9724–9737 *Nucleic Acids Research, 2021, Vol. 49, No. 17* *Published online 3 September 2021*
https://doi.org/10.1093/nar/gkab674

Double-stranded flanking ends affect the folding kinetics and conformational equilibrium of G-quadruplexes forming sequences within the promoter of *KIT* oncogene

Guglielmo Vesco[1], Marco Lamperti[2], Domenico Salerno[3], Claudia Adriana Marrano[3], Valeria Cassina[3], Riccardo Rigo[4], Enrico Buglione[3], Maria Bondani[5], Giulia Nicoletto[4], Francesco Mantegazza[3], Claudia Sissi[4,6,*] and Luca Nardo[3,*]

[1]Department of Science and High Technology, University of Insubria, 22100 Como, Italy, [2]Department of Physics, Polytechnic of Milan, 23900 Lecco, Italy, [3]School of Medicine and Surgery, BioNanoMedicine Center NANOMIB, University of Milano-Bicocca, 20854 Vedano al Lambro (MB), Italy, [4]Department of Pharmaceutical and Pharmacological Sciences, University of Padova, 35131 Padova, Italy, [5]Institute for Photonics and Nanotechnology, IFN-CNR, 22100 Como, Italy and [6]CRIBI Biotechnology Center, University of Padova, 35131 Padova, Italy

Received April 08, 2021; Revised July 13, 2021; Editorial Decision July 15, 2021; Accepted September 01, 2021

ABSTRACT

G-quadruplexes embedded within promoters play a crucial role in regulating the gene expression. KIT is a widely studied oncogene, whose promoter contains three G-quadruplex forming sequences, c-kit1, c-kit2 and c-kit*. For these sequences available studies cover ensemble and single-molecule analyses, although for kit* the latter were limited to a study on a promoter domain comprising all of them. Recently, c-kit2 has been reported to fold according to a multi-step process involving folding intermediates. Here, by exploiting fluorescence resonance energy transfer, both in ensemble and at the single molecule level, we investigated the folding of expressly designed constructs in which, alike in the physiological context, either c-kit2 or c-kit* are flanked by double stranded DNA segments. To assess whether the presence of flanking ends at the borders of the G-quadruplex affects the folding, we studied under the same protocols oligonucleotides corresponding to the minimal G-quadruplex forming sequences. Data suggest that addition of flanking ends results in biasing both the final equilibrium state and the folding kinetics. A previously unconsidered aspect is thereby unravelled, which ought to be taken into account to achieve a deeper insight of the complex relationships underlying the fine tuning of the gene-regulatory properties of these fascinating DNA structures.

INTRODUCTION

G-quadruplexes (G4s) are physiologically occurring nucleic acid structures not based on the canonical Watson-Crick pairing that can be assumed by sequences featuring guanine repeats (1–4). In the genome, they primarily occur at the telomeres, where they are involved in triggering the apoptotic/senescence response to cellular aging (5–7) and at gene promoters, where they act as epigenetic markers for the regulation of gene transcription (7–9).

The free energy required to locally melt the double-helical structure and to allow the folding of the G-rich strand into the tetrahelix is provided by two major contributions: Hoogsteen-like hydrogen bonding among four in-plane guanine bases, constituting the G-tetrad (or G-quartet), and stacking interactions between G-tetrads (1,2). The difference in energy between the double-helix and the G4 structures is sufficiently small to allow transitions between the two states *in vivo*. This shift is triggered by environmental changes including ions concentration gradients (particularly, increases in K^+ concentration shift the equilibrium towards the G4) (13,14) as well as nanomechanical solicitations (typically provided by proteins exerting torques on the double-helix) (15). Depending on the number of G-tetrads as well as on the sequence and length of the loops, G4s can assume a number of different conformations (10–12).

*To whom correspondence should be addressed. Tel: +39 031 238 6272; Fax: +39 031 238 6119; Email: luca.nardo@unimib.it
Correspondence may also be addressed to Claudia Sissi. Tel: +39 031 238 6272; Fax: +39 031 238 6119; Email: claudia.sissi@unipd.it
Present address: Luca Nardo, Department of Science and High Technology, University of Insubria, 22100 Como, Italy.

The folding into a G4 structure of a DNA sequence within a gene promoter frequently reduces its expression (8,9). For this reason, G4s have been recently proposed as ideal targets for gene-regulatory drugs (16–22). Foreseen applications cover cancer-cell specific antineoplastic agents able to silence overexpressed oncogenes through the induction/stabilization of G4s embedded at their promoters. The rational design of such therapeutics clearly relies on a detailed knowledge of the structure (10), stability (12), and folding pathways/kinetics of the G4 to be targeted (23). Moreover, a deeper comprehension of G4s folding energetics, dynamics and conformational equilibria would hopefully disclose new insight on their roles in the fine tuning of the cellular metabolism (3,4).

The promoter of the *KIT* proto-oncogene is among the most characterized to this regard (24). Upstream the transcription initiation site (TIS), it contains three distinct sites capable of folding into G4 structures, namely the sequence between bases −87 and −108, usually referred to as c-kit1 (25,26), the one comprised between bases −138 and −158, called c-kit2 (26–30), and that delimited by bases −115 and −136, named c-kit* (31,32). These three G4s are expected to play a key role in regulating the expression of *KIT*. The three G4s forming regions were characterized in the past by a panel of spectroscopic techniques, including UV-visible absorption (UV-Vis) (25–27,30,31), circular dichroism (CD) (25–27,30–32), nuclear magnetic resonance (NMR) (25,27,28,31,32). In addition, c-kit1 and c-kit2 were probed by single molecule techniques (26,29,33), including fluorescence resonance energy transfer (FRET), a widely applied technique to elucidate the conformational features of biomolecules (34–37). The c-kit* sequence, being the last one to be identified, has been investigated to a lesser extent, and, to the best of our knowledge, there exists only one study in which its folding has been characterized at the single-molecule level, namely by exploiting magnetic tweezers (33).

All available data emphasize the complex structural equilibria occurring at these regions. Among them, the most polymorphic appears to be c-kit2. In solution, it can fold into a monomeric and a dimeric G4 structures according to a folding pathway that comprises several kinetic intermediates as recently probed by native electrospray ionization mass spectrometry (ESI-MS), stopped-flow UV-Vis and CD experiments (30,38). Noteworthy, addition of a single base at the 3′ terminal of c-kit2 largely suppresses formation of the dimeric G4 in solution. However, further addition of bases at the 5′- and 3′-ends tends to destabilize also the monomeric G4 structure of c-kit2 (24,39) as well as of c-kit1 (40) and c-kit* (32). Interestingly, in c-kit*, the commonly studied minimal sequence comprises a 3′-terminal tail that, although not involved in G-quartet pairing, is fundamental to drive the folding process toward a unique antiparallel G4 with increased stability. However, in gene promoters the G4 forming tracts are actually embedded within a longer double-stranded DNA (dsDNA) sequence. Thus, at least on a quantitative standpoint, the *in vitro* characterizations of oligonucleotides encompassing only the bases actually involved in the G4 core might not properly model the *in vivo* behaviour

In this article, we present a detailed characterization of the conformational equilibria and folding kinetics of constructs designed expressly to probe the role of flanking ends on the G4 folding. Our works was based on advanced time-resolved and single-molecule FRET (SM-FRET) experiments on freely diffusing molecules exploiting a confocal spectroscopy setup. Consistently, we adopt a statistical approach, by sampling the folding pathways on a number of different molecules passing one at a time through the observation volume. Our approach differs from previously reported SM data on G4s where total internal reflection fluorescence microscopy was applied to analyse the dynamics of a single molecule fixed onto glass surfaces (26,29,34), and offers several advantages, including the possibility to monitor the folding dynamics in pseudo-physiological environment, far from surfaces that can possibly interfere with the oligonucleotide. Moreover, it avoids artefacts connected to different photobleaching kinetics of the donor and acceptor chromophores, as these photophysical mechanisms are negligible over the dwell time of molecules within the observation volume (<1 ms). To the best of our knowledge, this technique has been exploited only to monitor protein folding and peptide conformational dynamics at the single molecule level (35–37), but never on nucleic acids.

The constructs we investigated (c-kit2_LONG and c-kit*_LONG hereafter) encompass the G4-folding domain of c-kit2 or c-kit* with the addition of 21-nucleotides double-stranded flanking ends at both 3′- and 5′-terminals (Figure 1). Although the folding dynamics of a G4 *in vivo* can be affected by other DNA features, including, but not limited to, the presence of the complementary strand and the size of the local denaturation bubble required for G4 formation, our constructs constitute a better model of the real promoter with respect to the minimal length single stranded oligonucleotides which are commonly used in the current literature. Moreover, in comparison to previously reported strategies, the design of the building blocks of our constructs results into a highly versatile model, since the labelled strands can be easily shifted on different templates. Finally, bringing the chromophores further apart from guanine bases avoids their interaction with the G4 as well as non-radiative decay through excited-state electron transfer (i.e. guanine quenching, see [36] and references therein)

In order to perform a proper comparison of the so obtained data, the two short sequences corresponding to the minimal, single-stranded, G4-forming sequences c-kit2 and c-kit*, dual-labelled at the 3′- and 5′- ends with the same donor (D) and acceptor (A) (c-kit2_SHORT and c-kit*_SHORT hereafter), were also analysed. Interestingly, such comparison confirmed at a single molecule level that the presence of flanking ends induces substantial perturbations in the folding equilibrium of both c-kit2 and c-kit*, as well as in the folding kinetics of c-kit2. In this connection, our SM-FRET data were combined with time-resolved fluorescence experiments on molecular ensembles. The two techniques yielded compatible results even on a quantitative basis, thus conferring superior reliability to our findings.

The above observations disclose novel and partially unsuspected properties of G4, offering in perspective a less stylised picture of the complex interplays between sequence,

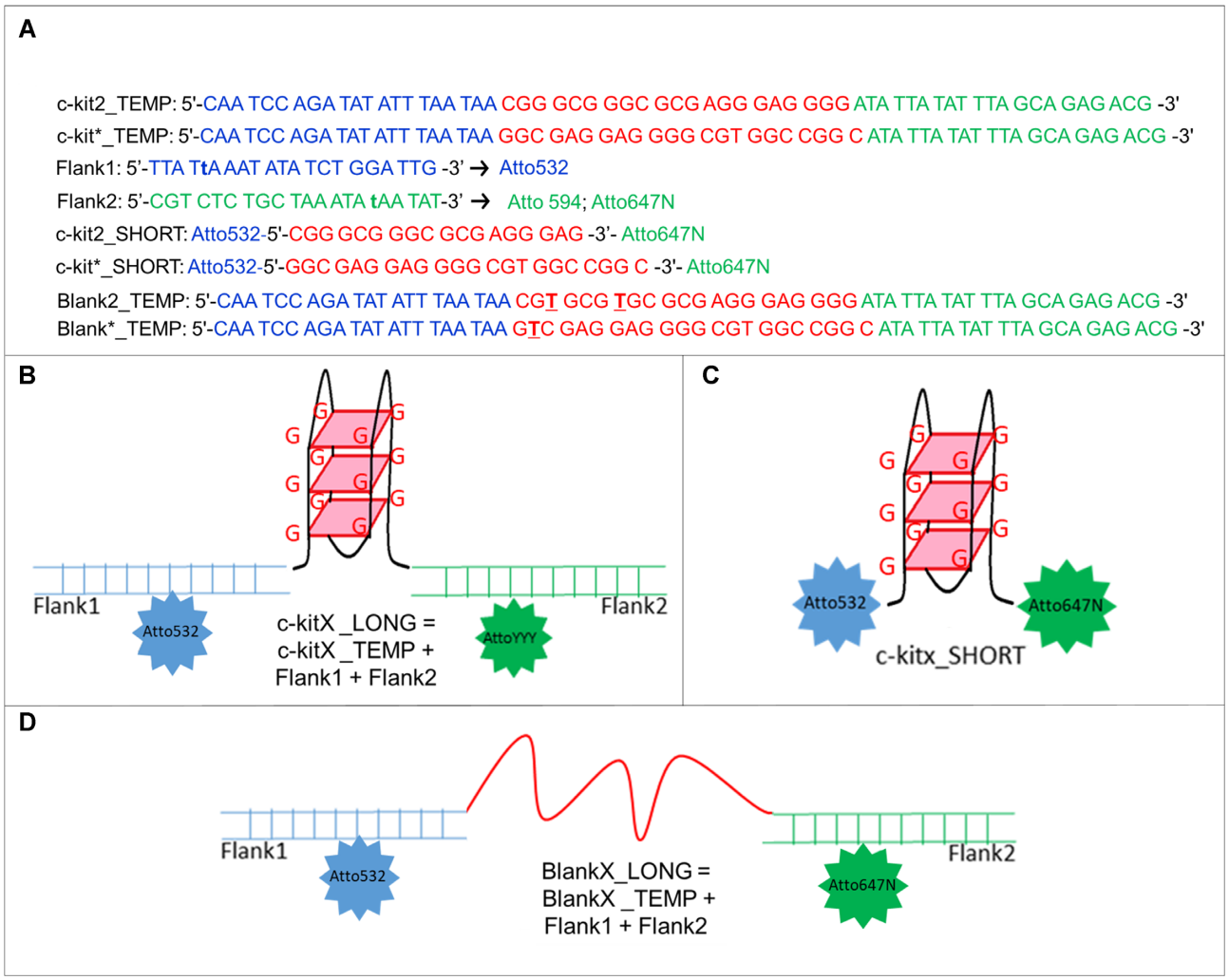

Figure 1. (**A**) Sequences used to build-up KIT models: c_kitX_TEMP (X = 2,*) are the 66 bp long oligonucleotides encompassing the G4-folding sequence (evidenced in red characters) between two single-stranded flanking ends (denoted by blue and green characters for the 5′ and 3′ extension, respectively). The latter are made anneal with complementary sequences (Flank1 and Flank2, respectively: colour codes are conserved), in order to build up the c-kitX_LONG (X = 2,*) constructs with double-stranded flanking ends. The c-kitX_SHORT (X = 2,*) oligonucleotides are those mimicking the minimal G4-forming sequences, and are labelled with the FRET donor and acceptor at their 5′ and 3′ end, respectively. BlankX_TEMP (X = 2,*) are oligonucleotides different from the corresponding c-kitX_TEMP by the G–T point mutations evidenced in bold underlined characters, which impair G4 folding. They are made to anneal with Flank1 and Flank2 to obtain the non-folding constructs BlankX (X = 2,*). Panels (**B**)–(**D**): sketches of (B) the c-kitX_LONG; (C) the c-kitX_SHORT and (D) the BlankX constructs.

structural properties and epigenetic functions of these amazingly versatile secondary structures of DNA.

MATERIALS AND METHODS

Oligonucleotides and sample preparation

Synthetic single strand DNA oligonucleotides (ssDNA) were purchased at Biomers.net and provided lyophilized with a HPLC grade of purity. A complete list of all sequences used in this study is reported in Table S2. As shown in Figure 1, three different classes of oligonucleotides were used:

(a) four ∼60 base-long oligonucleotides different in their central part (red in Figure 1) where the G4-forming sequences c-kit2 and c-kit* (c-kit2_TEMP and c-kit*_TEMP in Figure 1, respectively), or their non G4-forming mutated counterparts (Blank2_TEMP and Blank*_TEMP, respectively) where inserted. The inability of the mutated sequences to form G4 is demonstrated in (42) for Blank2 and in (31) for Blank*.

(b) two 21 base-long sequences (Flank1 and Flank2 in Figure 1) complementary to the 21 bases at the 5′- (blue in Figure 1) and 3′-end (green in Figure 1) of the four ∼60 base-long oligonucleotides. Flank1 and Flank2 were purchased both unlabelled (used for the UV-Vis absorption and CD experiments) and labelled with fluorophores (used for the FRET experiments). The D fluorophore was always the dye Atto532 (patented by Atto-Tec GmbH), and was linked to the T at position 5 of Flank1. In order to probe the independence of our results from the selected D–A pair, two different acceptors were compared (linked at the T at position 16 of Flank2): Atto647N was used to character-

ize all the constructs in any environmental condition, while Atto594 was utilized to replicate a subset of experiments.

(c) two oligonucleotides corresponding to the G4 forming sequences c-kit2 and c-kit* labelled at 5′ with Atto532 and at 3′ with Atto647N (c-kit2_SHORT and c-kit*_SHORT, respectively)

All oligonucleotides were resuspended in 10 mM Tris–HCl at pH 7.4 at a final concentration of 100 μM. Annealing of the _TEMP sequences with Flank1 and Flank2 was pursued by treating 1:1:1 mixtures of the oligonucleotides (diluted in Tris–HCl 10 mM to a final concentration of 20 μM) to the following annealing conditions in a Mastercycler nexus thermocycler (Eppendorf): fast heating to 98°C, hold for 5 min and cooling at −3°C/min till 20°C, with a 10 min stop at the annealing temperature of 35°C. When the annealing was performed in the presence of either 100 mM LiCl or 100 mM KCl, (see Results section for details) the annealing temperature was raised to 40°C as a result of the increased ionic strength. The obtained constructs are sketched in Figure 1B–D.

The samples used for the different experiments were diluted to the following final concentrations: 1–2 μM for the UV–Vis absorption measurements; 500 nM for the steady-state and time-resolved fluorescence measurements; 100 pM for the SM-FRET measurements. This concentration corresponds to an average value of 0.1 molecules within the observation volume, as determined by fluorescence correlation spectroscopy (FCS) measurements (data not shown).

Steady-state spectroscopy measurements

The UV–Vis absorption spectra of unlabelled oligonucleotides were recorded with a Lambda2 spectrophotometer (Perkin Elmer), equipped with a custom-made circulating-bath thermostated cell holder connected to a F12 mc refrigerated/heating circulator (Julabo). Temperature ramps were programmed by the Julabo Easytemp software. The thermostated bath temperature was raised from 25°C to 85°C at a rate of 0.25°C/min. The actual temperature of the cell was measured by means of a Testo 730 thermometer inserting the probe in a specific lodging within the cell holder, and acquired each second by exploiting the Testo Comfort X35 Software. Details on this temperature-control system are reported elsewhere (41). The samples were placed in hermetically sealed screw-top quartz cuvettes (Hellma) to avoid buffer evaporation at high temperatures. Absorbance was measured at regular time intervals in the 230–700 nm wavelength range.

Steady-state fluorescence spectra were acquired at 25°C on a Cary Eclipse spectrofluorimeter (Varian). They were automatically corrected with respect to differential excitation lamp spectral radiance and detector spectral quantum efficiency.

Circular dichroism

Circular dichroism spectra were acquired on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature controller. DNA samples were prepared at 2 μM in 10 mM Tris–HCl at pH 7.4; when required 100 mM KCl was included. Before use, all DNA solutions were annealed in the required buffer. Each spectrum represents the average of 3 scans acquired at a 20 nm/min scanning speed. Observed CD signals were converted to mean residue ellipticity [Θ] = deg cm^2 $dmol^{-1}$ (Mol. Ellip.).

Time-correlated single photon counting setup and fluorescence decay analysis

Time-resolved fluorescence measurements were obtained by time-correlated single photon counting (TCSPC), using a TCSPC setup that is fully described elsewhere (43–45) and is endowed with <30 ps temporal resolution. For the measurements presented in this paper, a mode-locked, frequency-doubled Ti:sapphire laser delivering 3 ps pulses at 48 MHz repetition rate (420 nm output wavelength) was used as the excitation source. This out-of-peak excitation was preferred to excitation in the green by means of the Nd:VAN laser used in previous works (41) because the fluorophore chosen as the D has a rather long fluorescence lifetime (≈4ns). The lower repetition rate of the Ti:sapphire with respect to the Nd:VAN assures complete decay of the D fluorescence within a pulse period even in the absence of the A. The free D lifetime and spectral line-shape proved to be insensitive to the change in excitation wavelength from 532 nm (D excitation spectrum peak wavelength) to 420 nm (data not shown). The samples were held in 1 × 1 cm^2 quartz cuvettes. Fluorescence was collected at 90 deg to the excitation beam through a (560 ± 4) nm band-pass filter (CVI).

Three decay patterns were reconstructed for each typology of sample, cumulating the count-statistics until 65 535 peak counts. The experimental counts vs time histograms were fitted to either a single- or a multi-exponential decay model using a customized Matlab routine based on the Levenberg–Marquardt χ^2 minimization algorithm. The number of decay components was fixed adding components one by one until the addition of a further component did not result in a significant improvement of the fit. The goodness of the fits was evaluated in term of the residual plots and by calculating the residuals autocorrelation function. The lifetime and relative amplitude values reported herein are the average of the fitting values over the three parallel measurements, with errors given by the pertaining standard deviations.

Single molecule fluorescence resonance energy transfer setup

The SM-FRET setup used for the measurements presented herein is an implementation of the FCS setup described in full detail elsewhere (41,44,45). The main modifications made are:

- Size of the pin-hole, in this article is set to 50 μm diameter.
- Filters at the income of the dark chamber (to remove residual excitation stray light from the fluorescence beam), in this article 3xFEL550 long-wavelength pass filters (Thorlabs) + BLP01-532R-25 532 nm EdgeBasic™ long-wavelength pass filter (Semrock) + PN 532 RB20, 532 nm NOTCH filter
- Beam splitter substitution with dichroic mirrors suitable to separate the D from the A fluorescent light (640DRSPXR

with Atto647N as the A, 594DCLP with Atto594 as the A)
- Additional filters on the objectives focusing the fluorescent light on the detectors, used to avoid fluorescence crosstalk between the D and the A detection channels: RPE590SP short-wavelength pass filter coupled with RPB560-590 band-pass filter on the D branch; either RPE660LP long-wavelength pass filter coupled with RPB660-700 band-pass filter (with Atto647N as the A) or FEL600 long-wavelength pass filter (Thorlabs) coupled with 630AF50 band-pass filter (with Atto594 as the A)
- Photon recording through two synchronized SPC150 TC-SPC boards (Becker & Hickl GmbH) used in the FIFO modality, instead of on-line temporal correlation through the hardware correlator Flex2k (correlator.com). This system enables to record the absolute time of detection of each photon in both the D and the A branch with respect to the experiment initiation time.

Whenever not explicitly quoted, the filters were purchased from Omega optics.

Single molecule fluorescence resonance energy transfer data acquisition and analysis

SM-FRET data were acquired by deposing a 100 μl droplet of sample onto a glass coverslip put on the microscope objective. Traces were acquired for 3 minutes. According to the supplier (https://www.atto-tec.com/images/ATTO/Katalog_Preisliste/Katalog_2019_2020.pdf), the typical photobleaching time for the D fluorophore at light intensities comparable to those used in our experiments (≈50 kW/cm^2) is >20 min. Five measurements were collected for each sample and at least six parallel samples were examined for each specimen. Overall, the evaluation of the FRET probability distributions was performed averaging at least 30 different traces.

In order to build up the FRET efficiency probability histograms, the FIFO traces of the D and A channels were analysed according to the following procedure, using a home-written Matlab routine:

- 'background' traces were acquired at both channels putting on the coverslip a droplet of buffer instead of the oligonucleotides solutions
- A sampling time t_S was chosen and photons detected within subsequent time bins $[nt_S, (n + 1)t_S]$, with n being an integer were counted. The value of t_S was set to 400 μs, i.e.: the typical diffusion time of the constructs through the observation volume, measured by FCS (data not shown).
- The average number of photon counts within a t_S time bin was determined for the background traces in both the D, B_D, and the A, B_A, channel, together with the corresponding standard deviations, σ_D and σ_A.
- The traces recorded for the different oligonucleotide specimens were binned. B_D and B_A were subtracted bin by bin to the D and A traces, respectively.
- In order to account for the differences in detection quantum efficiency and in the portion of spectrum selected by the filters in the two channels, the D trace was multiplied by a relative quantum yield factor, ϕ_{rel} (calculated to be 0.6 when Atto647N was used as the A, 1 when Atto594 was applied).
- The quantity $T = 10\sqrt{\sigma_D^2 + \sigma_A^2}$ was set as the threshold for the detection of fluorescence bursts. A bin interval was considered as corresponding to the detection of a construct within the observation volume whenever the sum of the entry of the D and the A channel overcame the threshold. Only traces displaying >1500 bins over threshold were used to build up the FRET probability distributions. It is worth noting that, because at least 30 valid traces were recorded per sample, the statistics accumulated in order to build up the FRET efficiency histograms presented in this article relies on a minimum of 45 000 bursts per histogram.
- For each bin over the threshold the FRET efficiency value was calculated according to the equation.

$$E_{FRET} = \frac{N_A}{N_A + N_D}$$

where N_A is the number of photons detected in the A channel in that bin, and N_D the number of photons detected in the *D* channel multiplied by ϕ_{rel}.

- The so-obtained E_{FRET} values were used to build up a frequency histogram.
- The histograms obtained by each valid trace were averaged column-by-column to obtain average E_{FRET} frequency histograms. The pertaining standard errors were assumed as error bars.
- The averaged histograms were fitted to a multi-gaussian peak function to unravel sub-populations characterized by different D-to-A distances (i.e.: distinct conformations of the construct)

RESULTS

Preliminary characterization of the oligo-constructs

Before acquisition of SM-FRET data, the c-kit2_LONG and c-kit*_LONG constructs were characterized by UV–Vis absorption spectroscopy in order to probe the correct annealing of Flank1 and Flank2 to c-kit2_TEMP and c-kit*_TEMP and to assess the G4 formation under selected experimental conditions. The results of these investigations are detailed in the Supplementary Materials Section (Figures S1–S3). Briefly, the annealing efficiency of the flanking ends was measured to be always >80%. The presence of a hyperchromic contribution at 295 nm was observed for c-kitX_LONG constructs only in the presence of 100 mM KCl, thus supporting G4 formation (46). In terms of overall folding topology, the addition of flanking ends (both in single stranded and double stranded form) did not alter c-kit2 G4 (Figure S4a). Conversely, in line with already reported data, a topological rearrangement was observed by comparing c-kit*_SHORT to its related c-kit*_TEMP or c-kit*_LONG (Figure S4b) (32).

Finally, putative effects of annealing of the D- and A-labelled flanks to the template sequences and of buffer interferences on the D- and A-fluorescence emission properties were ruled out by steady state fluorescence measurements (see Figure S5 in the Supplementary Materials Sec-

tion for exemplary fluorescence spectra). These measurements also demonstrate that both the D and the A fluorescence intensity are appreciably affected by conformational transition of the constructs from the unfolded state to the G4.

Single molecule FRET efficiency measurements

Assessment of fluorescence crosstalk on donor-only labelled constructs. The occurrence of undesired leakage of the D fluorescence in the A detection channel (also known as fluorescence crosstalk or blead-through) may significantly alter the detected FRET efficiencies, namely by systematically shifting them to higher values. Moreover, the presence of constructs with ineffective A was often reported, resulting in the build-up of a fictitious low-FRET efficiency peak in the SM-FRET distributions. To estimate the extent of this adverse phenomenon in our setup SM-FRET measurements on c-kit2_LONG constructs obtained by annealing of c-kit2_TEMP with Atto532-labelled Flank1 and unlabelled Flank2 were performed. The results are detailed in the Supplementary Materials Section, and shown in Figure S6. They stem in support of a negligible bias.

Conformational equilibrium in the absence of KCl. SM-FRET experiments were performed on both the c-kitX_LONG and the c-kitX_SHORT constructs (X = 2,*) in the absence of K^+ ions. The results are reported in the Supplementary Materials Section (see the devoted paragraph, specifically Table S1 and Figures S6–S10). The interchromophore distance distributions seem to depend only slightly on the environmental conditions (no salt versus 100 mM LiCl added) and on the specificities of the various constructs, suggesting a flexible, coiled conformation in which the flanking ends are quite free to reorient in the c-kitX_LONG samples.

Conformational equilibrium upon addition of KCl. The FRET efficiency distributions of both c-kitX_SHORT and c-kitX_LONG in TRIS-HCl 10 mM added with 100 mM KCl (see Figure 2) are qualitatively different from the corresponding ones measured in the absence of salt. Indeed, all the distributions obtained in the presence of KCl are bimodal whereas, as shown in Figure S9, panels A and B, the experiments performed on c-kitX_LONG upon addition of 100 mM LiCl resulted in monomodal distributions very similar to those obtained in the absence of salt. Moreover, as shown in Figure S10, panels C and D, even in the presence of 100 mM KCl, the BlankX (X = 2, *) constructs exhibit FRET probability distributions similar to those detected in the absence of salt.

The distribution obtained for c-kit2_LONG shows a major, broad peak with maximum at around 0.27 FRET efficiency and a shoulder extending from 0.7 to 0.9. At difference, the distribution collected for c-kit2_SHORT is dominated by two well resolved peaks, one relatively broad, with maximum at ≈0.2 FRET efficiency, the other sharper, with maximum occurring at ≈0.9 FRET efficiency. The same qualitative differences are observed between the c-kit*_LONG and c-kit*_SHORT distribution. Namely, the distribution of c-kit*_LONG is peaked at 0.25 FRET efficiency, with a shoulder from 0.5 to 0.8, while the FRET efficiency distribution of c-kit* short exhibits two peaks, one broader with maximum at ≈0.2 FRET efficiency and the other, sharper, at ≈0.94 FRET efficiency.

On a quantitative standpoint, both the c-kitX_LONG and c-kitX_SHORT FRET efficiency histograms are optimally fitted by the superposition of three peak functions. However, while for c-kitX_LONG a three-Gaussian model allows to attain accurate results, the c-kitX_SHORT frequency profiles are faithfully reproduced only by means of the composition of two Gaussians with a Lorentzian distribution (the inaccuracy of the three Gaussian model in this instance is shown in Figure S11). It should be noted that, in the absence of independent information on the number and nature of species coexisting in the sample, the choice of the fitting model to apply is somewhat arbitrary. It is thus a common practice in the community to fit the FRET histograms to the minimal number of peak functions yielding an accurate fit. The peak positions and the relative amplitudes of the three components are reported in Table 1. The fitting parameters clearly evidenced that the expectation values of the lowest-FRET-efficiency peak are very similar for the four samples. Moreover, they are very similar to those obtained by fitting to a single Gaussian model function the distributions pertaining to the same construct in the absence of KCl. On this basis, we associated this component to residuals of random-coiled construct. Indeed, substantial amounts of unfolded are generally detected in G4-folding experiments at the single molecule level (26,29,33). The second Gaussian component also has rather construct-independent expectation value, while the third population resolved in the FRET efficiency histograms has somewhat more construct-dependent characteristics. We denote the unfolded residual as U and the folded structures as S1 and S2, respectively.

The coexistence at equilibrium of more than one conformation is typical of a reversible chemical equilibrium, which configures a potential energy landscape where the free energies of the interconverting species are of comparable magnitude. Namely, in our experimental conditions, the free energy landscape should be shaped as a three-holed potential, with minima of comparable energy corresponding to U, S1 and S2, and with quite low barriers between one hole and another. This energetic picture matches with the transient nature of physiological G4s. Indeed, the latter are expected to be easily disrupted by application of modest forces or torques. An estimation of the free energy gap between the coiled sequence and the folded G4 configuration in the case of KIT promoter, based on magnetic tweezers experiments, has been reported in a recent publication (33).

In the case of c-kitX_LONG, S1 is detected with higher probability (about 42% relative abundance for c-kit2_LONG and 33% for c-kit*_LONG) than S2 (about 22% and 12% for c-kit2_LONG and c-kit*_LONG, respectively). Species U accounts for about 36% of the c-kit2_LONG construct and 54% of the c-kit*_LONG construct. Conversely, for c-kitX_SHORT the high-FRET S2 conformation is the major component (accounting for about 62% of oligos in the case of c-kit2_SHORT and 60% for c-kit*_SHORT) and S1 is the least abundant species

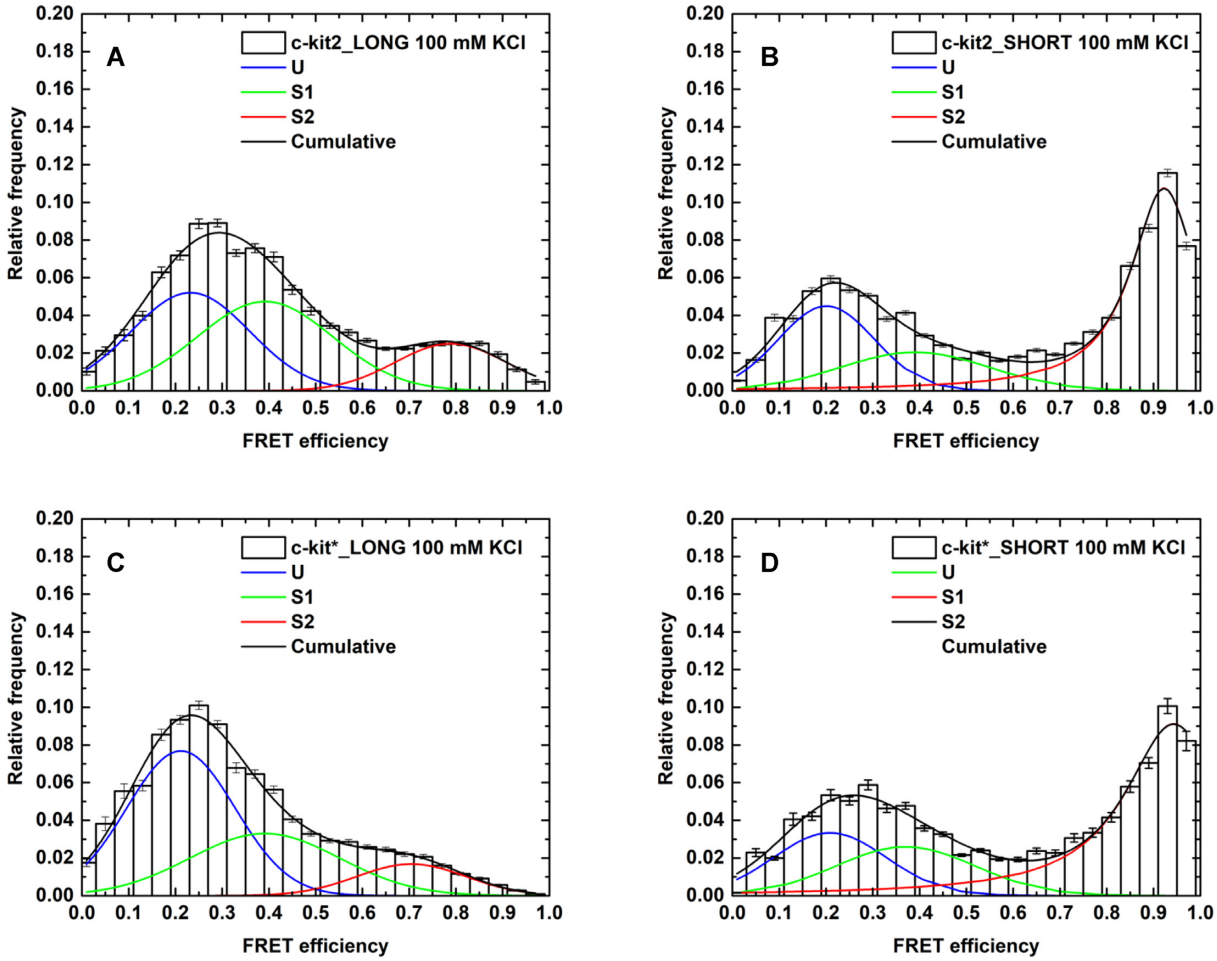

Figure 2. FRET efficiency distribution of (**A**) c-kit2_LONG; (**B**) c-kit2_SHORT; (**C**) c-kit*_LONG and (**D**) c_kit*_SHORT in Tris–HCl added with KCl 100 mM. Black solid lines represent the best-fitting curves according to the multi-peak model functions detailed in the main text. Coloured solid lines represent the single Gaussian/Lorentzian components of the multi-peak functions, as detailed in the text.

Table 1. Peak position, E_i, and relative amplitude, f_i, values of U, S1 and S2 for c-kitX_LONG and c-kitX_SHORT in 100 mM KCl

Sample	E_U(f_U)	E_{S1} (f_{S1})	E_{S2} (f_{S2})	
c-kit2_LONG	0.23 ± 0.01 (0.36 ± 0.02)	0.39 ± 0.02 (0.42 ± 0.02)	0.78 ± 0.03 (0.22 ± 0.02)	
c-kit2_SHORT	0.21 ± 0.01 (0.28 ± 0.02)	0.39 ± 0.02 (0.10 ± 0.02)	0.92 ± 0.01 (0.62 ± 0.02)	
c-kit*_LONG	0.22 ± 0.01 (0.54 ± 0.02)	0.39 ± 0.02 (0.33 ± 0.03)	0.72 ± 0.05 (0.13 ± 0.03)	
c-kit*_SHORT	0.21 ± 0.01 (0.22 ± 0.02)	0.37 ± 0.02 (0.18 ± 0.02)	0.94 ± 0.01 (0.60 ± 0.02)	

(about 10% and 18%, respectively). Consistently, the fraction of U is reduced in c-kit2_SHORT and c-kit*_SHORT (about 28% and 22%, respectively) when compared to the LONG constructs.

Overall, the presence of flanking ends seems to affect the folding equilibrium. Indeed, flanking ends appear to favour the folding into S1, a K^+ dependent conformation with a greater end-to-end distance, with reference to the shorter end-to-end distance S2. This effect, more pronounced in c-kit*_LONG than in c-kit2_LONG, might derive from the presence of the rigid and bulky double-stranded flanking ends directly at the terminals of the G4-forming sequence. *In vivo* a wider single stranded bubble can be expected during transcription or replication. However, the insertion of flexible single-stranded traits between the G4 ends and the double-stranded flanking ends would have significantly complicated the panel of configurations adoptable by the construct, possibly turning the interpretation of the FRET histograms cumbersome (e.g.: by making the introduction of a continuous interchormophore distance variation model function unavoidable). On the other hand, as derived from Figure S1, S3 and S4, c-kitX_LONG and c-kitX_TEMP retain the ability to fold in G4 although with some structural rearrangements with reference to the c-kitX_SHORT models that, in line with literature data, are more relevant on c-kit* (32).

Finally, for the S2 species the expectation value of the FRET efficiency is significantly higher in the c-

kitX_SHORT than in the c-kitX_LONG constructs, i.e.: the end-to-end distance is shorter for c-kitX_SHORT. This can derive from the different design of the two models. Indeed, in the c-kitX_LONG the two fluorophores are both located quite far from the terminals of the single stranded domains. Nevertheless, we cannot rule out a rearrangement of the constructs due to the sterically hindered tails in the c-kitX_LONG constructs or to the stacking of the chromophores to the external G-quartets in the c-kitX_SHORT constructs.

Folding kinetics of kit2. In a previously published article some of us reported that the folding process of the unlabelled oligonucleotide corresponding to the c-kit2 sequence comprises several kinetic intermediates whose formation was grouped into a fast and a slow step (30). The role of flanking ends in affecting such folding kinetics was quested here by time-lapse SM-FRET measurements. It is worth noting hereby that, whereas folding dynamics are commonly probed by SM-FRET by monitoring the behaviour of a specific molecule fixed at a surface as a function of time, this is to the best of our knowledge the first attempt to follow the progress of a folding reaction by applying SM-FRET on freely diffusing molecules. Namely, the process was started by adding 100 mM KCl to c-kit2_SHORT and c-kit2_LONG dissolved in 10 mM Tris–HCl buffer and the folding kinetics were reconstructed by acquiring FRET efficiency histograms on aliquots of the tested oligonucleotide solutions at increasing time lapses. After each acquisition, lasting for 180 s, the droplet of sample was substituted in order to minimize photobleaching. The histograms were fitted to a linear combination of U, S1 and S2 (keeping the values reported in Table 1 fixed) to estimate the fractional concentration of each population at every time step (which was the only parameter let free to vary). The obtained fractional concentrations over time are shown in Figure 3, panels A) and B) for c-kit2_LONG and c-kit2_SHORT, respectively.

To model the folding dynamics by means of differential rate equations and estimate transition rate constants, we assumed first order kinetics. This model is the simplest possible and pertains to systems in which aggregation dynamics are negligible and the chemical activities of the species involved in the reaction considered to be optimally approximated by the corresponding concentrations. Although c-kit2 dimer formation was detected at micromolar concentration range (30), it is more than reasonable that this phenomenon would negligibly take place in our experiments which are performed on extremely diluted solutions (100 pM). Moreover, considering that the equilibrium FRET efficiency distributions exhibit residuals of U and S1, the equilibrium must be the result of reversible conformational transitions. In view of these considerations, the most general rate equations describing a two-step folding kinetics are:

$$\begin{aligned}
\tfrac{d[\mathrm{U}]}{dt} &= -k_{\mathrm{U}\to\mathrm{S1}}[\mathrm{U}] - k_{\mathrm{U}\to\mathrm{S2}}[\mathrm{U}] + k_{\mathrm{S1}\to U}[\mathrm{S1}] + k_{\mathrm{S2}\to\mathrm{U}}[\mathrm{S2}] \\
\tfrac{d[\mathrm{S1}]}{dt} &= +k_{\mathrm{U}\to\mathrm{S1}}[\mathrm{U}] - k_{\mathrm{S1}\to\mathrm{U}}[\mathrm{S1}] - k_{\mathrm{S1}\to\mathrm{S2}}[\mathrm{S1}] + k_{\mathrm{S2}\to\mathrm{S1}}[\mathrm{S2}] \qquad (1) \\
\tfrac{d[\mathrm{S2}]}{dt} &= -k_{\mathrm{S2}\to\mathrm{S1}}[\mathrm{S2}] + k_{\mathrm{S1}\to\mathrm{S2}}[\mathrm{S1}] - k_{\mathrm{S2}\to\mathrm{U}}[\mathrm{S2}] + k_{\mathrm{U}\to\mathrm{S2}}[\mathrm{U}]
\end{aligned}$$

Extrapolation of kinetic constants and fitting curves was attempted by applying the least squares method to numerical solutions of the above rate equations. However, fitting the fractional concentrations to this model resulted computationally demanding. Although the convergence was cumbersome, it was apparent that direct conversion of U in S2 occurred at a negligible rate ($k_{\mathrm{U}\to\mathrm{S2}} < 10^{-8}$ s^{-1}). We thus proceeded to fit the data to the simplified model below:

$$\begin{aligned}
\tfrac{d[\mathrm{U}]}{dt} &= -k_{\mathrm{U}\to\mathrm{S1}}[\mathrm{U}] + k_{\mathrm{S1}\to U}[\mathrm{S1}] + k_{\mathrm{S2}\to\mathrm{U}}[\mathrm{S2}] \\
\tfrac{d[\mathrm{S1}]}{dt} &= +k_{\mathrm{U}\to\mathrm{S1}}[\mathrm{U}] - k_{\mathrm{S1}\to\mathrm{U}}[\mathrm{S1}] - k_{\mathrm{S1}\to\mathrm{S2}}[\mathrm{S1}] + k_{\mathrm{S2}\to S1}[\mathrm{S2}] \qquad (2) \\
\tfrac{d[\mathrm{S2}]}{dt} &= -k_{\mathrm{S2}\to\mathrm{S1}}[\mathrm{S2}] + k_{\mathrm{S1}\to\mathrm{S2}}[\mathrm{S1}] - k_{\mathrm{S2}\to\mathrm{U}}[\mathrm{S2}]
\end{aligned}$$

We note that numerical fitting of our data according to the above differential equations requires minimization of five independent parameters (the rates), and we dispose of 28 independent datapoints (the fractional concentrations of any two species, the third one being correlated). Thus, we can rely on 23 residual degrees of freedom. The best fit to the experimentally measured evolution of the *U, S1 and S2* populations over time is showed in Figure 3, while the inferred rate constants are listed in Table 2. For both samples, the overall folding process seems to advance through one fast transition, where S1 increases to the detriment of U, and a second slower transition, where S1 converts to S2. The values of $k_{\mathrm{U}\to\mathrm{S1}}$ and $k_{\mathrm{S1}\to\mathrm{U}}$ obtained from the fits are higher than the experiment time resolution, thus we can only safely state that these transitions occur on time scales of few minutes or faster. In spite of this indeterminacy, the estimated values of the $k_{\mathrm{S1}\to\mathrm{S2}}$, $k_{\mathrm{S2}\to\mathrm{S1}}$ and $k_{\mathrm{S2}\to\mathrm{U}}$ rate constants result to be stable with respect to parameters initialization. The values of $k_{S1\to\mathrm{S2}}$ and $k_{\mathrm{S2}\to S1}$ associated to the slower S1-to-S2 transition differ substantially in the c-kit2_LONG with respect to c-kit2_SHORT sample. In the former, they indicate that the time scale of both the forward and backward reaction is roughly hours. In the latter, a significantly faster forward transition and a negligibly improbable backward transition are determined. Moreover, the last estimated rate constant $k_{\mathrm{S2}\to\mathrm{U}}$ indicates that the back-transition from S2 to U is negligible in c-kit2_LONG, while it occurs within hours in c-kit2_SHORT. A more intuitive interpretation of these values can be given by considering the inverse reaction starting from a sample consisting in pure *S*2-folded construct. The thermodynamics of the system would dictate recovery of the three-species equilibrium with the same fractional concentrations measured in our experiments. The time needed to reach equilibrium would depend on the height of the potential barriers between the different conformations adopted by the constructs, and promises to be longer than that observed for the direct reaction, in view of the extrapolated values of the inverse rate constants $k_{\mathrm{S2}\to\mathrm{S1}}$ and $k_{\mathrm{S2}\to\mathrm{U}}$. However, the inverse reaction would reach equilibrium through different pathways in the case of ckit2_LONG and ckit2_SHORT. Namely, while the first construct would mainly adopt a two-step transition passing through the intermediate *S*1, the second would proceed towards direct unfolding of *S*2, followed by formation of *S*1 from the coiled state.

The distribution of U, S1 and S2 derived from kinetic experiments at 24 h perfectly matches with the data acquired on the equilibrated samples.

This overall picture fits with the evolution of the monomeric species derived by UV and CD experiments performed on the unlabelled short sequences (30).

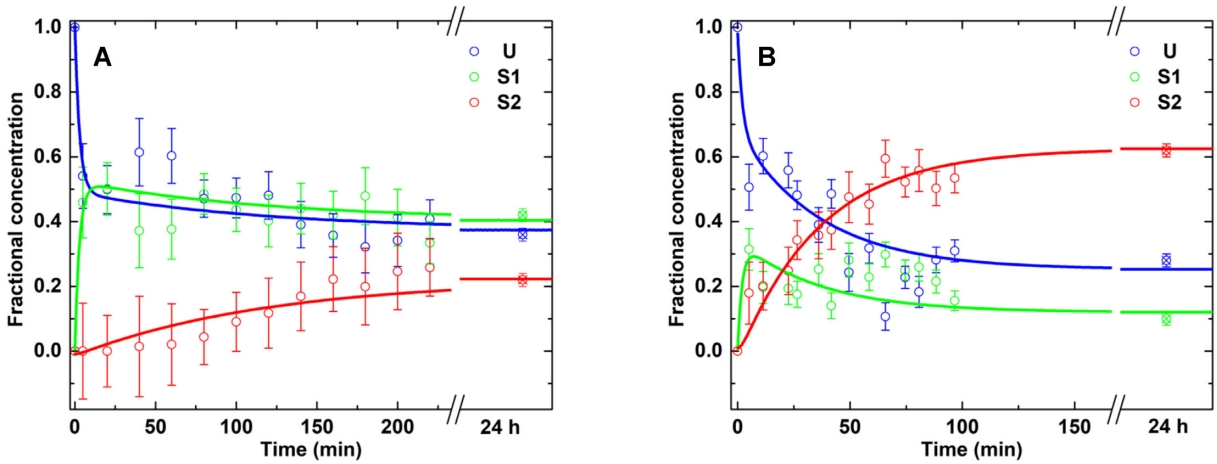

Figure 3. Relative amplitude of U, S1 and S2 versus time for (**A**) c-kit2_LONG and (**B**) c-kit2_SHORT. Dots represent experimental data; lines represent the fitting curves. The points at abscissa 24 h represent the equilibrium values reported in Table 2 for the samples at equilibrium: the fitting curves optimally interpolate these values. The solutions used for the kinetics experiments also attained the same value after overnight incubation at room temperature within the experimental errors.

Table 2. Kinetic constants at 25°C for the folding of c-kit2_SHORT and c-kit2_LONG upon addition of 100 mM KCl to the buffer.

Sample	$k_{U\rightarrow S1}$ ($\times 10^5$) [s^{-1}] $\tau_{U\rightarrow S1}$ [min]	$k_{S1\rightarrow U}$ ($\times 10^5$) [s^{-1}] $\tau_{S1\rightarrow U}$ [min]	$k_{S1\rightarrow S2}$ ($\times 10^5$) [s^{-1}] $\tau_{S1\rightarrow S2}$ [min]	$k_{S2\rightarrow S1}$ ($\times 10^5$) [s^{-1}] $\tau_{S2\rightarrow S1}$ [min]	$k_{S2\rightarrow U}$ ($\times 10^5$) [s^{-1}] $\tau_{S2\rightarrow U}$ [min]
c-kit2_SHORT	≈270 ≈ 6	≈480 ≈ 3	85.5 ± 1.2 ≈20	0.021 ± 0.003 8×10^4 (55 days)	16.4 ± 3 100
c-kit2_LONG	≈300 ≈ 5	≈280 ≈ 6	5.95 ± 0.14 4 h 40	9.82 ± 0.26 170	0.95 ± 0.08 1750

The folding dynamics of both c-kit*_SHORT and c-kit*_LONG could not be followed by means of single molecule FRET experiments because the samples reached dynamic equilibrium (i.e.: the FRET efficiency histograms are equal to those measured at equilibrium within the experimental errors) in a time shorter than 4 min, which is our temporal resolution.

Time-correlated single-photon counting

In parallel with SM-FRET studies, fluorescence decay measurements were conducted on the same specimens exploiting a TCSPC setup endowed with 30 ps time resolution. The fluorescence decay analysis is the most suited bulk technique to discriminate subpopulations within a heterogeneous ensemble by means of FRET. In this section the TCSPC data are commented and compared with those obtained by SM-FRET. In the absence of KCl, the fluorescence decay of c-kit2_LONG is nicely described by a single exponential, with a lifetime of ≈3.7 ns. In 100 mM KCl a double exponential model is necessary in order to accurately reproduce the fluorescence decay curve, with long decay time of ≈3.6 ns (relative amplitude 0.72) and short decay time of ≈1.2 ns (Figure 4A). A similar behaviour is displayed by the fluorescence decay patterns reconstructed for c-kit*_LONG (see Figure 4B). Namely, in the presence of KCl, a two-exponential model is required to reproduce the decay distribution accurately, with short decay time ≈1.2 ns (accounting for 29% of the oligos) and long decay time 3.5 ns. Conversely, the decay pattern obtained in the absence of salt is optimally fitted by a single-exponential, with time constant 3.7 ns (see Table S3 in the Supplementary Materials Section for details on the fitting parameters).

It is worth to recall that the FRET efficiency can be estimated starting from fluorescence decay data by applying the following equation:

$$E_{FRET} = 1 - \tau_{D,A}/\tau_D, \quad (3)$$

where $\tau_{D,A}$ and τ_D are the D lifetime values measured in the presence and in the absence of the A, respectively. The latter parameter can be estimated from the decay data in the absence of KCl, by assuming that the single decay time detected in this instance corresponds to the average FRET efficiency measured by SM-FRET. Namely,

$$\tau_D = \frac{\tau_{NoKCl}}{(1 - \langle E_{FRET}\rangle_{NoKCl})}. \quad (4)$$

Accordingly, the short decay component resolved upon addition of KCl can be ascribed to formation of the KCl-depending folded structures S1 and S2, and corresponds to the high-FRET efficiency peak appearing in these buffer conditions in the SM-FRET distributions. This is confirmed by the fact that the fluorescence decay distributions of BlankX are single exponential both in the presence and in the absence of KCl (see Figure S12 in the Supplementary Materials Section).

Taken into account that TCSPC, being an ensemble technique, is inherently less sensitive with respect to the ability to discriminate among different subpopulations within a specimen with respect to SM-FRET, the results obtained with the two techniques are compatible even on a quantita-

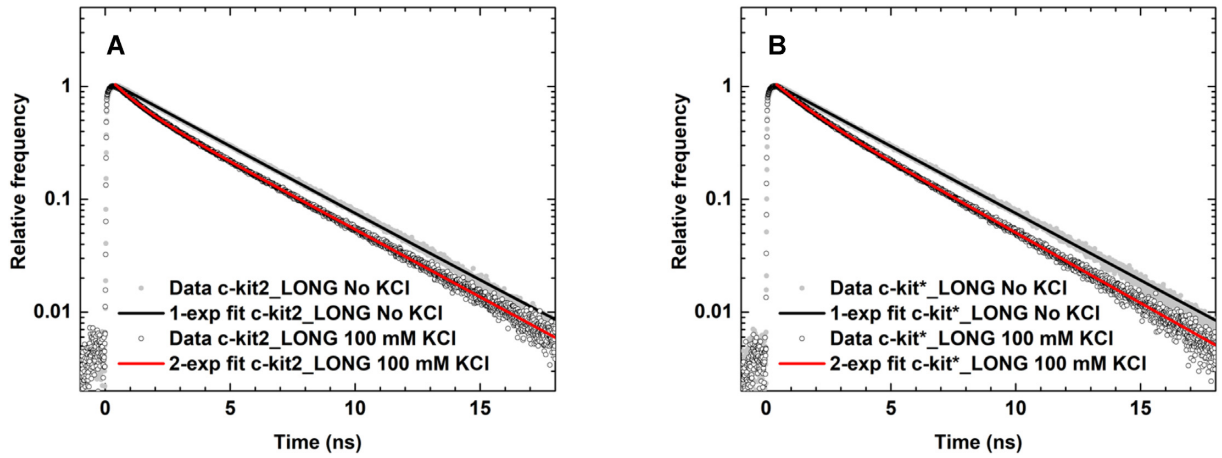

Figure 4. Decay patterns and fitting curves for (**A**) c-kit2_LONG and (**B**) c-kit*_LONG in the absence (full gray dots) and presence (black empty circles) of 100 mM KCl.

tive standpoint. Indeed, fitting the FRET efficiency distributions obtained for c-kitX_LONG in the presence of KCl to a two Gaussian model function (see Figure S13 in the Supplementary Materials Section) results in the retrieval of peak FRET efficiency values well compatible with those estimated from the pertaining decay time values by applying Equation (3), with similar relative amplitudes in the two instances (see Table S4 in the Supplementary Materials Section).

Moreover, although we did not manage to discriminate the three populations resolved in the SM-FRET experiments by means of an unconstrained three-exponential fit of the decay distributions obtained in the case of c-kitX_LONG in the presence of KCl, fixing (through application of Equations (3) and (4)) the decay times to values corresponding to the peak FRET efficiencies measured for U, S1 and S2 by SM-FRET experiments led to good-quality fits in which the (unconstrained) relative amplitudes converged to values similar to those estimated by multi-Gaussian analysis of the corresponding SM-FRET efficiency distributions (see Figure S14 and Table S5 in the Supplementary Materials Section).

The TCSPC data are more interlocutory in the case of c-kitX_SHORT, inasmuch as fluorescence decays are multi-exponential even in the absence of KCl. Efficient non-radiative decay of the D due to guanine quenching (40) may play a relevant role in biasing the excited-state dynamics.

Under these conditions, to probe consistency of the TCSPC with the SM-FRET data, we fitted the decays to multi-exponential model functions by setting the number of transients according to the criterion reported in Materials and Methods (see patterns and fits in Figure S15 and corresponding best-fitting parameters in Table S6 of the Supplementary Materials Section). We calculated the average fluorescence lifetime according to the formula

$$\langle \tau \rangle = \sum_{i=1}^{n} f_i \tau_i \qquad (5)$$

and compared the ratio of the average lifetimes obtained in the absence and presence of KCl with the corresponding $(1 - \langle E_{FRET} \rangle_{NoKCl})/(1 - \langle E_{FRET} \rangle_{KCl})$ ratios, where $\langle E_{FRET} \rangle_{NoKCl}$ and $\langle E_{FRET} \rangle_{KCl}$ are the average FRET efficiency values extrapolated by integration of the FRET efficiency distributions measured by SM-FRET in Tris–HCl and in 100 mM KCl, respectively. The values of these observables are very similar, as shown in Table 3, testifying the compatibility of the two datasets. This holds true also in the case of c-kitX_LONG.

In conclusion, the TCSPC data confirm by means of independent analyses the findings of the SM-FRET data, conferring robustness to the results discussed hereafter.

DISCUSSION

In this article the folding equilibria and dynamics of oligonucleotide constructs that mimic the sequences of two G4-forming sites (c-kit2 and c-kit*) located in the proximal promoter of the *KIT* proto-oncogene were investigated at the single molecule level. SM-FRET experiments were systematically run on two different construct models: c-kitX_SHORT, in which the D and A are directly tethered to the 5′- and 3′- ends of the minimal G4-folding sequence, and c-kitX_LONG, in which the G4-forming sequence is flanked at 3′ and 5′ by double-stranded stretches. On both models, a similar, single population was observed in the samples dissolved in TRIS-HCl with or without the addition of 100 mM LiCl. (Figure S10 in the Supplementary Materials Section). At difference, multi-peak analysis of the experimental FRET-efficiency probability distributions obtained in the presence of KCl revealed the coexistence of three different conformations (U, S1 and S2). We proved that these behaviours were independent from the specificities of the D-A pair. Indeed, SM-FRET measurements performed on c-kitX_LONG labelled with the pair Atto532-Atto594 (see Figure 1) in the presence/absence of KCl yielded FRET-efficiency histograms (see Figure S16 in the Supplementary Materials Section) very similar to those obtained with the pair Atto532-Atto647N (Figure 2, Supplementary Figures S7, S8).

The lowest-FRET-efficiency population that was resolved in the histograms acquired in the presence of KCl (U

Table 3. Comparison of mean fluorescence lifetime obtained by TCSPC and average FRET efficiency values measured by SM-FRET.

Sample	$< \tau >_{NoKCl}$ [ps]	$< \tau >_{KCl}$ [ps]	$\frac{\langle\tau\rangle_{NoKCl}}{\langle\tau\rangle_{KCl}}$	$<E>_{NoKCl}$	$<E>_{KCl}$	$\frac{1-\langle E\rangle_{NoKCl}}{1-\langle E\rangle_{KCl}}$
c-kit2_LONG	3608 ± 3	2900 ± 42	1.24 ± 0.02	0.246 ± 0.003	0.402 ± 0.004	1.26 ± 0.01
c-kit*_LONG	3661 ± 3	2860 ± 66	1.28 ± 0.03	0.225 ± .003	0.336 ± 0.003	1.17 ± 0.07
c-kit2_SHORT	2118 ± 66	1.200 ± 59	1.77 ± 0.10	0.235 ± 0.004	0.565 ± 0.004	1.76 ± 0.02
c-kit*_SHORT	1908 ± 10	1088 ± 30	1.75 ± 0.05	0.203 ± 0.002	0.563 ± 0.008	1.83 ± 0.03

in the Results Section) was very similar to the single population observed in K^+-lacking samples. Thus, it was ascribed to residuals of unfolded. In order to investigate the other two species (i.e.: those denoted as S1 and S2 above), measurements in Tris–HCl, Tris–HCl added with LiCl, and TRIS-HCl added with KCl were performed on the BlankX samples, which were unable to assume the G4 structure due to point mutations (Figure S9 and S10). These constructs exhibited FRET distributions virtually consisting of the U population only in any of the probed environments. These data point to S1 and S2 as potassium and sequence dependent species, thus allowing us to address them as G4 folded species.

To derive estimations of the D-A distances in the different conformations resolved within the FRET efficiency distributions from the pertaining peak FRET efficiency values we considered the Förster relation:

$$E = \frac{1}{1 + (R/R_0)^6} \tag{6}$$

A preliminary observation needs to be made in this regard. The interchromophore distance values R extracted from Equation (6) should be considered with caution, only as rough estimations of the G4s end-to-end distances. Indeed, the determination of R is affected by several experimental caveat. For this reason, SM-FRET has been historically exploited as a powerful method to determine the degree of heterogeneity in solutions containing multiple species, i.e.: number and fractional concentration of subpopulations (47), while it is common practice since the earliest times of SM-FRET experiments to give up quantitative assessments on the D-A distances (48). Indeed, fluorescence crosstalk might be in some instances a major source of systematic errors in the measurement of interchormophore distances, as photon counts in the A detection channel are overestimated due to this phenomenon. However, in our experimental conditions (D-A pairs and optical settings) the crosstalk has been proven to be negligible (Figure S6). Other systematics were taken into account by means of introduction in the data analysis of the ϕ_{rel} parameter (see Materials and Methods), including:

- Different detection quantum efficiency of the single-photon detectors in the D and A emission band
- Fabrication specificities of different detection modules even of the same brand and model
- Different spectral properties of the D and A dyes (spectral width, absorption cross section, fluorescence quantum yield…)
- Different optical properties of the filters used to select the D and A emission
- Different response of all the optical elements of the SM-FRET setup at the significantly different peak emission wavelengths of the D and A

Although the value attributed to the ϕ_{rel} correction parameter was carefully evaluated starting from extensive information on the chromophores, optics and detectors provided by the respective suppliers, unavoidable differences in optimization of the alignment of the collection optics and detector in the D versus A channel may bring about significant stochastic errors. Moreover, in order to recover correct end-to-end distance estimations from interchromophore distance calculations, the lengths of the tethers linking the D and A to the constructs should be known, but precise information on their values was not provided by the supplier. A direct comparison between the conformations of the c-kitX_LONG constructs and their homologous c-kitX_SHORT is also complicated by the different labelling position. However, perhaps the most severe sources of bias in the correct determination of the end-to-end distances adopted by the samples in the different conformations are the following, conformation dependent phenomena:

- Indeterminacy in the value of the k^2 reorientation factor appearing in the equation for R_0, which prevents determination of the Forster radius value whenever the D and A are not free to reorient with respect to each other, as it is likely in the case under discussion, when they are encompassed within a rigid and bulky biomolecular construct.
- Dye/base interactions (such as π-π stacking or guanine quenching) influencing unpredictably the spectroscopic properties of the D and A, specially their fluorescence quantum yield.

The best we could do in order to corroborate our findings on the c-kitX_LONG constructs was to check them upon substitution of the D-A pair. This is a very strong method to validate the qualitative features of our FRET distributions. Indeed, the benchmark properties claimed for the ckitX_LONG constructs are clearly evidenced with both Atto594 and Atto647N as the A. Moreover, the peak position of the three populations resolved with the two pairs are equal within experimental errors. This encouraged us to attempt a semi-quantitative analysis of our data in spite of the above-recalled caveats. Some interesting considerations arose from such attempt, which are detailed hereafter.

In the c-kitX_SHORT models, the high FRET efficiency species S2 were the most abundant Thus, we were tempted to associate them to the NMR solved available monomeric G4 structures for kit2 and kit*. Although quantitative assessment of the D-A distance R from the FRET efficiency value is reliable only for R values in the neighbourhood of the Förster radius R_0, i.e. for E values not

too distant from 0.5, we applied the Förster relation to test our hypothesis. From NMR data, the end-to-end distance of the monomeric parallel G4 of c-kit2 is ≈18 Å (49) and drop to 9.7 in the antiparallel two-tetrads G4 of c-kit*, thanks to the fold-back of the 3′ tail (32). Assuming that the theoretical value for the Förster radius of the Atto532-Atto647N D-A pair (59 Å, as calculated by the supplier (https://www.atto-tec.com/images/ATTO/Katalog_Preisliste/Katalog_2019_2020.pdf) holds true even when the dyes are linked to a DNA sequence, it would result in a FRET efficiency ≈100% for both G4s. However, considering that this rough estimation neglects putative dye-G4 stacking effects and tethers lengths, this value is not too different from the peak FRET efficiencies 0.92 and 0.94 measured in our experiments for c-kit2_SHORT and c-kit*_SHORT, respectively. Interestingly, the structure endowed with the higher D-A distance (c-kit2) actually produces a slightly lower peak FRET efficiency than the one characterized by nearest ends (c-kit*).

The peak FRET efficiency for the S2 species is shifted towards lower values in the c-kitX_LONG specimens. The lower FRET efficiency may reflect an incremented distance of the two fluorophores due to their insertion within the two negatively charged double stranded tails. Equation (6) gives an interchromophore distances of roughly 48 Å and 49 Å for the FRET efficiency values of 0.78 and of 0.76, detected for c-kit2_LONG and c-kit*_LONG, respectively. Taking into account that in these constructs the D and A are attached five base pairs (i.e. 3.3 Å/bp × 5 bp = 16.5 Å) upstream the 5′-end and six base pairs (i.e. 3.3 Å/bp × 6 bp = 19.8 Å) downstream the 3′-end of the G4 forming sequence, respectively, and assuming that the G4 topology is conserved upon flanking ends addition (50), the overall expected distance rises up to 54.5 Å and 46 Å for c-kit2_LONG and c-kit*_LONG, respectively. These values lead us to conclude that the double-stranded traits are on the average nearly maximally far apart to one another on both constructs. Interestingly, as shown in Figure S4a), for c-kit2 the chiroptical features of the long and short constructs overlap with each other, thus pointing to S2 as the conserved parallel G4 previously solved for the minimal-length oligonucleotide. Thus, for c-kit2, the existence of repulsive interactions between the double-stranded flanking ends does not seem to have a relevance in tuning the S2 conformation. It rather has a notable effect on the stability of S2 relative to S1, since the fractional population of S2 is remarkably reduced at the advantage of the S1 fractional population in c-kit2_LONG with respect to c-kit2_SHORT, as detailed in Table 1.

It is worth to compare these result to those previously reported by Balasubramanian's group (26,29) that used kit2 fused to a double stranded segment only at 3′. On these constructs a folded species with FRET efficiency comparable to the one measured herein for c-kit2_SHORT was detected. Only when a flank at the 5′ terminal was added the efficiency was reduced to a value comparable to those herein reported for c-kit2_LONG. Interestingly, it occurs either when the central G-rich domain is fully paired to the complementary one (as in Balasubramanian model) or not (as in this work): this suggests that the presence of the unpaired C-rich strand does not significantly affect the folding distribution of G-quadruplexes in c-kit2.

We can try to extend the same evaluation to c-kit*. It was already observed that both elongation at 5′ and the impairment of proper folding back of the 3′ tail can significantly affect the G4 folding of this trait (32). Accordingly, the destabilization of the S2 folded fraction is even more evident than for c-kit2. Moreover, as shown in Figure S4b), the CD profile showed by c-kit*_LONG suggests a conformational rearrangement that reminds those observed when the above mentioned modifications occur. Thus, in this case, the S2 form might not be shared by ckit*_LONG and ckit*_SHORT.

Another interesting aspect of this work refers to the application of SM-FRET experiments on freely diffusing molecules to kinetic analyses. Interestingly, the kinetic parameters describing the folding process of kit2_SHORT and kit2_LONG derived herein are in good agreement with those previously reported for the short sequence without labelling groups (30,38). As such, we can attribute S1 to the intermediate folding species that has been reported to present a higher antiparallel content. Remarkably, SM-FRET allowed us to prove that in the c-kit2_LONG construct this intermediate has a reduced ability to convert into the thermodynamically more stable S2 form if compared to the isolated single stranded c-kit2. Although application of this technique to elucidate conformational dynamics is presently hindered by poor temporal resolution, this proof of principle could foster the development of technical strategies and instrumentation capable of overcoming this limit.

On an exquisitely technical basis, the above reported analyses exemplify the superior power of single-molecule techniques in discriminating subpopulations within a heterogeneous sample with respect to ensemble methods (e.g.: TCSPC). On a more fundamental and general viewpoint, the measurements presented in this article put into question the widely applied assumption that G4 topologies can be safely probed by analysing *in vitro* the conformations adopted by single-stranded oligonucleotides comprising only the minimal sequence required for G4 formation, with no consideration for the highly ordered DNA matrix they are inserted in, a condition highly relevant in particular for those G4s occurring within gene promoters. This evidence might have a strong impact on the methodological approaches that should be adopted in G4s structural studies, which in turn may influence the knowledge on their structure/function relationships as well as the development of G4 targeting drugs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

University of Padova [SISS_SID19_01]; C.A.M. acknowledges partial support of the Italian Ministry of University and Research (MIUR)—Department of Excellence project PREMIA (PREcision MedIcine Approach: bringing biomarker research to clinic). Funding for open access charge: Standard institutional funds from University of Milan Bicocca.

Conflict of interest statement. None declared.

REFERENCES

1. Huppert,J.L. (2008) Hunting G-quadruplexes. *Biochimie*, **90**, 1140–1148.
2. Huppert,J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, **37**, 1375–1384.
3. Hansel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
4. Kwok,C.K. and Merrick,C.J. (2017) G-quadruplexes:prediction, characterization, and biological application. *Trends Biotech.*, **35**, 997–1013.
5. Juranek,S.A. and Paeschke,K. (2012) Cell cycle regulation of G-quadruplex DNA structures at telomeres. *Curr. Pharmac. Des.*, **18**, 1867–1872.
6. Wang,Q., Liu,J., Chen,Z., Zheng,K., Chen,C., Hao,Y. and Tan,Z. (2011) G-quadruplex formation at the 3′ end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.*, **39**, 6229–6237.
7. Johnson,J.E., Smith,J.S., Kozak,M.L. and Johnson,F.B. (2008) In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie*, **90**, 1250–1263.
8. Eddy,J., Vallur,A.C., Varma,S., Liu,H., Reinhold,W.C., Pommier,Y. and Maizels,N. (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.*, **39**, 4975–4983.
9. Kim,N. (2019) The interplay between G-quadruplex and transcription. *Curr. Med. Chem.*, **26**, 2898–2917.
10. Lightfoot,H.L., Hagen,T., Tatum,N.J. and Hall,J. (2019) The diverse structural landscape of quadruplexes. *FEBS Lett.*, **593**, 2083–2102.
11. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
12. Yang,D. and Chen,Y. (2012) Sequence, stability, and structure of G-quadruplexes and their interactions with drugs. *Curr. Prot. Nucl. Acids Chem.*, **50**, 17.5.1–17.5.17.
13. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
14. Dettler,J.M., Buscaglia,R., Le,V.H. and Lewis,E.A. (2011) DSC deconvolution of the strucutral complexity of c-MYC P1 promoter G-quadruplex. *Biophys. J.*, **100**, 1517–1525.
15. Sun,D. and Hurley,L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplexes and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **52**, 2863–2874.
16. Ruggiero,E. and Richter,S.N. (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.*, **46**, 3270–3283.
17. Che,T., Wang,Y.-Q-., Huang,Z.-L., Tan,J.-H., Huang,Z.-S. and Chen,S.-B. (2018) Natural alkaloids and heterocycles as G-quadruplex ligands and potential anticancer agents. *Molecules*, **23**, 493.
18. Sengupta,A., Ganguly,A. and Chowdhury,S. (2019) Promise of G-quadruplex structure binding ligands as epigenetic modifiers with anti-cancer effects. *Molecules*, **24**, 582.
19. Asamitsu,S., Obata,S., Yu,Z., Bando,T. and Sugiyama,H. (2019) Recent progress of targeted G-quadruplex-preferred ligands towards cancer therapy. *Molecules*, **24**, 429.
20. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug. Discov.*, **10**, 261–275.
21. Chen,B., Liang,J., Tian,X. and Liu,X. (2008) G-quadruplex structure: a target for anticancer therapy and a probe for detection of potassium. *Biochem.*, **73**, 853–861.
22. Muller,S., Laxmi-Reddy,K., Jena,P.V., Baptiste,B., Dong,Z., Godde,F., Ha,T., Rodriguez,R., Balasubramanian,S. and Huc,I. (2014) Targeting DNA G-quadruplexes with helical small molecules. *ChemBioChem*, **15**, 2563–2570.
23. Harkness,R.W. and Mittermaier,A.K. (2017) G-quadruplex dynamics. *Biochim. Biophys. Acta – Proteins And Proteomics*, **1865**, 1544–1554.
24. Ceschi,S. and Sissi,C. (2020) Kit promoter: Structure, function and targeting. In: Neidle,S. (ed). *Quadruplex Nucleic Acid as Target for Medicinal Chemistry*. Elsevier, Amsterdam, Vol. **LIV**, pp. 409–439.
25. 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.
26. Fegan,A., Shirude,P.S., Ying,L. and Balasubramanian,S. (2010) Ensemble and single molecule FRET analysis of the structure and unfolding kinetics of the c-kit promoter quadruplexes. *Chem. Commun.*, **46**, 946–948.
27. 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. *Biochem.*, **45**, 7854–7860.
28. Hsu,S.-T. D., 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.
29. Shirude,P.S., Okumus,B., Ying,L., Ha,T. and Balasubramanian,S. (2007) Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc.*, **129**, 7484–7485.
30. Rigo,R., Dean,W.L., Gray,W.D., Chaires,J.B. and Sissi,C. (2017) Conformational profiling of a G-rich sequence within the c-KIT promoter. *Nucleic Acids Res.*, **45**, 13056–13067.
31. 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.
32. 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.
33. Buglione,E., Salerno,D., Marrano,C.A., Cassina,V., Vesco,G., Nardo,L., Dacasto,M., Rigo,R., Sissi,C. and Mantegazza,F. (2021) Nanomechanics of G-quadruplexes within the promoter of the c-kit oncogene. *Nucleic Acids Res.*, **49**, 4564–4573.
34. Noer,S.L., Preus,S., Gudnason,D., Aznauryan,M., Mergny,J.L. and Birkedal,V. (2016) Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na+ solutions by single molecule FRET microscopy. *Nucleic Acids Res.*, **44**, 464–471.
35. König,I., Zarrine-Afsar,A., Aznauryan,M., Soranno,A., Wunderlich,B., Dingfelder,F., Stüber,J.C., Plückthun,A., Nettels,D. and Schuler,B. (2015) Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nature Meth.*, **12**, 773–779.
36. Soranno,A., Holla,A., Dingfelder,F., Nettels,D., Makarov,D.E. and Schuler,B. (2017) Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E1833–E1839.
37. Schuler,B, Soranno,A, Hofmann,H and Nettels,D. (2016) Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Ann. Rev. Biophys.*, **45**, 207–231
38. Ceschi,S., Largy,E., Gabelica,V. and Sissi,C. (2020) A two-quartet G-quadruplex topology of human KIT2 is conformationally selected by a perylene derivative. *Biochimie*, **179**, 77–84.
39. Rigo,R. and Sissi,C. (2017) Characterization of G4-G4 crosstalk in the c-KIT promoter region. *Biochemistry*, **56**, 4309–4312.
40. Arora,A., Nair,D.R. and Maiti,S. (2009) Effect of flanking bases on quadruplex stability and Watson-Crick duplex competition. *The FEBS Journal*, **276**, 3628–3640.
41. Nardo,L., Lamperti,M., Salerno,D., Cassina,V., Missana,N., Bondani,M., Tempestini,A. and Mantegazza,F. (2015) Effects of non-CpG site methylation on DNA thermal stability: a fluorescence study. *Nucleic Acids Res.*, **43**, 10722–10733.
42. Da Ros,S., Nicoletto,G., Rigo,R., Ceschi,S., Zorzan,E., Dacasto,M., Giantin,M. and Sissi,C. (2021) G-Quadruplex modulation of SP1 functional binding sites at the *KIT* proximal promoter. *Int. J. Mol. Sci.*, **22**, 329.
43. Lilletvedt,M., Tønnesen,H.H., Høgset,A., Kristensen,S. and Nardo,L. (2010) Time-domain evaluation of drug-solvent interactions of the photosensitizers TPCS2a and TPPS2a as part of

physicochemical characterization. *J. Photochem. Photobiol., A: Chem.*, **214**, 40–47.
44. Nardo,L., Re,F., Brioschi,S., Cazzaniga,E., Orlando,A., Minniti,S., Lamperti,M., Gregori,M., Cassina,V., Brogioli,D., Salerno,D. and Mantegazza,F. (2016) Fluorimetric detection of the earliest events in amyloid β oligomerization and its inhibition by pharmacologically active liposomes. *Biochim. Biophys. Acta Gen. Subj.*, **1860**, 746–756.
45. Stromqvist,J., Nardo,L., Broekmans,O., Kohn,J., Lamperti,M., Santamato,A., Shalaby,M., Sharma,G., Di Trapani,P., Bondani,M. and Rigler,R. (2011) Binding of biotin to streptavidin: a combined fluorescence correlation spectroscopy and time-resolved fluorescence study. *Eur. Phys. J. Special Topics*, **199**, 181–194.
46. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
47. Weiss,S. (1999) Fluorescence spectroscopy of single biomolecules. *Science*, **283**, 1676–1683.
48. Ying,L., Wallace,M.I., Balasubramanian,S. and Klenerman,D (2000) Ratiometric analysis of single-molecule fluorescence resonance energy transfer using logical combinations of threshold criteria: a study og 12-mer DNA. *J. Phys. Chem. B*, **104**, 5171–5178.
49. 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.
50. Salsbury,A.M., Dean,T.J. and Lemkul,J.A. (2020) Polarizable molecular dynamics simulations of two c-kit oncogene promoter G-quadruplexes: effect of primary and secondary structure on loop and ion sampling. *J. Chem. Theory Comput.*, **16**, 3430–3444.