salomon-Ferrer, A. W. Götz, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory. Comput.*, 2013, **9**, 3878–3888.
S9. A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. L. Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
S10. S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
S11. P. Turq and F. Lantelme, *J. Chem. Phys.*, 1977, **66**, 3039–3044.
S12. H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
S13. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
S14. D. R. Roe and T. E. Cheatham III, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
S15. S. Manna, D. Sarkar and S. G. Srivatsan, *J. Am. Chem. Soc*., 2018, **140**, 12622–12633.
S16. N. V. Anantha, M. Azam and R. D. Sheardy, *Biochemistry*, 1998, **37**, 2709–2714.
S17. C. Wei, G. Jia, J. Yuan, Z. Feng and C. Li, *Biochemistry*, 2006, **45**, 6681-6691.
S18. R**.** Hänsel, S. Foldynová-Trantírková, F. Löhr, J. Buck, E. Bongartz, E. Bamberg, H. Schwalbe, V. Dötsch and L. Trantírek, *J. Am. Chem. Soc.*, 2009, **131**, 15761–15768.
S19. R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch, *Nucleic Acids Res.*, 2011, **39**, 5768–5775.
S20. G. Wu and H. Han, *Methods Mol. Biol.*, 2019, **2035**, 223–231.