Published online 7 August 2021

Nucleic Acids Research, 2021, Vol. 49, No. 15 ***8947–8960***
https://doi.org/10.1093/nar/gkab659

c-kit2 G-quadruplex stabilized via a covalent probe: exploring G-quartet asymmetry

Kateřina Peterková[1,2,3], Ivo Durník[2,4], Radek Marek[2,4,5], Janez Plavec[1,3,6] and Peter Podbevšek[1,*]

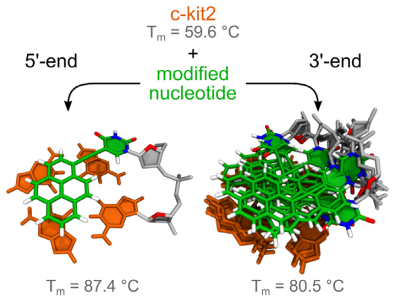
[1]Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia, [2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, CZ-625 00 Brno, Czechia, [3]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia, [4]CEITEC-Central European Institute of Technology, Masaryk University, Kamenice 5, CZ-62500 Brno, Czechia, [5]Department of Chemistry, Faculty of Science, Masaryk University, Kamenice 5, CZ-62500 Brno, Czechia and [6]EN-FIST Centre of Excellence, Trg OF 13, SI-1000 Ljubljana, Slovenia

Received March 30, 2021; Revised July 12, 2021; Editorial Decision July 14, 2021; Accepted July 22, 2021

ABSTRACT

Several sequences forming G-quadruplex are highly conserved in regulatory regions of genomes of different organisms and affect various biological processes like gene expression. Diverse G-quadruplex properties can be modulated via their interaction with small polyaromatic molecules such as pyrene. To investigate how pyrene interacts with G-rich DNAs, we incorporated deoxyuridine nucleotide(s) with a covalently attached pyrene moiety (U^{py}) into a model system that forms parallel G-quadruplex structures. We individually substituted terminal positions and positions in the pentaloop of the c-kit2 sequence originating from the KIT proto-oncogene with U^{py} and performed a detailed NMR structural study accompanied with molecular dynamic simulations. Our results showed that incorporation into the pentaloop leads to structural polymorphism and in some cases also thermal destabilization. In contrast, terminal positions were found to cause a substantial thermodynamic stabilization while preserving topology of the parent c-kit2 G-quadruplex. Thermodynamic stabilization results from π–π stacking between the polyaromatic core of the pyrene moiety and guanine nucleotides of outer G-quartets. Thanks to the prevalent overall conformation, our structures mimic the G-quadruplex found in human KIT proto-oncogene and could potentially have antiproliferative effects on cancer cells.

GRAPHICAL ABSTRACT



INTRODUCTION

G-rich sequences are not randomly distributed throughout the human genome. They are concentrated in regions implicated in essential cellular processes such as initiation of DNA replication, transcription and telomere maintenance (1,2). Due to their biological significance, extensive effort has been directed towards eliciting a therapeutic response by manipulating non-B-DNA structures found in telomeric and promoter regions (3,4). The KIT receptor is a transmembrane protein that belongs to a family of growth factor receptors with intrinsic tyrosine kinase activity and participates, upon activation by endogenous ligands, in a variety of physiological processes such as cell growth, proliferation, migration and survival (5). Specific mutations and/or overexpression of the KIT proto-oncogene have been implicated in several human cancers including gastrointestinal stromal tumours, pancreatic cancer, melanoma and haematological neoplastic diseases (6,7). Due to its role in the pathogenesis of cancer, KIT is an attractive target for anti-cancer treatment (8–10). The human KIT proto-oncogene promoter contains three G-rich regions, c-kit1, kit* and c-kit2, which are capable of folding into non-B-DNA structures, specifically G-quadruplexes (11,12). These four-stranded struc-

*To whom correspondence should be addressed. Tel: +386 1 4760 240; Fax: +386 1 4760 300; Email: peter.podbevsek@ki.si

tures consist of stacked G-quartets, planar arrangements of four guanines held together by Hoogsteen hydrogen bonds organized around a central cation (13). Structures of G-quadruplexes are highly polymorphic and their fold is influenced by various factors such as oligonucleotide sequence, loop arrangements, crowding conditions and presence of other biomolecules. The ability of c-kit1, kit* and c-kit2 sequences to fold into G-quadruplexes has been demonstrated recently, and their high-resolution structures are available in the Protein Data Bank (14–19). The three G-rich sequences in KIT are clustered and spaced only few nucleotides apart. Interestingly, the function of each G-quadruplex unit is strongly affected by formation or disruption of the remaining two (20). Furthermore, kit* and c-kit2 interact and form a tandem G-quadruplex unit (21). The promoter segment comprising kit* and c-kit2 contains putative binding sites for Sp1 and AP2 transcription factors (12,22). Apart from binding to a double-stranded consensus motif, Sp1 can also bind to a G-quadruplex (22). Considering that Sp1 binding is critical for activity of the human KIT promoter (23), highly stable G-rich oligonucleotides mimicking G-quadruplexes from KIT could be used as decoys to sequester these proteins and modulate KIT expression.

During the last decade extensive effort has been directed towards developing a variety of DNA probes based on functionalized oligonucleotides (24). Approaches for manipulating promoter G-quadruplex structures *in situ* often utilize small molecules that effect their thermodynamic properties and/or interfere with potential protein–DNA interactions (13). G-quadruplexes are especially suitable for targeting with planar organic compounds with an extended aromatic core structure that can stack through π–π interactions on outer G-quartets (25,26). Unfortunately, development of new non-covalent binders is hindered by serious selectivity issues (27). Alternatively, polyaromatic groups can be covalently attached to short DNA oligonucleotides at desired position(s) and delivered to cells where they elicit a therapeutic response.

Pyrene is a polycyclic aromatic hydrocarbon, which versatility can be extended with chemical modifications. Thanks to its unique fluorescence properties, pyrene-modified oligonucleotides can act as microenvironment-sensitive fluorescent probes (28,29), excimer- or exciplex-generating molecules (30–36), nucleic acid complex intercalators (37–40) and π–π stacking hydrophobic moieties (41,42). Photophysics of pyrene-labelled compounds has been studied extensively, whereas only a limited number of studies focused on the structure of pyrene-conjugated nucleic-acid-based constructs. Recently, we have shown that incorporation of a pyrene-conjugated uridine building block into the thrombin binding aptamer (TBA) can lead to thermal stabilization and increased resistance against nuclease degradation under human serum conditions (43). Although TBA is often used as a model system, its antiparallel chair-like topology is unique. We have also shown that short pyrene-conjugated oligonucleotides can recover the G-quadruplex fold of G-rich oligonucleotides after oxidative damage (44). On the other hand, Xodo *et al.* utilized a pyrene-conjugated G-rich sequence derived from the KRAS oncogene as a stable G-quadruplex decoy for sequestering transcription factors (45). A reduction of pancreatic tumour xenografts in mice supported the decoy approach as a promising cancer therapy.

In the current study, we focus on the G12T/G21T mutant of promoter c-kit2 sequence, d(CGGGCGGGCGCT AGGGAGGGT), (hereby referred to as c-kit2) as a representative of a monomeric parallel G-quadruplex. In parallel G-quadruplex topologies with propeller loops, outer G-quartets are more exposed to the solvent compared to G-quadruplexes with diagonal or edge-type loops. We hypothesized that large exposed surface areas of guanine nucleobases would be ideal for stacking with polyaromatic moieties such as pyrene. Introduction of a pyrene-conjugated nucleotide, 5-(1-pyrenylethynyl)-2′-deoxyuridine (U^{py}), in c-kit2 was expected to increase thermal stability. However, as the pyrene moiety in U^{py} is covalently tethered to the uracil nucleobase, careful positioning of this bulky building block will likely to be required not to perturb the G-quadruplex structure. Positions in the structure with more degrees of freedom such as overhangs and long non-structured loops should tolerate substitutions with U^{py} without causing alterations in G-quadruplex topology. We individually substituted U^{py} at terminal positions and all positions of the pentaloop in the c-kit2 sequence and evaluated the effect of pyrene moieties on the G-quadruplex forming ability, structure and thermodynamic stability. U^{py} at all positions adjacent to the core of G-quadruplex was expected to stack on outer G-quartets and result in thermodynamic stabilization. However, substitutions in the pentaloop caused structural polymorphism. On the other hand, introduction of pyrene moieties to individual or both termini of the c-kit2 sequence resulted in highly stable G-quadruplex structures. Unexpectedly, pyrene moiety at the 3′ terminus featured increased dynamics that hindered detailed NMR analysis. Therefore, we had to bridge the gap by state-of-the-art simulation techniques to describe its complex structural properties with atomistic resolution.

MATERIALS AND METHODS

Sample preparation

The 5-pyrene-dU-CE phosphoramidite and other standard phosphoramidites were purchased from Glen Research. Isotopically labelled phosphoramidites were purchased from Cambridge Isotope Laboratories. All oligonucleotides were synthesized on a K&A Laborgeraete H-8 DNA/RNA Synthesizer employing standard phosphoramidite chemistry. Natural abundance oligonucleotides were prepared in DMT-on mode, deprotected with AMA solution (1:1 mixture of aqueous 30% ammonium hydroxide and 40% aqueous methylamine) at 65°C for 30 min and purified with Glen-Pak cartridges. During the purification process failed sequences were eliminated and the DMT was subsequently removed. The solvent was evaporated and samples were dissolved in 100 mM LiCl following desalting using FPLC. Site-specific 10% ^{15}N- and ^{13}C-labelled samples were prepared in DMT-off mode, deprotected with AMA solution at 65°C for 30 min, dissolved in 100 mM LiCl and purified with FPLC. All oligonucleotides were dissolved in a solution containing 10% D_2O and 20 mM KCl, 5 mM K-phosphate buffer (pH 7.0) with DNA concentration

ranging from 0.5 to 1.0 mM, if not stated otherwise. NMR titration of 100 μM c-kit2, c-kit2_py1 and c-kit2_py21 samples with the sodium salt of 1-pyrenesulfonic acid and hydroxypyrene was performed at 0.0, 0.5, 1.0, 2.0, 3.0 and 4.0 molar equivalents. NMR samples with different ratios of oligonucleotide and hydroxypyrene were prepared by dissolving dry hydroxypyrene in DMSO-d_6 and titrating 15 mM hydroxypyrene solution in DMSO-d_6 into an aqueous solution of the oligonucleotide. Concentrations of DMSO-d_6 used in this study did not cause any spectral differences, which could account for structural changes of DNA influenced by DMSO-d_6.

Oligonucleotide concentration was determined by measuring ultraviolet absorbance at 260 nm using the Varian CARY-100 BIO UV-VIS UV/VIS Spectrophotometer with 1.0 cm light path cells. An extinction coefficient of 202 200 M^{-1} cm^{-1} was calculated with the nearest neighbour method for the non-substituted c-kit2 oligonucleotide and used for all samples.

UV melting profiles

Samples for UV melting experiments were prepared by diluting NMR samples with a blank solution containing 20 mM KCl and 5 mM K-phosphate buffer (pH 7.0). UV melting experiments were performed on a Varian CARY 3500 UV-VIS spectrophotometer with the Cary Win UV Thermal program using 1.0 cm light path cells. Samples were heated/cooled at 0.5°C min^{-1} in the range of 10–95°C, and absorbance at 295 nm was measured at 0.5°C steps. UV melting experiments were replicated 5 × for each oligonucleotide. $T_{1/2}$ of each measurement was determined from the first derivation of A_{295} versus temperature plot. Final $T_{1/2}$ was calculated as the average of 5 measurements. The van't Hoff enthalpies (ΔH_m) were determined by fitting the slope of melting curves at $T_{1/2}$ and using the equation $\Delta H_m = -4RT_m^2(\delta\alpha/\delta T)$ (46).

Fluorescence spectroscopy

Fluorescence properties of 150 μM c-kit2, c-kit2_py1, c-kit2_py21 and c-kit2_py1/21. G-quadruplexes in 20 mM KCl and 5 mM K-phosphate buffer (pH 7.0) solution were measured on QuantaMaster (PTI) spectrofluorometer at 25°C using a quartz cuvette with 1.0 cm path length and excitation wavelength of 330 nm.

NMR experiments

NMR data were collected on Bruker AVANCE III HD and AVANCE NEO 600 MHz NMR spectrometers equipped with cryogenic probes in the temperature range from 0 to 80°C. For suppression of water signals, the excitation sculpting pulse sequence element was used. NMR spectra were externally referenced to Sodium trimethylsilylpropanesulfonate (DSS). Translation diffusion coefficients were obtained with the use of a stimulated echo experiment using bipolar gradients. DOSY was performed at 15°C due to reduced resonance overlap and a favourable population ratio. 1D ^{15}N-edited HSQC and 2D ^{13}C-edited HSQC experiments were performed with ^{15}N-, ^{13}C-labelled oligonucleotides. 2D NOESY spectra were acquired at mixing times 70, 100, 150, 200 and 250 ms, 2D TOCSY at mixing time 65 ms and 2D ROESY at mixing time 120 ms. ^{1}H-^{13}C HSQC spectra measured on natural abundance samples were used to distinguish between *syn* and *anti* conformations of glycosidic bonds. NMR spectra were processed and analysed using TopSpin (Bruker). Resonance assignment and integration were done with Sparky (NMRFAM) software (47).

Interproton distances were calculated from 250 ms NOESY spectra, which were chosen from NOE build-up curves with C9 H5-H6 distance used as a reference (2.479 Å). Based on calculated distances, cross-peaks were divided into four categories (strong, medium, weak, very weak), which corresponded to plateaus in applied flat bottom potentials (1.8–3.6 Å, 2.6–5.0 Å, 3.5–6.5 Å, 4.5–7.5 Å) during simulated annealing. Hydrogen bonds within G-quartets were restrained (>3.0 Å). Further, potassium cations were placed in between adjacent G-quartets and restrained to guanine O6 atoms (>4.0 Å) as a physically sound alternative to planarity restraints. Harmonic potential was applied above and below these ranges with a force constant of 20 kcal mol^{-1} $Å^{-2}$.

Flat bottom potentials for dihedral angles χ were -90° to 90° for *syn* and 90° to 270° for *anti* nucleotides. Restraints were also applied to backbone dihedral angles α (−120° to 120°), β (150° to 210°), γ (30° to 90°), δ (130° to 190°), ε (170° to 300°) and ζ (−120° to 120°). Employed force constant was 200.0 kcal mol^{-1} rad^{-2}. Sugar puckering was restrained via ν_0, ν_1, ν_2, ν_3 and ν_4 dihedral angles, which were calculated from pseudorotation phase set between 144° and 180° using standard AMBER procedure. Dihedral angle definition is shown in Supplementary Figure S1. All sugar puckers were determined to be of S-type (C2'-endo) based on 2D TOCSY NMR spectra.

Molecular dynamics

A simulated annealing protocol was used for c-kit2_py1 using GPUs and pmemd.cuda 18.0 and OL15 force field (48–51) with Generalized Born/Surface Area implicit solvent model (igb = 1). Additionally, U^{py} was fitted with one extra pseudo-torsion term to correctly describe rotation of pyrene moiety and uracil nucleobase with respect to the linker (Supplementary Data, Force Field Parameterization, Supplementary Figure S2). A starting structure was prepared by substituting U^{py} into position 1 of c-kit2 NMR structure (PDB ID 2KYP). Before each run, the structure was annealed at 1200 K for 200 ps using all restraints except NOE distance restraints that involved $U^{py}1$ to remove the bias from a single starting structure (Supplementary Figure S3). Simulation following annealing was 200 ps long and consisted of several phases: 5 K → 1200 K (1 ps), 1200 K (39 ps), 1200 K → 600 K (50 ps), 600 K → 300 K (50 ps), 300 K → 0 K (60 ps) and final geometry optimization (5 000 steps). In total, 1000 repeats were performed. Ten structures with the lowest sum of potential and restraint energy were further refined by molecular dynamics (MD) in the explicit solvent.

Each selected structure was immersed into a truncated octahedral box filled by the TIP3P water with potassium and chloride ions. Resulting concentration of KCl was approximately 0.14 mM and radius of the largest inscribed

sphere was about 24 Å. Temperature and pressure were kept at 300 K and 100 kPa, respectively. NOE and hydrogen bond distance restraints as well as χ torsion angles and torsion angles within the ribose ring were active during the refinement. After 1 μs of simulation, the systems were cooled to 0 K during 10 ns at a constant volume and then geometry optimized (5000 steps) leading to the final NMR ensemble. Alongside, a same set of simulations was performed without any restraints to assess their effect on the obtained structure.

Molecular dynamics simulations of c-kit2_py21 closely followed settings used during refinement of the NMR structures in the explicit solvent (vide supra). The initial configurations of c-kit2 and c-kit2_py21 systems were taken from the previous tests (Supplementary Data, Testing Molecular Dynamics, Supplementary Figures S4–S6). Collected trajectories were 1 μs long. Because exchange between states observed in NMR is far beyond the timescale of regular MD, we decided to investigate possible stacking modes of U^{py} in c-kit2_py21 using biased MD providing also free energies of sampled configurations. The bias was introduced along geometrical parameters (collective variables, CVs) developed by the analysis of the testing runs and are shown in Supplementary Figure S7 including used restraint. Biased MD was done in a modified pmemd program coupled with PMFLib (https://pmflib.ncbr.muni.cz) using CPUs due to complexity of employed CVs and PMFLib design.

Corresponding free energy surface was calculated by the adaptive biasing force (ABF) method (52) enhanced by the multiple-walker approach (MWA) (53). The calculated mean forces were integrated using Gaussian process regression (GPR) (54) to get the final free energy landscape. ABF/MWA simulation accumulated 3 μs of sampling, which ensured converged free energy. Standard deviations (errors) of the free energy were calculated by Gaussian process from standard deviations of mean forces. Statistical inefficiency due to correlation in time series was evaluated by optimizing a GPR hyperparameter, which was achieved by maximizing logarithm of marginal likelihood. Free energies are reported with confidence intervals provided at three standard deviations (99.7% confidence level). More details on the procedure are provided in (55).

Analysis of simulations

Trajectories were processed and analysed using the cpptraj module of AMBER and the CATs package. Further, CATs were connected with PMFLib and served for finer filtering of the ABF/MWA trajectories by values of collective variables and calculated free energies. Thermodynamic states corresponding to the local minima found on the free energy surface were represented by ensembles of structures sampled in the ABF/MWA trajectories that were ≤0.3 kcal mol^{-1} (0.5RT at 300 K) above the free energy minima. If a free energy minimum contained differing conformations, structures were clustered. Visual representation of each thermodynamic state was prepared by randomly selecting and partially optimizing a structure from these ensembles. Average structures could not have been used due to flexibility of loop nucleotides. All structures were rendered using VMD (56).

RESULTS

Position of U^{py} substitution influences c-kit2 fold

Terminal positions and all positions of the pentaloop in c-kit2 were individually substituted with U^{py} and the capability of constructs to form G-quadruplex was evaluated through examination of 1D ^{1}H NMR spectra (Supplementary Figures S8 and S9). Oligonucleotides c-kit2_py9, c-kit2_py10, c-kit2_py11, c-kit2_py12 and c-kit2_py13 (hereby referred to as loop U^{py} analogues) exhibited broad NMR resonances suggesting formation of multiple structures in solution and/or aggregation (Supplementary Figure S8). On the other hand, favourable spectral properties were attained with oligonucleotides c-kit2_py1, c-kit2_py21 and c-kit2_py1/21 (hereby referred to as terminal U^{py} analogues), in which U^{py} was substituted at the 5′ and/or 3′ terminus of c-kit2 sequence (Figure 1 and Supplementary Figure S9). Guanine imino resonances of c-kit2_py1 and c-kit2_py21 were unambiguously assigned with the help of ^{15}N-edited HSQC spectra acquired on partially (10%) residue-specifically ^{15}N-enriched oligonucleotides (Supplementary Figures S10 and S11). Imino region of the c-kit2_py1 spectrum with its 12 resonances indicates formation of a G-quadruplex structure with three G-quartet planes. Differences in linewidths amongst imino resonances of c-kit2_py1 ($\Delta\nu_{1/2} \approx 7$ Hz for G2 and 15 Hz for G6) suggest different levels of exposure to exchange with bulk solvent and/or increased local dynamics (vide infra). Moreover, a set of broader, low intensity signals (<10%) suggests the presence of an additional minor structure. The major and minor structures of c-kit2_py1 exhibit different translational diffusion coefficients of 9.3×10^{-11} and 7.4×10^{-11} $m^2\ s^{-1}$ at 15°C, respectively, which suggest their monomeric and dimeric nature (Supplementary Figure S12). ^{1}H NMR spectrum of c-kit2_py21 exhibits nine major guanine imino resonances in the range from δ 11.0 to 12.0 ppm and four lower intensity imino resonances in the range from δ 10.3 to 11.0 ppm. Interestingly, one major and two lower intensity imino resonances were assigned to G20 suggesting presence of three distinct conformations that are in slow exchange on the NMR time scale. Two upfield imino resonances of lower intensity in the spectrum of c-kit2_py21 were assigned to G4 and G8, while no imino resonance for G16 was observed at 25°C. Favourable spectral properties were also attained when c-kit2 was double substituted at positions 1 and 21 (Figure 1 and Supplementary Figure S9). Imino region of c-kit2_py1/21 spectrum contains two sets of resonances in the range from δ 10.2 to 11.9 ppm, which suggest presence of major (>85%) and minor structures (<15%). Imino resonances corresponding to the major structure were assigned to 11 G-quartet forming guanines, while no imino resonance for G16 was observed at 25°C. Interestingly, three imino resonances were assigned to G20 suggesting presence of three distinct conformations in the major structure of c-kit2_py1/21.

Robustness of the c-kit2 fold was evaluated by extending the parent sequence and terminal U^{py} analogues by up to three nucleotides in each direction. Additional nucleotides at the 5′- or 3′-end of c-kit2 resulted in only minor chemical shift changes of imino resonances indicating that the G-

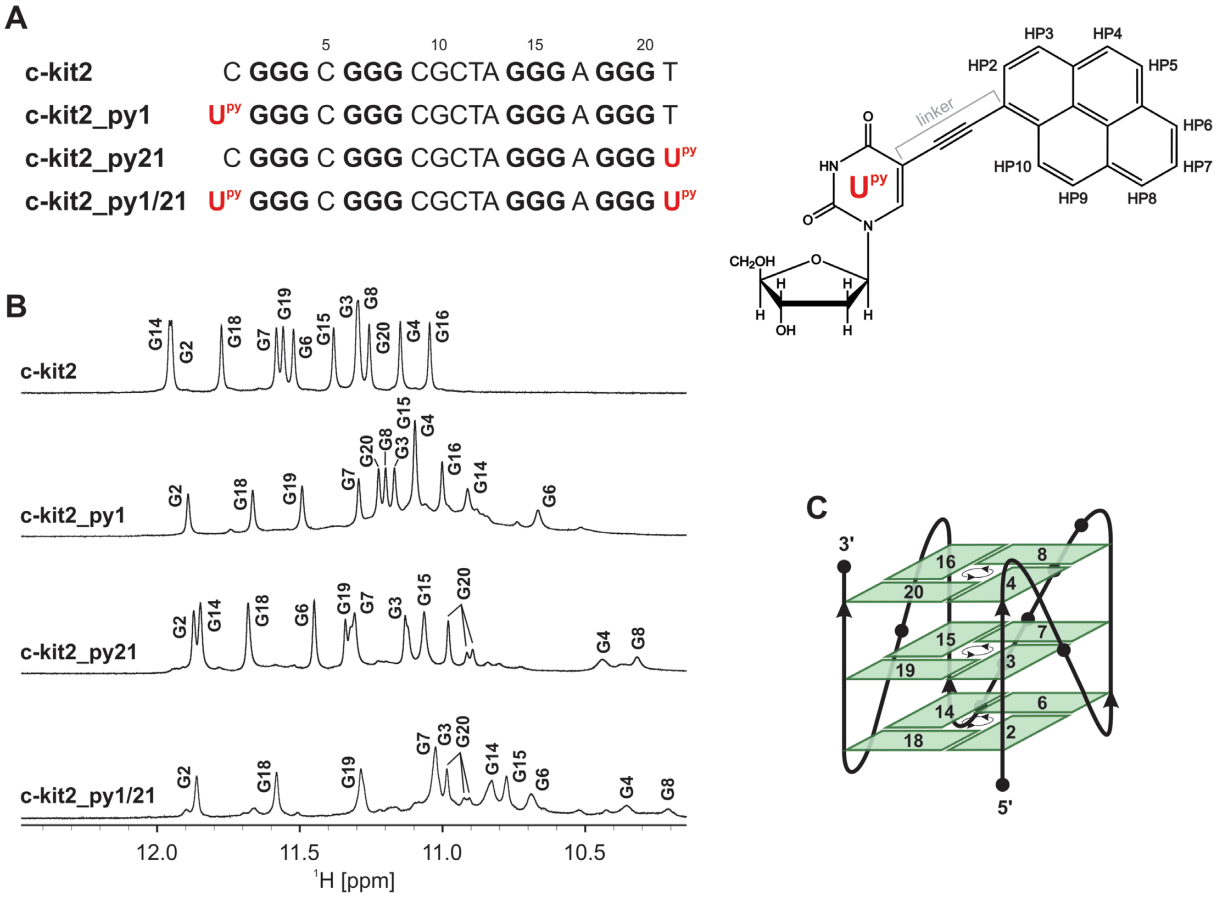


Figure 1. (**A**) Sequences of c-kit2 and terminal U^{py} analogues (U^{py} highlighted in red) (left). Guanines involved in G-quartet formation are in bold. Chemical formula of 5-(1-pyrenylethynyl)-2′-deoxyuridine (U^{py}) with atom numbering of pyrene moiety (right). (**B**) Imino regions of 1D ^{1}H NMR spectra of c-kit2 and terminal U^{py} analogues with assignment. Spectra were acquired in 90% H_2O and 10% D_2O, 20 mM KCl and 5 mM K-phosphate buffer, pH 7 on a 600 MHz NMR spectrometer. Oligonucleotide concentrations ranged from 0.7 to 1.0 mM. (**C**) Schematic representation of the c-kit2 promoter G-quadruplex topology. Hydrogen-bond directionality is indicated with arrows in the centre of each G-quartet.

quadruplex fold is not perturbed (Supplementary Figures S13 and S14). While extensions at 5′-end of all three terminal U^{py} analogues resulted in occurrence of multiple structures in solution and/or aggregation (Supplementary Figures S15–S17), extensions at 3′-end provided well-resolved 1D ^{1}H NMR spectra with a single major structure in solution (Supplementary Figures S18–S20). Interestingly, addition of nucleotide(s) at the 3′-end of U^{py}21 in c-kit2_py21 and c-kit2_py1/21 resulted in sharpening of upfield G4 and G8 imino resonances (Supplementary Figures S19 and S20). This suggests reduced dynamics of U^{py}21 most likely caused by stacking with 3′-flanking nucleotide(s).

UV absorbance and fluorescence spectra of c-kit2 and its terminal U^{py} analogues were acquired in the wavelength region characteristic for the pyrene moiety (Supplementary Figure S21). Expectedly, c-kit2 does not exhibit any fluorescent properties and any absorbance above 320 nm. On the other hand, all three terminal U^{py} analogues show absorption between 360 and 420 nm being two times higher for c-kit2_py1/21 than for c-kit2_py1 and c-kit2_py21. Respective fluorescence spectra exhibit a fluorescence maximum near 450 nm. Compared to c-kit2_py1/21, fluorescence maximum of c-kit2_py1 and c-kit2_py21 is blue-shifted by 13 and 5 nm, respectively. While c-kit2_py21 exhibits only slightly higher fluorescence intensity than c-kit2_py1/21, c-kit2_py1 exhibits the highest intensity of all terminal U^{py} analogues (Supplementary Figure S21).

U^{py} substitutions can result in substantial thermal stabilization

UV-melting experiments showed that all terminal U^{py} analogues exhibit higher thermal stability compared to parent c-kit2. While the melting profile of c-kit2 is classified as monophasic, melting profiles of c-kit2_py1 and c-kit2_py21 are more complex and display two transitions (Figure 2). Low temperature transitions observed at approximately 55.0 and 60.0°C for c-kit2_py1 and c-kit2_py21, respectively, exhibit small absolute changes in absorbance and may correspond to melting of minor species. High temperature transitions occur at 87.4 and 80.5°C for c-kit2_py1 and c-kit2_py21, which represents a 27.8 and 20.9°C increase compared to c-kit2, respectively (Supplementary Table S1). Expectedly, c-kit2_py1/21 exhibits an even larger gain in thermal stability. However, its melting temperature could not be extracted from UV data due to limitations related to the boiling point of the solvent (Supplementary Figure S22). We observed a hysteresis in the melting profiles of

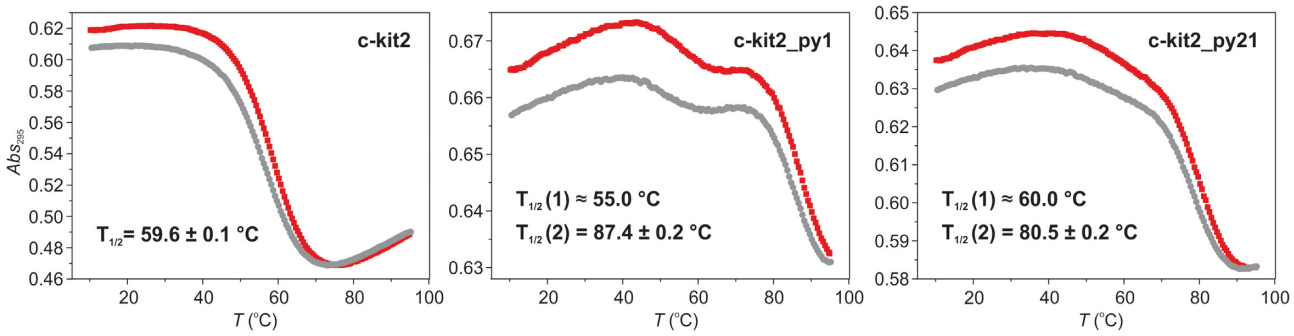


Figure 2. UV melting profiles of c-kit2, c-kit2_py1 and c-kit2_py21. Data were acquired for ∼10 μM DNA with 20 mM KCl and 5 mM K-phosphate buffer, pH 7. UV-melting experiments were initiated by equilibrating samples at 10°C for 10 min before being heated to 95°C (red curve) then after 10 min at 95°C the temperature was reduced to 10°C (grey curve). Temperature was ramped at 0.5°C/min and absorbance was followed at 295 nm. $T_{1/2}$ values were extracted form red curves.

the parent c-kit2 and all terminal U^{py} analogues indicating that the structural transition is slow relative to the temperature gradient. A simple thermodynamic analysis revealed that the pyrene moiety at the 5′-end (c-kit2_py1) is enthalpically favourable compared to the 3′-end (c-kit2_py21) by 9.2 kcal mol^{-1} (Supplementary Table S1).

Although none of the substitutions in the pentaloop of c-kit2 results in well-resolved NMR spectra (Supplementary Figure S8), three loop U^{py} analogues (c-kit2_py9, c-kit2_py10 and c-kit2_py13) exhibit higher $T_{1/2}$ values than c-kit2. The highest stabilization by 16.5 and 5.5°C is observed for U^{py} substitutions at positions 9 and 13, respectively (Supplementary Figure S22). Interestingly, these correspond to the 5′- and 3′-ends of the pentaloop in the parent G-quadruplex structure. On the other hand, U^{py} substitutions at positions 11 and 12 result in 1.5 and 15.0°C decrease in $T_{1/2}$ (Supplementary Figure S22).

Parallel fold is retained upon U^{py} substitutions

Imino, aromatic and anomeric resonances of c-kit2_py1 and c-kit2_py21 were assigned by analysing NOE connectivities (Figure 3 and Supplementary Figure S23). Assignment of guanine aromatic protons was confirmed through acquisition of ^{13}C-edited HSQC spectra using partially (10%) residue-specifically isotopically enriched oligonucleotides (Supplementary Figure S24). The majority of imino, aromatic and anomeric resonances of c-kit2_py1/21 could be assigned through sequential connectivities and by comparing their chemical shifts with those for c-kit2_py1 and c-kit2_py21. Sequential walks (H1'$_n$-H8/H6$_{n+1}$) in NOESY spectra of all terminal U^{py} analogues were interrupted at G4-C5, C5-G6, G16-A17 and A17-G18 steps due to large interproton distances resulting from single nucleotide double-chain-reversal loops. Furthermore, a lack of C9H6-G8H1', G10H8-C9H1' and C11H6-G10H1' cross-peaks in NOESY spectra of all terminal U^{py} analogues was attributed to dynamic behaviour of the pentaloop. NOESY spectra of c-kit2_py1 contained intense exchange cross-peaks between major and minor components, which complicated the signal assignment. However, they could be unequivocally distinguished from true NOE cross-peaks with the help of ROESY spectra (Supplementary Figure S25).

Examination of inter-nucleotide NOE connectivities between imino-imino and imino-aromatic protons revealed that topology of the parent c-kit2 G-quadruplex remained unperturbed despite introduction of U^{py} substitutions. NMR data show that all terminal U^{py} analogues adopt a parallel G-quadruplex topology consisting of G2-G6-G14-G18, G3-G7-G15-G19 and G4-G8-G16-G20 quartets (Figure 1C). All guanines forming the G-quadruplex core adopt *anti* glycosidic conformations and G-quartets exhibit counter-clockwise hydrogen bond donor-acceptor directionalities. The parallel G-strands of terminal U^{py} analogues form four medium width grooves and are linked by three double-chain-reversal loops.

c-kit2_py1 features a single stacking mode of U^{py}

^{1}H NMR resonances of the pyrene moiety in the monomeric G-quadruplex adopted by c-kit2_py1 were found in the range from δ 7.5 to 8.4 ppm. Although pyrene resonances are heavily overlapped with aromatic resonances of nucleobases, their unique spin-spin coupling network (resulting in eight doublets and one doublet of doublet) allowed us to unequivocally identify and assign them with TOCSY and DQF-COSY experiments. An intense H6-H1' cross-peak in a NOESY spectrum acquired with a short mixing time of 70 ms suggested the *syn* conformation for U^{py}1, whereas remaining nucleotides in c-kit2_py1 adopt *anti* conformations. The *syn* conformation of U^{py}1 is supported by a higher-frequency shift of its C6 resonance ($\delta^{13}C \approx 145$ ppm) with respect to $\delta^{13}C \approx 135$ ppm observed for guanine nucleotides (Supplementary Figure S26) (57).

Quantum mechanical calculations revealed that rotation of the pyrene moiety with respect to the uracil nucleobase of U^{py}1 is not free and two coplanar conformations (M1 and M2) are preferred (Supplementary Figure S2). Conformation M1 is characterized by a short distance between U^{py}1H6 and HP2. Rotation of the pyrene moiety with respect to the uracil nucleobase by 180° results in conformation M2 with U^{py}1H6 facing towards HP10 (see Figure 1). Observation of the U^{py}1H6-HP10 NOE cross-peak in NMR spectra showed that U^{py}1H6 is positioned closer to HP10 than to HP2, which suggest that U^{py} in c-kit2_py1 adopts conformation M2. This experimental observation does not agree with quantum mechanical calcula-

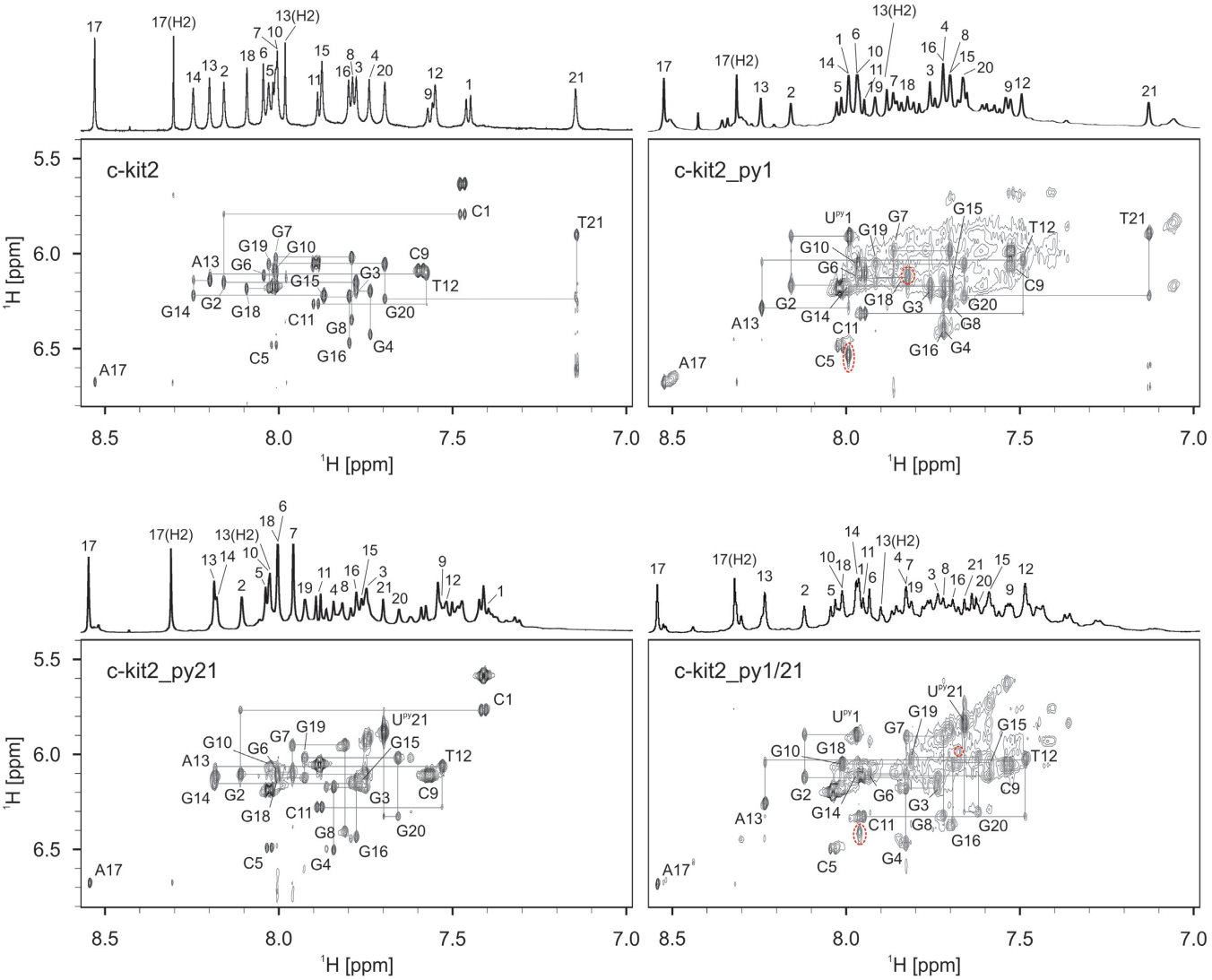


Figure 3. Aromatic-anomeric regions of 2D NOESY spectra (τ_m = 250 ms) with traced sequential walks for c-kit2 and terminal U^{py} analogues. The intra-nucleotide H8/H6(n)-H1'(n) cross-peaks are labelled. Red dashed circles in the spectra of c-kit2_py1 and c-kit2_py1/21 mark exchange cross-peaks between different species. Spectra were acquired in 90% H_2O and 10% D_2O, 20 mM KCl and 5 mM K-phosphate buffer, pH 7 on a 600 MHz NMR spectrometer. Oligonucleotide concentrations ranged from 0.7 to 1.0 mM.

tions, which showed slight energetic preference for conformation M1 most likely due to the stabilizing contact HP10-O4 and the absence of other supramolecular interactions in our isolated model. Besides U^{py}1H6-HP10 cross-peak, we observed five inter-nucleotide NOE contacts involving pyrene protons, which defined orientation of the pyrene moiety and its stacking with the G2-G6-G14-G18 quartet (Supplementary Figures S23 and S27).

A solution-state structure of c-kit2_py1 was determined on the basis of NMR observables using a restrained simulated annealing protocol. A series of 1000 simulations resulted in a well converged set of structures. Ten structures of c-kit2_py1 with the lowest sum of potential and restraint energy (NMR ensemble) were selected and refined via explicit solvent simulation before further analysis (Figure 4A). c-kit2_py1 forms a well-defined G-quadruplex consisting of three G-quartets, parallel strand orientations and propeller-type loops. The C9-G10-C11-G12-A13 pentaloop in c-kit2_py1 (average RMSD of pentaloop 1.62 ± 0.86 Å) retains its propensity for internal stacking. Due to the U^{py} substitution at position 1, the C1:A13 non-canonical base pair, observed in the parent c-kit2, cannot be formed. Consequently, the G2-G6-G14-G18 quartet is available for stacking interaction with the pyrene moiety and uracil nucleobase of U^{py}1. The pyrene moiety was found to be located predominantly above G6 and G14, while the uracil nucleobase is located mostly above G18 (Figure 4B). The value of U^{py}1 glycosidic torsion angle χ averages to 34° for ten structures of the NMR ensemble, which agrees with the experimentally determined *syn* conformation. Interestingly, both pyrene moiety and uracil nucleobase of U^{py}1 exhibit lateral motion above the G2-G6-G14-G18 quartet (Figure 4C). All sugars in c-kit2_py1 adopt the S-type conformation (Supplementary Table S2).

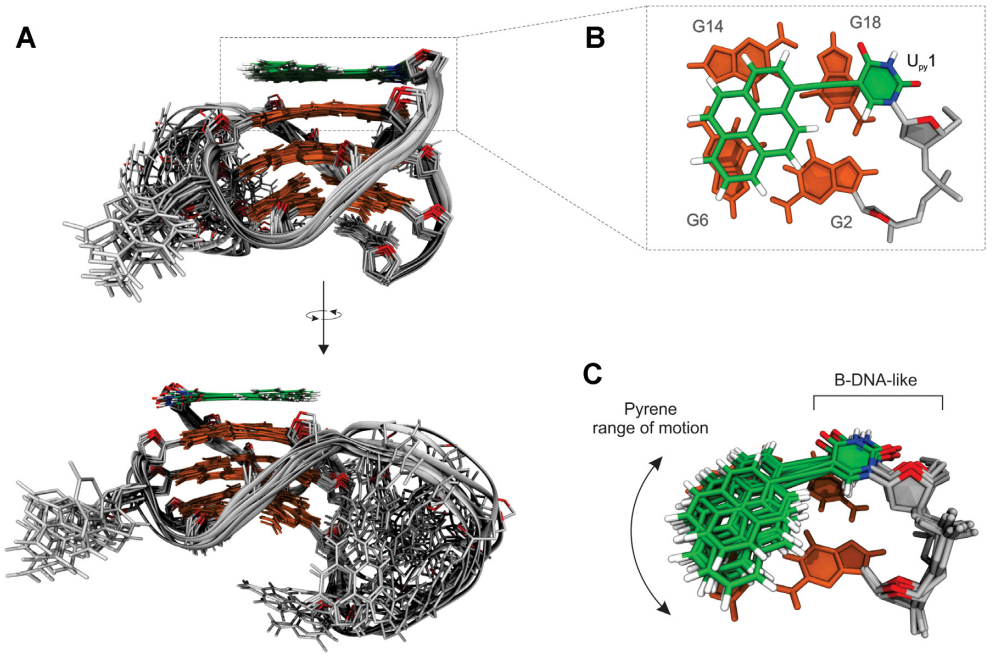


Figure 4. Solution-state structure of c-kit2_py1. (**A**) Superposition of 10 lowest energy structures from the NMR ensemble refined in the explicit solvent. (**B**) Details of $U^{py}1$ stacking on the top of G2-G6-G14-G18 quartet. (**C**) Superposition of $U^{py}1$ conformations demonstrating range of motion sampled by the pyrene moiety, while deoxyuridine linker adopts a B-DNA-like conformation. $U^{py}1$ is shown in green and G-quartets in orange. Loops and backbone are in grey with sugar O4' atoms highlighted in red.

Structures of the c-kit2_py1 NMR ensemble were also subjected to 1 μs long unrestrained explicit solvent MD simulations. Stability of the systems without restraints suggested good convergence of the NMR ensemble (Supplementary Figure S28). Fast $U^{py}1$ *syn* to *anti* transitions were the only observed difference compared to the distance restrained trajectories. The less populated (23%) *anti* conformation results from rotating the sugar 5′ terminus of $U^{py}1$ towards G2 (Supplementary Figure S28C) and might represent a short-lived higher energy state. Overall, the $U^{py}1$ *syn–anti* ratio in the unrestrained simulation (χ averaged to 58° for *syn* and -164° for *anti*) is in line with the NOE-based assessment (Supplementary Figure S29). All sugars averaged to the S-type conformation but with a wider range of values (Supplementary Table S3). Elimination of restrains also resulted in pronounced dynamic behaviour of the pentaloop during MD simulations (average RMSD of pentaloop 3.65 ± 0.55 Å). We observed transient stacking of the 3′-end residue of the pentaloop (A13) with the pyrene moiety and of the 5′-end residue (C9) with the G4-G8-G16-G20 quartet (Supplementary Figure S28A). A nucleus-independent chemical shift (NICS) approach showed that G6 and G14 imino resonances shift to lower frequencies upon stacking of the pyrene moiety (Supplementary Figure S30 and Supplementary Table S4). This agrees with NMR data where imino resonances of G6 and G14 in c-kit2_py1 were found shielded with respect to those in c-kit2 (Figure 1B). However, not so clear agreement was seen for the ensemble refined with restraints. This suggests that the time-averaged restraints might not capture well transient positions of pyrene. Although, exact values of NMR shifts could not be reproduced due to limitation of the computational method (Supplementary Table S4).

U^{py} stacking modes in c-kit2_py21 are in dynamic equilibrium

Presence of three distinct stacking modes of $U^{py}21$ with the G4-G8-G16-G20 quartet was identified in ^{1}H NMR spectra of c-kit2_py21 measured above 40°C. These exhibited three sets of imino resonances for G20, G16 and G8, and two (possibly overlapped) resonances for G4 (Figure 5A and Supplementary Figure S31A). Additionally, we observed doubling of G3 and G19 imino resonances from the central G-quartet (Supplementary Figure S31A). Two out of three stacking modes exhibited a downfield C6 chemical shift ($\delta^{13}C \approx 145$ ppm) for the $U^{py}21$ residue, which suggest its *syn* glycosidic conformation (57). On the other hand, the last $U^{py}21$C6-H6 cross-peak exhibited a chemical shift of $\delta^{13}C \approx 135$ ppm suggesting its *anti* glycosidic conformation (Figure 5B).

Analysis of NOESY spectra (Figure 3) revealed that c-kit2_py21 adopts the same parallel G-quadruplex topology consisting of three G-quartets as the parent c-kit2. However, all imino resonances of the G4-G8-G16-G20 quartet were found to be shifted upfield in c-kit2_py21 compared to c-kit2, which suggest that all distinct conformations of $U^{py}21$ feature a shielding effect of the stacked pyrene moiety. This was supported by a single NOE contact between the pyrene moiety and G8H2'. Unfortunately, dynamic equilibrium between distinct stacking modes of $U^{py}21$ resulted in poor chemical shift dispersion and lack of NOE contacts between $U^{py}21$ and the G-quadruplex core. Consequently, position of the pyrene moiety with respect to the core of the c-kit2_py21 G-quadruplex could not be determined on the basis of NMR observables. In order to obtain a more detailed picture on dynamic processes in c-kit2_py21, we built a model from the parent

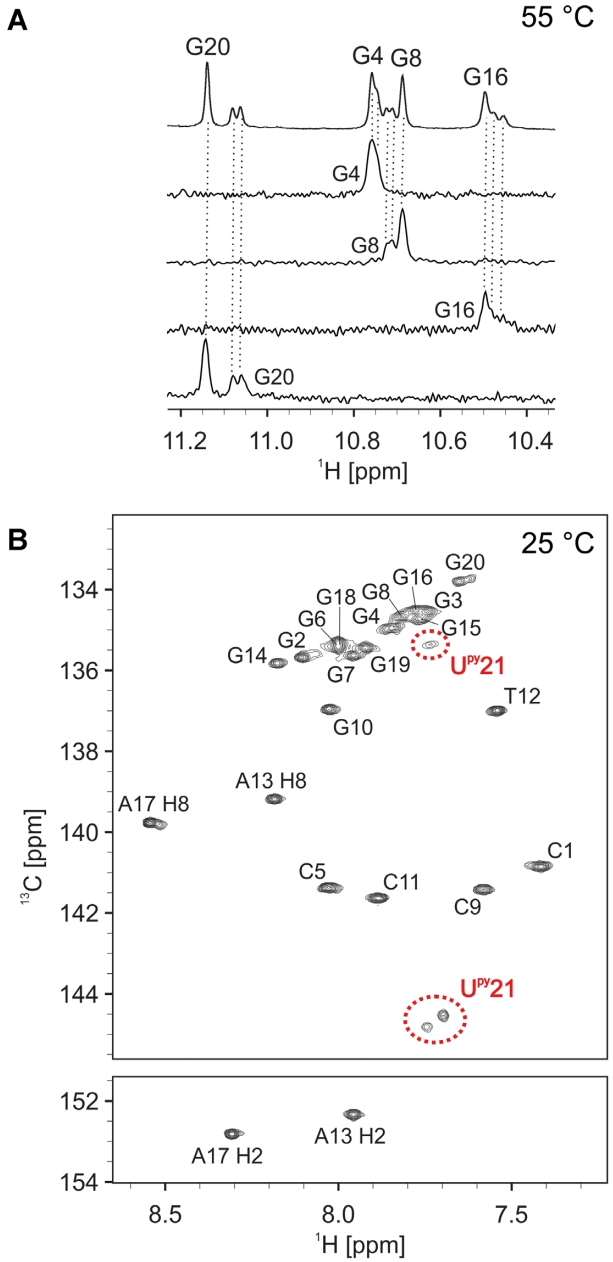


Figure 5. (**A**) Imino region of G4-G8-G16-G20 quartet of a 1D ^{1}H NMR spectrum of c-kit2_py21 with 1D ^{15}N-edited HSQC spectra below. Spectra were acquired at 55°C. Oligonucleotide concentration varied from 0.6 to 0.9 mM. (**B**) 2D ^{1}H-^{13}C HSQC spectrum of c-kit2_py21 acquired at 25°C with assignment. Three signals observed for U^{py}21 are marked by red dashed circles. The spectrum was acquired with 90% H_2O and 10% D_2O, 20 mM KCl, and 5 mM K-phosphate buffer, pH 7, on a 600 MHz NMR spectrometer.

c-kit2 structure (PDB ID: 2KYP) with U^{py}21 positioned above the G4-G8-G16-G20 quartet and performed a set of MD simulations (20 μs in total, Supplementary Figures S4–S6). G-quadruplex structure of c-kit2_py21 was stable and the pyrene moiety remained stacked during all simulations. While orientation of the pyrene moiety was found to be dependent on the starting conformation and did not change during MD (suggesting a high energy barrier), the uracil-sugar moiety was able to rotate with respect to the linker. We found a geometrical description, i.e. collective variables, discriminating both rotation of pyrene moiety and rotation of uracil-sugar moiety with respect to the linker (Supplementary Figure S7).

Next, we explored multiple conformations of U^{py}21 in c-kit2_py21 via biased MD. We calculated the free energy surface (Supplementary Figure S32) from which we extracted a minimum free energy path (Figure 6). Interestingly, five free energy minima (thermodynamic states) were predicted on the path (**1**, **2**, **2***, **1*** and **3** in Figure 6), in contrast to three distinct conformations observed experimentally by NMR (Figure 5). However, both computational and experimental results suggested that one state is preferred. Calculated free energies transferred to 94% Boltzmann population of state **1** at 300 K.

States **1** and **2** have the pyrene moiety positioned mostly above G16 (Figure 6), which agrees with shielding of its imino resonance with respect to c-kit2. U^{py}21 backbone conformation in **1** and **2** remains relaxed B-DNA-like as in the case of c-kit2_py1. State **1** differs from **2** by a 180° rotation of pyrene moiety around the linker. Interestingly, the most stable state **1** has an identical U^{py}21 conformation as the quantum mechanical global minimum in vacuo, which was not the case for c-kit2_py1. States **1*** and **2*** are distinguished from **1** and **2**, respectively, by rotating uracil-sugar moiety with respect to the linker by 180°. This rotation brings the pyrene moiety above G8 (Figure 6) and explains the NMR shielding of the G8 imino resonance of c-kit2_py21. State **3** differs from the rest by featuring a non-planar pyrene-uracil arrangement that is stabilized by interaction of the uracil nucleobase with the preceding phosphate (Figure 6). However, population of **3** is most likely insignificant due to its higher energy.

Surprisingly, a detailed structural analysis revealed that states **1*** and **2*** are represented by two different conformations, which are undistinguishable by employed collective variables. Without biasing this degree of motion, U^{py}21 sugar spontaneously flipped on the backbone and started to interact with the preceding phosphate via its 3′-OH group. **1*** and **2*** ensembles were split between non-flipped and sugar-flipped conformations with approximate ratios of 40:60 and 60:40, respectively. This suggests a low energy barrier for the interconversion and similar free energies. The sugar-flip brings the pyrene moiety above G4, which agrees with shielding of the G4 imino resonance in NMR spectra of c-kit2_py21. If only B-DNA-like structures **1** and **2** were populated, G4 would not experience shielding. Furthermore, substantial backbone reorganization in **1*** and **2*** could explain doubling of G19 and G3 imino resonances (Supplementary Figure S31A).

In total, we identified seven different stacking modes of U^{py}21 as representative structures of the calculated free energy minima (Figure 6). However, sets of NMR resonances could not be assigned to individual stacking modes of U^{py}21 due to complex averaging of all motions.

Observation of three sets of imino resonances for G20, G16, G8 and G4 in temperature-dependent 1D ^{1}H NMR spectra of c-kit2_py1/21 suggested preservation of distinct dynamics of U^{py} on opposite sides of the G-quadruplex core in the double-substituted terminal U^{py} analogue (Sup-

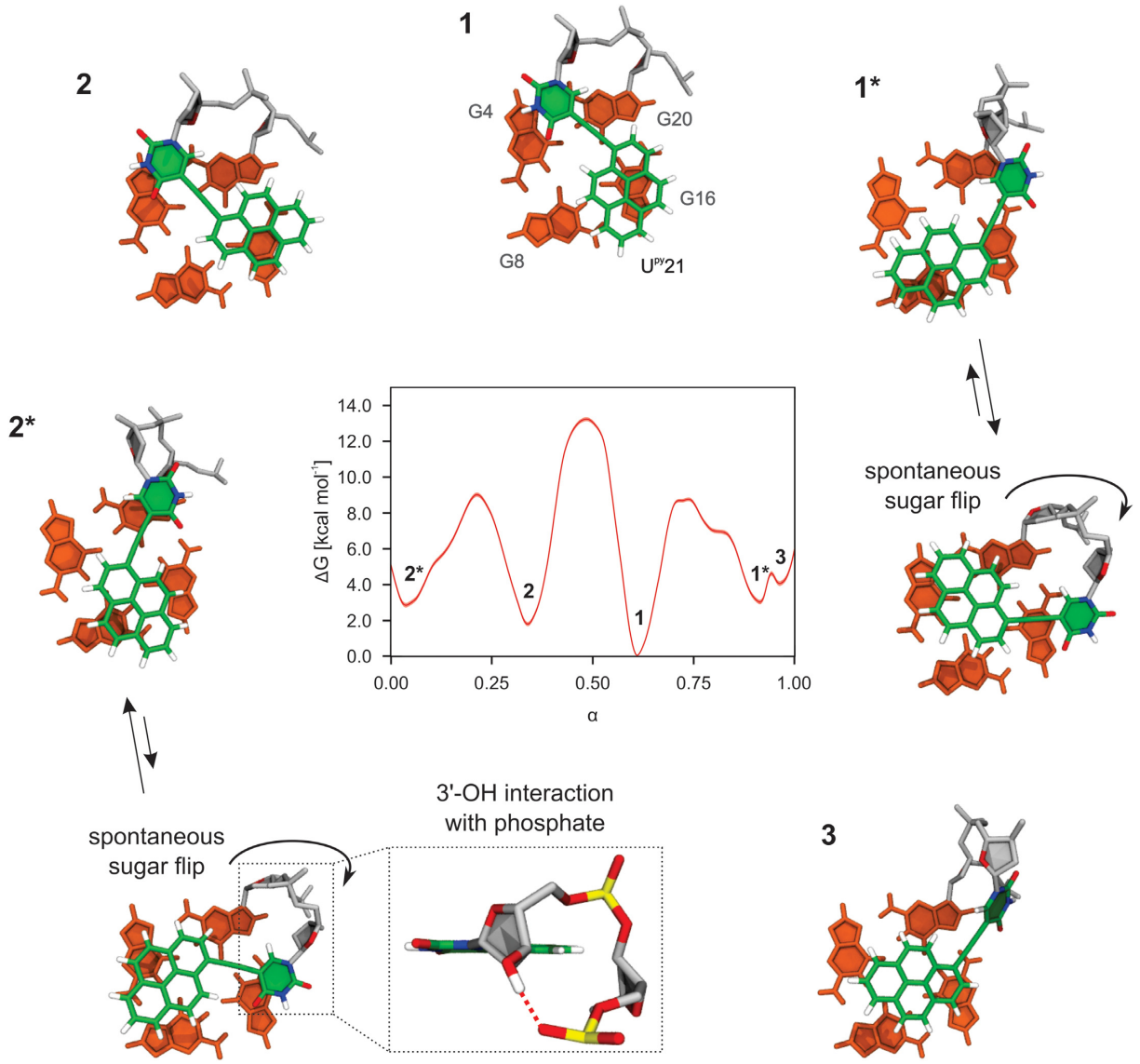


Figure 6. Minimum free energy path connecting five detected thermodynamic states **1**, **1***, **2**, **2*** and **3** with representative structures showing different stacking modes of pyrene moiety in c-kit2_py21. Minimum free energy path was extracted from the free energy surface shown in Supplementary Figure S32. Dimensionless parameter α describes position along the path.

plementary Figure S33). This was corroborated by overlaying ^{13}C-edited HSQC spectra of c-kit2_py1, c-kit2_py21 and c-kit2_ py1/21 (Supplementary Figure S34). While the chemical-shift dispersion of cross-peaks corresponding to pyrene moieties showed to be almost identical in both carbon and proton dimensions in all three HSQC spectra, the majority of cross-peaks in c-kit2_ py1/21 was slightly shifted towards lower frequencies in the ^{1}H dimension compared to c-kit2_py1 and c-kit2_py21 (Supplementary Figure S34).

Introduction of U^{py}21 reduces structural heterogeneity

The c-kit2 G-quadruplex dimerizes with time, while c-kit2_py1 and c-kit2_py21 remain unaffected even after several months in solution (Supplementary Figure S35). Compared to the sequence found in the human KIT gene, the c-kit2 oligonucleotide and its U^{py} analogues contain two G-to-T substitutions, which are known to reduce heterogeneity and suppress formation of dimeric species. Interestingly, we found that introduction of U^{py} at position 21 of the genomic KIT sequence eliminates heterogeneity and formation of dimeric species without the need for any G-to-T substitution (Supplementary Figure S36). On the other hand, samples with U^{py} at position 1 are highly polymorphic in the absence of G-to-T substitutions (Supplementary Figure S36).

Free pyrene exhibits a binding preference for G-quadruplex structure

Titration of a c-kit2 G-quadruplex solution with a sodium salt of 1-pyrenesulfonic acid resulted in shifting and broadening of several c-kit2 imino resonances in 1D NMR spec-

tra (Supplementary Figure S37). Changes were detected for the G4-G8-G16-G20 quartet with a low-frequency shift of all guanine imino resonances and a maximum of 0.07 ppm for G4 (Supplementary Figure S40). On the other hand, minor high-frequency shifts were observed for methyl and aromatic resonances of T21 and aromatic resonances of G20 and C9. No new signals suggesting formation of new G-quadruplex species arose during the titration. Similar behaviour was observed during titration of c-kit2_py1. We detected broadening and shielding of imino resonances from the G4-G8-G16-G20 quartet, and deshielding of T21CH_3, T21H6 and G20H8 resonances during titration of c-kit2_py1 (Supplementary Figure S38). Absolute values of chemical shift differences were smaller than in the case of c-kit2 with the maximum of 0.05 ppm for G4 and G20 (Supplementary Figure S40). Surprisingly, there were no observable changes in NMR resonances of the c-kit2_py21 G-quadruplex during its titration with the sodium salt of 1-pyrenesulfonic acid (Supplementary Figures S39 and S40).

Since 1-pyrenesulfonic acid is negatively charged we have also tested the interaction with the neutral hydroxypyrene. Titration of c-kit2 and c-kit2_py1 solutions with hydroxypyrene resulted in significant shifting and broadening of imino resonances assigned to the G4-G8-G16-G20 quartet (Supplementary Figures S41 and S42). On the other hand, no shifting and broadening of imino resonances was observed during titration with c-kit2_py21 (Supplementary Figure S43). However, we detected sharpening of G4 and G8 imino resonances as well as appearance of G16 imino resonance with increasing concentration of hydroxypyrene in the c-kit2_py21 solution. Similar sharpening was observed in ^{1}H NMR spectra measured above 30°C (Supplementary Figure S31A).

Our data suggest that 1-pyrenesulfonic acid and hydroxypyrene interact with the c-kit2 G-quadruplex and its analogues only when the G4-G8-G16-G20 quartet is available for interaction. While stacking of the 1-pyrenesulfonic acid and hydroxypyrene to the G2-G6-G14-G18 quartet might be energetically favourable, this G-quartet in c-kit2 and c-kit2_py21 is blocked by the C1:A13 non-canonical base pair or already stacked with the covalently attached pyrene moiety in c-kit2_py1.

DISCUSSION

Polyaromatic moieties can be employed for modulating G-quadruplex properties via their stacking with G-quartets. We recently showed that a single residue substitution with a pyrene conjugated nucleotide (U^{py}) in TBA can cause a moderate increase in the thermal stability of the G-quadruplex structure (43). However, the chair-like topology of TBA consisting of two G-quartets is unique and those results cannot be generalized and easily extrapolated to modulate properties of promoter G-quadruplexes. On the other hand, the genomic c-kit2 sequence originating from the KIT proto-oncogene exhibits features typical for G-quadruplex forming sequences in human promoter regions. Genomic c-kit2 contains G-tracts of different lengths and additional Gs in loop forming regions, which result in multiple G-quadruplex structures in solution (58). Different mutations were employed in order to reduce the level of structural polymorphism in c-kit2. Herein, we focused on G12T/G21T mutant of genomic c-kit2 sequence forming a single monomeric three-quartet G-quadruplex, as a representative of parallel G-quadruplex structures.

We individually incorporated U^{py} at all positions in the pentaloop and terminal positions of the c-kit2 sequence. While some of the loop substitutions resulted in destabilization of the G-quadruplex structure, most of them caused an increase in thermal stability. Due to extended structure of U^{py}, stacking of pyrene moiety is less likely if U^{py} is substituted in the middle of the loop. Expectedly, a notable stabilization effect was achieved, when U^{py} was placed at 5′- or 3′-ends of the pentaloop. We believe that this is due to partial and structurally heterogenous π–π stacking interactions of the pyrene moiety with adjacent G-quartets. A substantial increase in thermal stability of up to 27.8°C was achieved when terminal C1 and T21 were individually substituted with U^{py}. Initial observations suggested that pyrene moieties in c-kit2_py1 and c-kit2_py21 G-quadruplexes are exposed to similar environments, which resulted in only minor differences in fluorescent properties and thermal stability. However, a detailed analysis revealed major differences in structural dynamics of U^{py} between the two terminal analogues. While c-kit2_py1 appeared structurally rigid with one well-defined stacking mode of the pyrene moiety, c-kit2_py21 demonstrated U^{py} in multiple conformational states. In total, we identified seven ensembles with distinct stacking modes of U^{py}21. Although one stacking mode is predominant, remaining six likely serve as a pool of structures, which play a role in the stabilization of the c-kit2_py21 G-quadruplex. Differences in thermal stability of c-kit2_py1 and c-kit2_py21 might stem from different efficacy of π–π stacking interactions between the pyrene moiety and adjacent guanine nucleobases. Considering its size, the pyrene moiety can effectively stack on two guanines of an adjacent G-quartet to maximize π–π stacking interactions. While the pyrene moiety in c-kit2_py1 is located above two guanine nucleobases, this is not the case in c-kit2_py21 where it was found to stack only on one guanine of an adjacent G-quartet. Therefore, less efficient stacking in c-kit2_py21 results in slightly lower thermal stability and larger structural heterogeneity compared to c-kit2_py1.

Efficient stacking of the pyrene moiety requires a planar stacking partner. Therefore, we analysed planarity of G-quartets in c-kit2, c-kit2_py1 and c-ki2_py1 by comparison of root-mean-square deviation (RMSD) from a plane, which was best-fitted individually to each G-quartet. The 5′ G-quartet was uniformly found to be the most planar, whereas the 3′ G-quartet is the least planar in all three systems (Supplementary Figure S44, c-kit2_py1: average RMSD of 5′ G-quartet 0.28 ± 0.06 Å, middle G-quartet 0.35 ± 0.06 Å, 3′ G-quartet 0.50 ± 0.07 Å). Positioning of a fully planar, fused ring system of pyrene next to a more planar 5′ G-quartet in c-kit2_py1 was found to be advantageous and correlated with tighter stacking of U^{py}1. On the other hand, the less planar 3′ G-quartet in c-kit2_py21 probably promoted conformational tumbling and various stacking modes of U^{py}21. We believe that the contrast between structural dynamics of U^{py}1 and U^{py}21 might stem from an intrinsic asymmetry of c-kit2 G-quartets. This way U^{py} acts as a probe for local G-quadruplex dynamics, which

is true especially for c-kit2, where outer G-quartets are exposed and U^{py} interactions with propeller loops are minimized. This is a vice-versa effect to the binding of ligands that comprises unfused aromatic rings to G-quadruplexes, where ligand planarity is key for efficient stacking (59). Similar conclusions can be drawn from comparing U^{py} substitutions in TBA and c-kit2. While TBA G-quartets are protected by two TT and one TGT loop, bulky U^{py} was able to sandwich between G-quartets and loop residues. However, the maximum thermal stabilization achieved for TBA was 6°C, whereas c-kit2 exhibits up to 29°C increase. In the case of TBA, stabilization gained by U^{py} stacking appears to have been largely lost by steric clashes with the loop nucleotides. Modulation of G-quadruplex properties via covalent attachment of polycyclic aromatic hydrocarbon molecules, therefore, needs to consider mostly local geometry as the stability of different conformations depends on the balance between stacking, deformation of the linker residue and steric interactions.

Incorporation of two pyrene moieties into the TBA sequence results in disruption of the hybrid G-quadruplex structure or oligomerization facilitated by pyrene–pyrene interactions (43). On the other hand, the parallel c-kit2 G-quadruplex can accommodate two pyrene moieties at opposite sides of the G-quadruplex, which does not perturb the overall topology. Sandwiching the G-quadruplex core between two pyrene moieties provided additive gain in thermal stabilization compared to single substitutions, while preserving distinct dynamic behaviour of polyaromatic groups on opposite sides of the structure. c-kit2_py1/21 was expected to show propensity for dimerization/oligomerization via pyrene–pyrene stacking. Surprisingly, we did not observe any higher-order structure formation in c-kit2_py1/py21 despite the high concentration needed for NMR studies (mM range). Additionally, we showed that single incorporation of U^{py} at position 21 of genomic c-kit2 region (with no G-to-T substitution) prevents formation of the dimeric species. Therefore, covalent attachment of a polyaromatic group at the appropriate position in the sequence can restrict formation of multimeric G-quadruplex species and reduce structural heterogeneity as an alternative to guanine substitutions or modulation of sample conditions.

KIT mutations and overexpression not only cause dysfunction leading to tumours (60–63), but aberrant expression of KIT is associated with poor prognosis in patients with lung cancer (64). Unfortunately, regulation of KIT expression is complex and relatively little is known about regulatory factors that are involved. Clinical drugs directly targeting c-KIT kinase generally become less effective due to the appearance of c-KIT secondary mutations (8,65). Therefore, alternative ways of regulation of KIT expression are needed. Higher-order arrangement of the entire G-rich domain in the core-promoter region of the KIT gene is expected to regulate protein recruitment, in particular transcription factors. The specificity protein 1 (Sp1) is a prototype member of a family of related transcription factors that modulate gene expression (66). Sp1 with the binding site located in the promoter portion comprising c-kit2 and kit*, generally functions as a transcriptional activator, which contributes to the proliferative and metastatic tumour phenotypes (66). Considering that Sp1 recognizes the G-quadruplex DNA structure as an alternative binding motif (22), we believe that modified oligonucleotides mimicking KIT G-quadruplexes have potential to saturate Sp1-binding sites and prevent the protein from binding to the promoter in genomic DNA. After transfer into the cells, our engineered oligonucleotides could down-regulate KIT transcription by sequestration of Sp1 and consequently decrease cell growth. Besides that, interaction of so-called G4 decoys with transcription factors might expand on our knowledge about the processes involved in regulation of KIT proto-oncogene transcription. Using covalently attached ligands could reduce frequently occurring off-targeting of multiple G-quadruplexes commonly observed for non-covalent therapeutic binders. Spatial distribution of fluorescent G-quadruplexes and their complexes with transcription factors could be visualized in cells with fluorescent microscopy. Furthermore, detailed structural data could also prove useful for studying fluorescence with high-level ab initio computational methods.

In conclusion, using U^{py} substitutions we have achieved a considerable thermal stabilization of the c-kit2 G-quadruplex while preserving its original topology. Three substituted c-kit2 G-quadruplexes varying in the accessibility of outer G-quartets expand the portfolio of possible interactions with biological targets. Additionally, high thermal stability of our structures allows their use as platforms for attachment of additional functional groups (24). We believe that development of approaches to improve stability of genomic G-quadruplexes and of methods for their characterization could bring improved therapeutics for treatment of cancer or other G-quadruplex related disorders.

DATA AVAILABILITY

Atomic coordinates and list of chemical shifts for the c-kit2_py1 NMR structures have been deposited with the Protein Data bank under accession number 7NWD.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are very grateful to Dr. Petr Kulhanek for numerous consultations of the computational part of the work and to Dr. Jan Novotny for fruitful discussions about some NMR experiments. We would like to thank the CERIC-ERIC Consortium for access to their experimental facilities.

FUNDING

This work was supported by the European Programme H2020 MSCA ITN [grant number 765266-LightDyNAmics project]. The authors also acknowledge financial support from the Slovenian Research Agency [grants P1-0242 and J1-1704] and the project CEITEC 2020 [Grant No. LQ1601] with financial support from the Ministry of Education, Youth and Sports of the Czech Republic (MEYS CR) under the National Sustainability Programme II. CIISB, Instruct-CZ Center of

Instruct-ERIC EU consortium, funded by MEYS CR infrastructure project LM2018127, is gratefully acknowledged for the financial support of the measurements at the Core Facility Josef Dadok National NMR Center. The computational resources were supported by MEYS CR from the Large Infrastructures for Research, Experimental Development and Innovations project "e-Infrastructure CZ" [LM2018140]. Funding for open access charge: H2020 Marie Skłodowska-Curie Actions [765266].
Conflict of interest statement. None declared.

REFERENCES

1. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
2. Maizels,N. and Gray,L.T. (2013) The G4 genome. *PLoS Genet.*, **9**, e1003468.
3. De Cian,A., Lacroix,L., Douarre,C., Temime-Smaali,N., Trentesaux,C., Riou,J.F. and Mergny,J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
4. Balasubramanian,S. and Neidle,S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
5. Edling,C.E. and Hallberg,B. (2007) c-Kit—A hematopoietic cell essential receptor tyrosine kinase. *Int. J. Biochem. Cell Biol.*, **39**, 1995–1998.
6. Metcalfe,D.D. (2008) Mast cells and mastocytosis. *Blood*, **112**, 946–956.
7. Gregory-Bryson,E., Bartlett,E., Kiupel,M., Hayes,S. and Yuzbasiyan-Gurkan,V. (2010) Canine and human gastrointestinal stromal tumors display similar mutations in c-KIT exon 11. *BMC Cancer*, **10**, 559–568.
8. Ashman,L.K. and Griffith,R. (2013) Therapeutic targeting of c-KIT in cancer. *Expert Opin. Investig. Drugs*, **22**, 103–115.
9. Lennartsson,J. and Rönnstrand,L. (2006) The stem cell factor receptor/c-Kit as a drug target in cancer. *Curr. Cancer Drug Targets*, **6**, 65–75.
10. Fabbro,D., Ruetz,S., Buchdunger,E., Cowan-Jacob,S.W., Fendrich,G., Liebetanz,J., Mestan,J., O'Reilly,T., Traxler,P., Chaudhuri,B. *et al.* (2002) Protein kinases as targets for anticancer agents: from inhibitors to useful drugs. *Pharmacol. Ther.*, **93**, 79–98.
11. Rankin,S., Reszka,A.P., Huppert,J., Zloh,M., Parkinson,G.N., Todd,A.K., Ladame,S., Balasubramanian,S. and Neidle,S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
12. Yamamoto,K., Tojo,A., Aoki,N. and Shibuya,M. (1993) Characterization of the promoter region of the human c-kit proto-oncogene. *Jpn. J. Cancer Res.*, **84**, 1136–1144.
13. Neidle,S. (2020) *Quadruplex Nucleic Acids As Targets For Medicinal Chemistry 1st edition*. Academic Press, Cambridge, USA.
14. Kuryavyi,V., Phan,A.T. and Patel,D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
15. Hsu,S.T., Varnai,P., Bugaut,A., Reszka,A.P., Neidle,S. and Balasubramanian,S. (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.*, **131**, 13399–13409.
16. Wei,D., Husby,J. and Neidle,S. (2015) Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic Acids Res.*, **43**, 629–644.
17. Wei,D., Parkinson,G.N., Reszka,A.P. and Neidle,S. (2012) Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.*, **40**, 4691–4700.
18. Kotar,A., Rigo,R., Sissi,C. and Plavec,J. (2019) Two-quartet kit* G-quadruplex is formed via double-stranded pre-folded structure. *Nucleic Acids Res.*, **47**, 2641–2653.
19. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
20. Ducani,C., Bernardinelli,G., Högberg,B., Keppler,B.K. and Terenzi,A. (2019) Interplay of Three G-Quadruplex Units in the KIT Promoter. *J. Am. Chem. Soc.*, **141**, 10205–10213.
21. Rigo,R. and Sissi,C. (2017) Characterization of G4–G4 Crosstalk in the c-KIT Promoter Region. *Biochemistry*, **56**, 4309–4312.
22. Raiber,E.A., Kranaster,R., Lam,E., Nikan,M. and Balasubramanian,S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.*, **40**, 1499–1508.
23. Park,G.H., Plummer,H.K. and Krystal,G.W. (1998) Selective Sp1 binding is critical for maximal activity of the human c-kit promoter. *Blood*, **92**, 4138–4149.
24. Krasheninina,O.A., Novopashina,D.S., Apartsin,E.K. and Venyaminova,A.G. (2017) Recent advances in nucleic acid targeting probes and supramolecular constructs based on pyrene-modified oligonucleotides. *Molecules*, **22**, 2108.
25. Haider,S.M., Neidle,S. and Parkinson,G.N. (2011) A structural analysis of G-quadruplex/ligand interactions. *Biochimie*, **93**, 1239–1251.
26. Arola,A. and Vilar,R. (2008) Stabilisation of G-quadruplex DNA by small molecules. *Curr. Top. Med. Chem.*, **8**, 1405–1415.
27. Asamitsu,S., Obata,S., Yu,Z., Bando,T. and Sugiyama,H. (2019) Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules*, **24**, 429.
28. Yao,C., Kraatz,H.-B. and Steer,R.P. (2005) Photophysics of pyrene-labelled compounds of biophysical interest. *Photochem. Photobiol. Sci.*, **4**, 191–199.
29. Østergaard,M.E. and Hrdlicka,P.J. (2011) Pyrene-functionalized oligonucleotides and locked nucleic acids (LNAs): Tools for fundamental research, diagnostics, and nanotechnology. *Chem. Soc. Rev.*, **40**, 5771–5788.
30. Astakhova,I.K., Pasternak,K., Campbell,M.A., Gupta,P. and Wengel,J. (2013) A Locked Nucleic Acid-Based Nanocrawler: Designed and Reversible Movement Detected by Multicolor Fluorescence. *J. Am. Chem. Soc.*, **135**, 2423–2426.
31. Kumar,T.S., Myznikova,A., Samokhina,E. and Astakhova,I.K. (2013) Rapid genotyping using pyrene−perylene locked nucleic acid complexes. *Artif DNA PNA XNA*, **4**, 58–68.
32. Bichenkova,E.V., Gbaj,A., Walsh,L., Savage,H.E., Rogert,C., Sardarian,A.R., Etchells,L.L. and Douglas,K.T. (2007) Detection of nucleic acids in situ: novel oligonucleotide analogues for target-assembled DNA-mounted exciplexes. *Org. Biomol. Chem.*, **5**, 1039–1051.
33. Bichenkova,E.V., Sardarian,A.R., Wilton,A.N., Bonnet,P., Bryce,R.A. and Douglas,K.T. (2006) Exciplex fluorescence emission from simple organic intramolecular constructs in non-polar and highly polar media as model systems for DNA-assembled exciplex detectors. *Org. Biomol. Chem.*, **4**, 367–378.
34. Bichenkova,E.V., Savage,H.E., Sardarian,A.R. and Douglas,K.T. (2005) Target-assembled tandem oligonucleotide systems based on exciplexes for detecting DNA mismatches and single nucleotide polymorphisms. *Biochem. Biophys. Res. Commun.*, **332**, 956–964.
35. Ensslen,P. and Wagenknecht,H.-A. (2015) One-dimensional multichromophor arrays based on DNA: from self-assembly to light-harvesting. *Acc. Chem. Res.*, **48**, 2724–2733.
36. Teo,Y.N. and Kool,E.T. (2012) DNA-multichromophore systems. *Chem. Rev.*, **112**, 4221–4245.
37. Filichev,V.V. and Pedersen,E.B. (2005) Stable and selective formation of hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic sonogashira solid-phase coupling reactions. *J. Am. Chem. Soc.*, **127**, 14849–14858.
38. Hrdlicka,P.J., Kumar,T.S. and Wengel,J. (2005) Targeting of mixed sequence double-stranded DNA using pyrene-functionalized 2′-amino-α-L-LNA. *Chem. Commun.*, **34**, 4279–4281.
39. Nakamura,M., Fukunaga,Y., Sasa,K., Ohtoshi,Y., Kanaori,K., Hayashi,H., Nakano,H. and Yamana,K. (2005) Pyrene is highly emissive when attached to the RNA duplex but not to the DNA duplex: the structural basis of this difference. *Nucleic Acids Res.*, **33**, 5887–5895.
40. Lou,C., Dallmann,A., Marafini,P., Gao,R. and Brown,T. (2014) Enhanced H-bonding and π-stacking in DNA: a potent duplex-stabilizing and mismatch sensing nucleobase analogue. *Chem. Sci.*, **5**, 3836–3844.

41. Crescenzo,A.D., Ettorre,V. and Fontana,A. (2014) Non-covalent and reversible functionalization of carbon nanotubes. *Beilstein J. Nanotechnol.*, **5**, 1675–1690.
42. Lemek,T., Mazurkiewicz,J., Stobinski,L., Lin,H.M. and Tomasik,P. (2007) Non-covalent functionalization of multi-walled carbon nanotubes with organic aromatic compounds. *J. Nanosci. Nanotechnol.*, **7**, 3081–3088.
43. Kovačič,M., Podbevšek,P., Tateishi-Karimata,H., Takahashi,S., Sugimoto,N. and Plavec,J. (2020) Thrombin binding aptamer G-quadruplex stabilized by pyrene-modified nucleotides. *Nucleic Acids Res.*, **48**, 3975–3986.
44. Takahashi,S., Kim,K.T., Podbevšek,P., Plavec,J., Kim,B.H. and Sugimoto,N. (2018) Recovery of the formation and function of oxidized G-quadruplexes by a pyrene-modified guanine Tract. *J. Am. Chem. Soc.*, **140**, 5774–5783.
45. Cogoi,S., Zorzet,S., Rapozzi,V., Géci,I., Pedersen,E.B. and Xodo,L.E. (2013) MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice. *Nucleic Acids Res.*, **41**, 4049–4064.
46. Breslauer,K.J. (1995) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods in Enzymology, Energetics of Biological Macromolecules*. Academic Press, Cambridge, USA, Vol. **259**, pp. 221–242.
47. Lee,W., Tonelli,M. and Markley,J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
48. Zgarbová,M., Šponer,J., Otyepka,M., Cheatham,T.E., Galindo-Murillo,R. and Jurečka,P. (2015) Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.*, **11**, 5723–5736.
49. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
50. Krepl,M., Zgarbová,M., Stadlbauer,P., Otyepka,M., Banáš,P., Koča,J., Cheatham,T.E., Jurečka,P. and Šponer,J. (2012) Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.*, **8**, 2506–2520.
51. Zgarbová,M., Luque,F.J., Šponer,J., Cheatham,T.E., Otyepka,M. and Jurečka,P. (2013) Toward improved description of DNA backbone: revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.
52. Comer,J., Gumbart,J.C., Hénin,J., Lelièvre,T., Pohorille,A. and Chipot,C. (2015) The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B*, **119**, 1129–1151.
53. Minoukadeh,K., Chipot,C. and Lelièvre,T. (2010) Potential of mean force calculations: a multiple-walker adaptive biasing force approach. *J. Chem. Theory Comput.*, **6**, 1008–1017.
54. Mones,L., Bernstein,N. and Csányi,G. (2016) Exploration, sampling, and reconstruction of free energy surfaces with Gaussian process regression. *J. Chem. Theory Comput.*, **12**, 5100–5110.
55. Bouchal,T., Durník,I., Illík,V., Réblová,K. and Kulhánek,P. (2020) Importance of base-pair opening for mismatch recognition. *Nucleic Acids Res.*, **48**, 11322–11334.
56. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
57. Fonville,J.M., Swart,M., Vokáčová,Z., Sychrovský,V., Šponer,J.E., Šponer,J., Hilbers,C.W., Bickelhaupt,F.M. and Wijmenga,S.S. (2012) Chemical shifts in nucleic acids studied by density functional theory calculations and comparison with experiment. *Chem. Eur. J.*, **18**, 12372–12387.
58. Fernando,H., Reszka,A.P., Huppert,J., Ladame,S., Rankin,S., Venkitaraman,A.R., Neidle,S. and Balasubramanian,S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*, **45**, 7854–7860.
59. Hou,J.-Q., Tan,J.-H., Wang,X.-X., Chen,S.-B., Huang,S.-Y., Yan,J.-W., Chen,S.-H., Ou,T.-M., Luo,H.-B., Li,D. *et al.* (2011) Impact of planarity of unfused aromatic molecules on G-quadruplex binding: Learning from isaindigotone derivatives. *Org. Biomol. Chem.*, **9**, 6422–6436.
60. Ma,Y.-Y., Yu,S., He,X.-J., Xu,Y., Wu,F., Xia,Y.-J., Guo,K., Wang,H.-J., Ye,Z.-Y., Zhang,W. *et al.* (2014) Involvement of c-KIT mutation in the development of gastrointestinal stromal tumors through proliferation promotion and apoptosis inhibition. *Onco Targets Ther.*, **7**, 637–643.
61. Yamanoi,K., Higuchi,K., Kishimoto,H., Nishida,Y., Nakamura,M., Sudoh,M. and Hirota,S. (2014) Multiple gastrointestinal stromal tumors with novel germline c-kit gene mutation, K642T, at exon 13. *Hum. Pathol.*, **45**, 884–888.
62. Boldrini,L., Ursino,S., Gisfredi,S., Faviana,P., Donati,V., Camacci,T., Lucchi,M., Mussi,A., Basolo,F., Pingitore,R. *et al.* (2004) Expression and mutational status of c-kit in small-cell lung cancer: prognostic relevance. *Clin. Cancer Res.*, **10**, 4101–4108.
63. Meng,D. and Carvajal,R.D. (2019) KIT as an oncogenic driver in melanoma: an update on clinical development. *Am. J. Clin. Dermatol.*, **20**, 315–323.
64. Sakabe,T., Azumi,J., Haruki,T., Umekita,Y., Nakamura,H. and Shiota,G. (2017) CD117 expression is a predictive marker for poor prognosis in patients with non-small cell lung cancer. *Oncol. Lett.*, **13**, 3703–3708.
65. Kee,D. and Zalcberg,J.R. (2012) Current and emerging strategies for the management of imatinib-refractory advanced gastrointestinal stromal tumors. *Ther. Adv. Med. Oncol.*, **4**, 255–270.
66. Safe,S. and Abdelrahim,M. (2005) Sp transcription factor family and its role in cancer. *Eur. J. Cancer*, **41**, 2438–2448.