



**HAL**  
open science

## Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex

Phong Lan Thao Tran, Martin Rieu, Samar Hodeib, Alexandra Joubert, Jimmy Ouellet, Patrizia Alberti, Anthony Bugaut, Jean-François Allemand, Jean-Baptiste Boulé, Vincent Croquette

### ► To cite this version:

Phong Lan Thao Tran, Martin Rieu, Samar Hodeib, Alexandra Joubert, Jimmy Ouellet, et al.. Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex. Nucleic Acids Research, 2021, 49 (9), pp.5189-5201. 10.1093/nar/gkab306 . hal-03214603

**HAL Id: hal-03214603**

**<https://hal.science/hal-03214603v1>**

Submitted on 2 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex

Phong Lan Thao Tran<sup>1</sup>, Martin Rieu<sup>2,3</sup>, Samar Hodeib<sup>2,3</sup>, Alexandra Joubert<sup>1</sup>, Jimmy Ouellet<sup>4</sup>, Patrizia Alberti<sup>1</sup>, Anthony Bugaut<sup>1</sup>, Jean-François Allemand<sup>2,3</sup>, Jean-Baptiste Boulé<sup>1,\*</sup> and Vincent Croquette<sup>2,3,5,\*</sup>

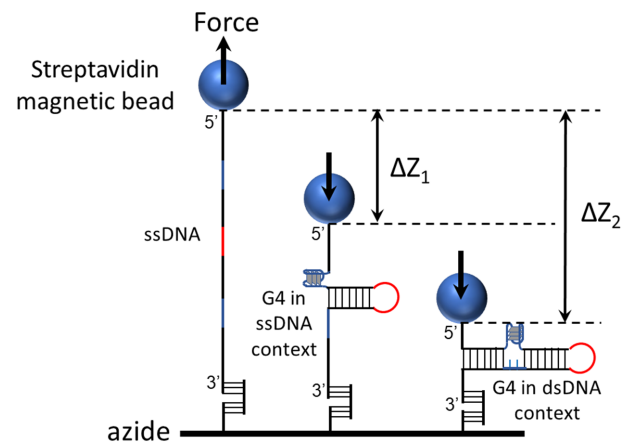
<sup>1</sup>Structure et Instabilité des Génomes, Museum National d'Histoire Naturelle, INSERM, CNRS, Alliance Sorbonne Université, 75005 Paris, France, <sup>2</sup>Laboratoire de physique de L'École Normale Supérieure de Paris, CNRS, ENS, Université PSL, Sorbonne Université, Université de Paris, 75005 Paris, France, <sup>3</sup>Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France, <sup>4</sup>Depixus SAS, 3-5 Impasse Reille, 75014 Paris, France and <sup>5</sup>ESPCI Paris, Université PSL, 75005 Paris, France

Received October 21, 2020; Revised April 09, 2021; Editorial Decision April 13, 2021; Accepted April 14, 2021

## ABSTRACT

G-quadruplex (G4) DNA structures have emerged as important regulatory elements during DNA metabolic transactions. While many *in vitro* studies have focused on the kinetics of G4 formation within DNA single-strands, G4 are found *in vivo* in double-stranded DNA regions, where their formation is challenged by the complementary strand. Since the energy of hybridization of Watson-Crick structures dominates the energy of G4 folding, this competition should play a critical role on G4 persistence. To address this, we designed a single-molecule assay allowing to measure G4 folding and persistence times in the presence of the complementary strand. We quantified both folding and unfolding rates of biologically relevant G4 sequences, such as the cMYC and cKIT oncogene promoters, human telomeres and an avian replication origin. We confirmed that G4s are found much more stable in tested replication origin and promoters than in human telomere repeats. In addition, we characterized how G4 dynamics was affected by G4 ligands and showed that both folding rate and persistence time increased. Our assay opens new perspectives for the measurement of G4 dynamics in double-stranded DNA mimicking a replication fork, which is important to understand their role in DNA replication and gene regulation at a mechanistic level.

## GRAPHICAL ABSTRACT



## INTRODUCTION

Nucleic acid sequences with four or more stretches of multiple guanines ( $G_nN_xG_nN_xG_nN_xG_n$ ;  $n \geq 2$ ,  $x \leq 7$ ) can form non-canonical four-stranded structures called G-quadruplexes (G4s). These structures are assembled by stacking of coplanar guanine's quartets (G-quartets) that are stabilized by monovalent cations, such as  $K^+$  or  $Na^+$ . G4 structures are well described at the structural level. Depending on their sequence, they can adopt different folding patterns and can form from the same (intramolecular) or distinct (intermolecular) nucleic acid strands (1–3). Potential G4-forming motifs are widespread across the genomes of a large spectrum of organisms from bacteria to mammals (4–8), and the formation of stable G4 structures *in vivo* has been linked to replication (9–11), oncogene expression (12–14),

\*To whom correspondence should be addressed. Tel: +33 140795616; Fax: +33 140793705; Email: jean-baptiste.boule@mnhn.fr  
Correspondence may also be addressed to Vincent Croquette. Email: vincent.croquette@phys.ens.fr

translation (15–17), DNA repair (18,19) telomere maintenance (20) or epigenetic marking of a DNA locus (21–23). Questions around the physiological conditions conducive to their formation and their biological roles have made G4 structures the focus of extensive *in vitro* and *in vivo* studies over the last two decades.

Single-stranded (ss) nucleic acid sequences, such as the telomeric 3'-overhang or RNAs, are prone to G4 formation with no challenge from a complementary strand. However, G4 forming sequences are allegedly also found in double-stranded (ds) DNA regions, such as promoters (24), minisatellites (25) or replication origins (10). These structures, if stable and unprocessed by helicases (26), may hinder the replication fork and transcription bubble progression, causing fork pausing or promoting DNA breakage (25,27,28). Despite causing potential roadblock for molecular motors, current wisdom is that their formation has been harnessed by evolution to encode structural information within DNA, at the expense of evolving protein motors able to remove the stable structures challenging replication (26,29).

Their potential biological effect would require their persistence post replication in a context where they are embedded in a dsDNA bubble. Persistence of G4 structures within dsDNA is largely unknown, but may depend on their sequence, the genomic context and/or the presence of a bound protein (30–36). Convincing evidence of G4 formation *in vivo* has indeed been reported near several transcription start sites (37) or near-replication origins where G4 structures could participate in transient fork pausing or loading of proteins (10,11,38). All these roles suggest that once formed these structures have a significant persistence time in a dsDNA context.

To date, our knowledge of the conditions of G4 formation is largely inferred from *in vitro* biophysical studies of different model G4-forming single-stranded sequences. These approaches have contributed largely to our understanding of G4 stability and structural diversity, but are limited in providing a complete biologically relevant picture of the dynamics of G4 structures at the molecular level.

Folding/unfolding of several G4s (such as human telomeric or cMYC promoter sequences) have been studied in recent years using magnetic or optical tweezers (39–43). Most of these studies address the dynamics of G4 in a ssDNA context. In a non-replicative dsDNA context, the presence of a complementary strand causes a competition between the G4 and the duplex structure. At the thermodynamic level, the energy of hybridization of a Watson-Crick duplex containing a G-rich DNA sequence is much larger (~50 kcal/mol) than the energy of G4 folding (~4–8 kcal/mol) (1). Therefore, the competition with a complementary strand should play a major role in the formation and persistence of G4 structures *in vivo*, as recently pinpointed by Chalikian *et al.* (35). Two reports using single-molecule FRET or optical tweezers have shown that G4 can readily compete with the reannealing of dsDNA *in vitro* (44,45). However, the persistence time of the G4 structure embedded in duplex DNA remains largely unknown.

Here, we developed an original and label-free single-molecule setup using magnetic tweezers in order to gain insights into this aspect of G4 dynamics. This setup allows measuring formation and persistence of G4 structures in an

alternating context between ssDNA and dsDNA, by opening and closing periodically single molecule dsDNA hairpins. This label-free assay allowed us to measure the folding time ( $T_f$ ) of G4 structures as well as their persistence time ( $T_p$ ) under a defined force cycle. Using this assay, we compared four well-studied model G4 sequences from different genomic regions, namely human telomeric, human promoters cMYC and cKIT, and a replication origin G4 from the chicken genome. Our experiments show that these sequences form G4 structures as expected but show great variation in their folding and persistence times in ways that are not fully described while considering thermal melting parameters from bulk experiments.

Finally, this system also allowed us to assess the effect of a chemical G4 ligand (360B) (46) and an anti-G4 single-chain antibody (BG4) (47,48) on G4 dynamics. We observed that both molecules favor G4 structure formation by reducing the apparent folding time and increasing the persistence time of the G4. This observation confirms the basic but important idea that some G4s revealed experimentally by such G4 binders (specifically BG4 and 360B) may not form significantly in the absence of ligand.

## MATERIALS AND METHODS

### Oligonucleotides and G4 binders

The oligonucleotides used in this study are described in Supplementary Tables S1 and S2. All oligonucleotides were purchased from Eurogentec (Seraing, Belgium). Lyophilized oligonucleotides were resuspended in distilled water and stored at  $-20^{\circ}\text{C}$ . The G4 ligand 360B was synthesized in the laboratory by Patrick Mailliet. 360B is similar to the G4 ligand 360A, but the iodine counter ions in 360A are replaced by sulphonates (46,49) (Supplementary Figure S12A). This modification improves solubility of the compound in aqueous buffers, without affecting its specificity and affinity toward G4 (Patrick Mailliet and Jean-François Riou, personal communication). The BG4 single chain antibody was purified and provided by the laboratory of Prof. Kevin D. Raney (50).

### Experimental setup

The PicoTwist<sup>®</sup> magnetic tweezers instrument used to manipulate individual DNA hairpins tethered between magnetic beads and a coverslip was described previously (51). Briefly, the magnetic bead was held by a force due to a vertical magnetic field gradient generated by a pair of permanent magnets. Controlling the distance of the magnets from the sample surface allows applying a precisely calibrated variable force on the samples. The movement of these magnets was achieved with the help of DC-motor, which allowed us to exert a force with sub-picoNewton accuracy. The force was calculated from the Brownian fluctuations of the tethered bead (51). The DNA substrate used in the single-molecule studies consisted of a 1.1 kb hairpin containing the G4 forming sequence at ~640 nucleotides from the ssDNA–dsDNA junction of the molecule. The hairpin also contains a 5'-tri-biotinylated ssDNA tail and a 3' end with 43-nucleotide-long ssDNA tail (Supplementary Figure S1 and S2). We attached the hairpin substrate at the 5' end

to a streptavidin-coated magnetic bead and at the 3' end to a glass cover slip through annealing with ssDNA oligonucleotides covalently attached on the surface by click chemistry. Relevant oligonucleotides for hairpin synthesis are described in Supplementary Table S1. To image the beads, a CMOS camera at the image plane of the objective was used to track the position of the magnetic beads in three dimensions with nanometer resolution. Thus, tracking the  $z$ -axis fluctuations of the tethered beads enabled us to monitor the change in extension of the tethered DNA hairpin in real time with an accuracy of about 5 nm. The extension of the DNA changed in accordance with the applied force or by the presence of hybridized oligonucleotides or a G4 structure. Between 50 and 100 beads were tracked simultaneously in real time in order to obtain statistically significant measurements of single molecule events.

### Experimental conditions for single molecule G4 structure manipulation

All experiments were performed at 25°C in 10 mM Tris–HCl (pH 7.5), with 100 mM KCl or 100 mM LiCl, as indicated in the text and figure legends. Potassium ions (K<sup>+</sup>) strongly stabilize G4s by coordination with the carbonyl oxygens of guanines in the central cavity. To confirm that the detected structures were G4s, control experiments were systematically carried out in 100 mM LiCl. Lithium ions (Li<sup>+</sup>) are too small to fit in the G4 central channel and, hence, stabilizes G4 to a much lesser extent than potassium (52). The experimental system was based on manipulating multiple dsDNA hairpins tethered to a glass slide on one end and to magnetic beads on the other end. Time resolved experiments were started with injection of 10 nM of a 7-base-long oligonucleotide ('blocking oligonucleotide') complementary to the loop of the hairpin. In the case of fast folding and persistent G4 structures formed by ori  $\beta^A$ , the blocking oligonucleotide concentration was lowered to 5 nM. In experiments using G4 ligands, ligands and blocking oligonucleotides were injected concomitantly.

### Data collection and analysis

The image of the bead displayed diffraction rings that were used to estimate its 3D position as previously described (53). From position fluctuations of the bead, both the mean elongation of the molecule and the force applied to it could be deduced (54). The  $z$ -axis fluctuations were acquired at 16 Hz. A typical experimental acquisition spanned almost 24 h, corresponding to  $\sim 2000$  open/close cycles. Bead position at each force (2, 7 or 20 pN) were automatically recorded and analyzed using a software developed in house (Supplementary Figure S3–S5). The experimental mean folding time ( $\bar{T}_f$ ) of each G4 sequence was inferred by dividing the whole time spent in the unfolded state (the cumulative time of oligonucleotide blockage before a structure is observed), summed over all beads and over the entire acquisition, by the number of times a G4 structure was formed ( $\bar{T}_f = \Sigma$  oligo blocking time spent in unfolded state/total number of folded G4). This average is the best estimator for the parameter of a single exponential law if one considers G4 folding and unfolding as Poisson processes (see

Supplementary Figure S7 for explanations on calculations of the parameters and their independence on experiment time). The mean unfolding time, or persistence time ( $\bar{T}_p$ ), was similarly obtained by dividing the whole time spent in the folded state ( $\geq T_{\text{hold}} = 15$  s) by the number of times a G4 is unfolded ( $\bar{T}_p = \Sigma$  time spent in folded state ( $\geq 15$  s)/total number of unfolded G4).

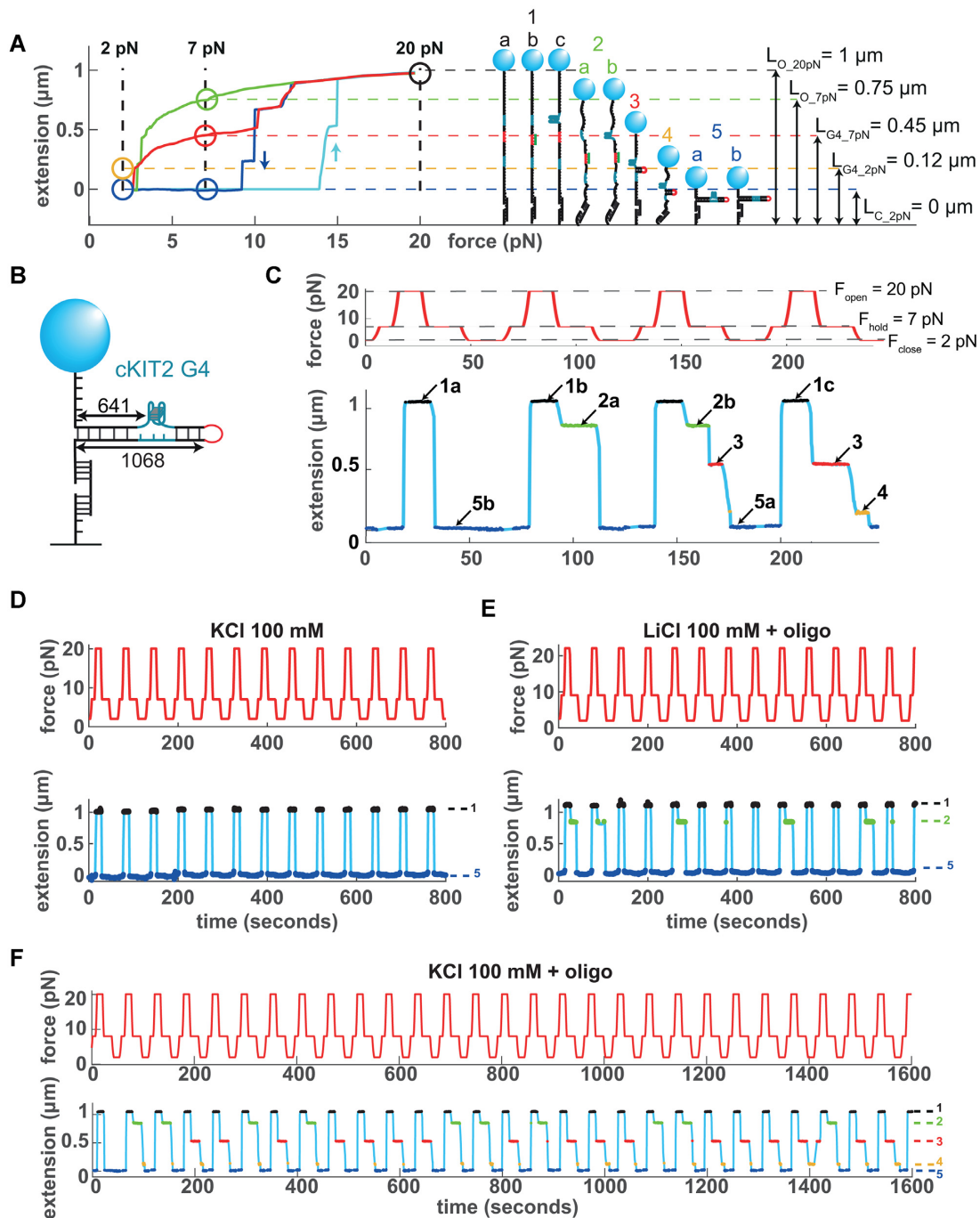
The relative error made on these estimations is  $\sqrt{N}$ , where  $N$  is the number of observed events (respectively unfolding and folding events). All results presented below were computed from at least 100 independent G4 events. When two successive cycles showed a blockage at the position of the G4, we assumed that it was due to the same structure and that no unfolding and refolding took place during one single hairpin opening phase. If the typical  $\bar{T}_f$  was much larger than  $T_c$  (cycle time), the probability that a G4 unfolded and then refolded during the opening phase was so low that this assumption does not bias the estimation of  $\bar{T}_f$  and  $\bar{T}_p$ . However, in the case where  $T_c/\bar{T}_f$  was not negligible, the  $\bar{T}_p$  of the G4 could be slightly overestimated as two successive structures could be mistaken for the same structure. The extent of the relative overestimation of  $\bar{T}_p$  is less than  $1 + (k_f + k_u)\alpha_i^2 T_c/2$ , where  $k_f = 1/\bar{T}_f$  ( $k_u = 1/\bar{T}_p$ ) is the real folding (unfolding) rate of the G4 structure and  $\alpha_i = 0.83$ . This correcting factor was considered in the estimation of the errors, which are for this reason asymmetric. In the same manner, the probability that a G4 structure was formed but could not be detected was considered in the errors relative to its folding time  $\bar{T}_f$ . Details methods and calculations used to determined error estimations of folding and persistence times is detailed in Supplementary Figure S8.

## RESULTS

### A single molecule assay to visualize G4 formation in real-time

To decipher G4 folding/unfolding dynamics at the molecular level, in a context where G4 structure exists in competition with a dsDNA structure, we developed a single molecule assay based on manipulation of DNA hairpin under a magnetic tweezer setup (54). Here, our molecules are  $\sim 1.1$  kb and contain one G4 forming sequence (or its non-G4 mutated version) (Figure 1A, Supplementary Figure S1, S2). In a magnetic tweezer experiment, the read-out is the molecule extension, which depends on the force applied to the bead and the DNA conformation. Assuming that the formation of the G4 structure requires a single stranded intermediate to form, the assay consisted of submitting the DNA hairpin to a predetermined force cycle so as to periodically open and close the molecule. Thus, the DNA molecules alternates between single-stranded and double stranded DNA configurations. The principle of our assay is presented in Figure 1 with a G4 forming sequence from the promoter of the human cKIT oncogene (cKIT-2, supplementary Table S2) (55).

Figure 1A shows the mean force–extension trajectories of the cKIT-2 hairpin at increasing and decreasing forces. At the beginning of the experiment, starting from the situation of a closed hairpin and no force applied, when we increase the force the hairpin remains closed until the force reaches 14 pN and then the hairpin quickly opens, with a maximum



**Figure 1.** Single molecule assay for detection of G4 formation. (A) Left panel: Force-extension curves of a 1068 bp DNA hairpin containing a cKIT-2 G4 forming sequence. Open circles mark the different molecule extensions observed at the three test forces. Cyan: hairpin opening at increasing forces. Blue: hairpin closing with no G4 or blocking oligonucleotide at decreasing forces. Green: hairpin closing with a blocking oligonucleotide. Red: hairpin closing with a folded G4.  $L_{O-20\text{pN}}$ : fully open hairpin at 20 pN.  $L_{O-7\text{pN}}$ : oligonucleotide-blocked hairpin at 7 pN.  $L_{G4-7\text{pN}}$ : G4-blocked hairpin at 7 pN.  $L_{G4-2\text{pN}}$ : G4-blocked hairpin at 2 pN.  $L_{C-2\text{pN}}$ : fully closed hairpin at 2 pN. Right panel: structures inferred from the force-extension curves. (B) Schematics of the hairpin molecule, indicating the size in bp of the full dsDNA hairpin and the position of the G4 sequence. (C) Force cycles (red) and extension traces (blue) illustrating the corresponding molecular states described in (A). (D) Force cycles (800 s, red) and corresponding extension traces (blue) of cKIT-2 hairpin in KCl only (color-coded numbers correspond to structures described in (A)). (E) Same as (D) in 100 mM LiCl plus blocking oligonucleotide. (F) Same as (D) in 100 mM KCl with blocking oligonucleotide.

extension of  $\sim 1 \mu\text{m}$  at 20 pN (structure 1a) corresponding to a fully unzipped hairpin (cyan curve). Conversely, starting from an extended hairpin at 20 pN and decreasing the force, the hairpin refolds when the force reaches 12 pN, and is completely re-annealed at 9 pN (blue curve). However, simply opening and closing the hairpin by alternating between ssDNA and dsDNA in KCl buffer did not allow to observe G4 formation (Figure 1D).

To favor formation of the G4 structure in the ssDNA state, we added a small 7-base long oligonucleotide complementary to the hairpin loop (blocking oligonucleotide, Supplementary Table S1). If during the opened phase at  $F_{\text{open}} = 20 \text{ pN}$ , the blocking oligonucleotide hybridized to the hairpin loop (structure 1b), it sequestered the loop, which is the hairpin nucleation spot (56). During oligonucleotide binding, the hairpin remains open until the force is reduced so much that the oligonucleotide gets ejected from the hairpin, allowing it to completely refold. In the trace shown in Figure 1A (green curve), this occurs when the force is lowered to 3 pN. The green curve therefore represents the elasticity of a 1068 bp hairpin in ssDNA state at forces from 20 to 3 pN.

Now, how do we detect the presence of a G4 structure in the hairpin? This situation is illustrated by the red curve in Figure 1A, which is observed for the cKIT-2 G4 only in the presence of KCl plus blocking oligonucleotide (Figure 1F), but not in 100 mM LiCl plus oligonucleotide condition, where G4 formation is disfavored (Figure 1E) (52). The trajectories for the red (containing a folded G4) and blue (without folded G4) curves show consistent patterns while the force is reduced from 20 to 10 pN (Figure 1A). However, in the case of the red curve, closing of the hairpin is blocked at  $\sim 0.5 \mu\text{m}$  extension at around 10 pN. The full closing of the hairpin is delayed until the force reaches 3 pN. We attribute this difference between the red and blue curves to the formation of G4, which prevents complete closing of the hairpin in the 10–3 pN force range.

To observe the different states of the molecule, we used the force–extension trajectories to choose empirically three specific forces (Figure 1C). At  $F_{\text{open}} = 20 \text{ pN}$ , the hairpin is fully open ( $\sim 1 \mu\text{m}$ ). Since hybridization with the blocking oligonucleotide and formation of a G4 do not significantly impact on the molecule extension at 20 pN, the  $1 \mu\text{m}$  extension corresponds to three indistinguishable structures: open hairpin without a bound oligonucleotide nor a folded G4 (structure 1a), open hairpin with a bound oligonucleotide (structure 1b), or open hairpin with a folded G4 (structure 1c). At  $F_{\text{hold}} = 7 \text{ pN}$ , we observe three different extension states of the molecule: (i) oligo blockage (structure 2a or 2b,  $\sim 0.75 \mu\text{m}$ , which corresponds to the full length molecule in ssDNA form containing or not a G4 structure), (ii) G4 blockage (structure 3,  $\sim 0.45 \mu\text{m}$ ) or (iii) no blockage (neither by a G4 nor by an oligonucleotide) (structure 5,  $0 \mu\text{m}$ ). Consistent with the blockage at  $0.45 \mu\text{m}$  corresponding to a G4 structure, the ratio between the two extensions at 7 pN ( $0.45/0.75$ ) represents 60% of the total extension, which for this cKIT-2 hairpin of 1068 bp long corresponds exactly to the distance between the G4 sequence and the ssDNA–dsDNA junction ( $0.6 \times 1068 \text{ bp} = 641 \text{ bp}$ ) (Figure 1B). At  $F_{\text{close}} = 2 \text{ pN}$ , the hairpin is fully closed (structure 5a or 5b,  $0 \mu\text{m}$ , which corresponds to a dsDNA hairpin containing

or not a folded G4). In addition, only in the cases where a G4 structure is formed, a  $0.12 \mu\text{m}$  extension state can be observed at 2 pN, which can be attributed to a relaxed form of structure 3 (Figure 1A, C, F, structure 4).

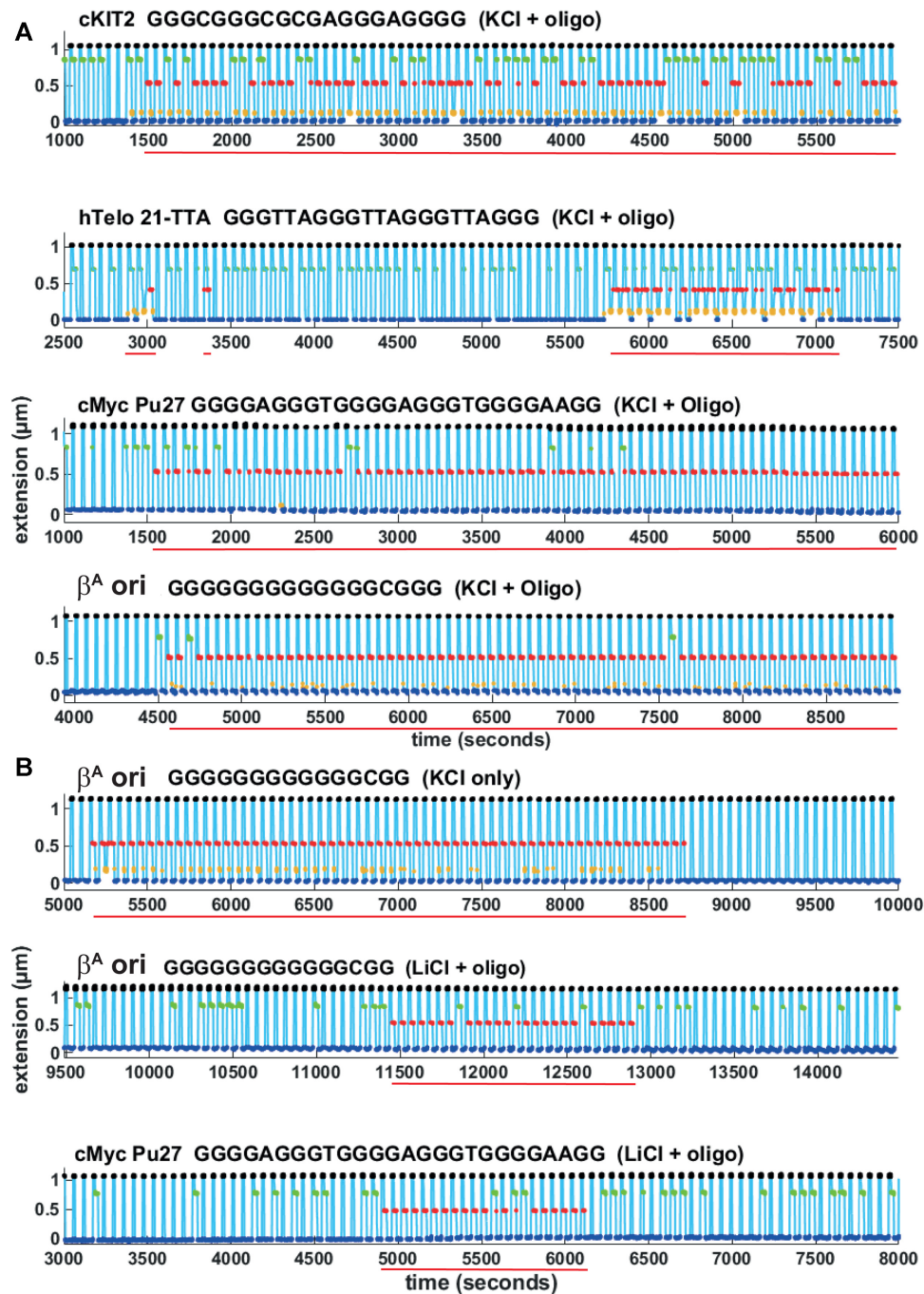
Practically, the assay consisted on imposing the three forces successively on the molecule during multiple cycles (Figure 1C–F). Figure 1C illustrates the extensions corresponding to different molecular states detected during force cycles. Evidence of the persistence of the G4 throughout multiple cycles is then provided by successive blockage at  $0.45 \mu\text{m}$ . These data would suggest that the G4 gets embedded in a dsDNA bubble during the phase at 2 pN (structure 5a) and resists to the pulling force at 20 pN (structure 1c). Different scenarios to determine the persistence of a G4 from extension traces is described in Supplementary Figure S9.

### Folding of G4 structures from human telomeres, human oncogene promoters and an avian replication origin

We used this assay to compare the behavior of several well studied G4 forming sequences in our assay in addition to cKIT-2, namely human telomeric repeats (hTelo 21-TTA) (57), the promoter of the human oncogene cMYC (cMYC-Pu27) (58), and a replication origin found in the chicken genome (ori  $\beta^A$ ) (10). We performed experiments lasting up to 16 hours representing several thousands force cycles. Figure 2A shows 5000 s long snapshots of extension traces obtained with the four sequences, in presence of 100 mM KCl and blocking oligonucleotide. Formation of G4 structures could be observed for all sequences on most tested beads, demonstrating the generality of the assay to study formation of intramolecular G4s. Recordings and automatic detection of molecule states from the full course of representative experiments can be found in Supplementary Figure S3–S5. Experiments were also performed in control conditions (100 mM LiCl with or without blocking oligonucleotide, or 100 mM KCl without blocking oligonucleotide). No blockage corresponding to G4 structures was observed in either control conditions for cKIT-2 and hTelo 21-TTA. We could, however, detect formation of G4 structures, while less frequent and short-lived, in 100 mM LiCl plus blocking oligonucleotide for cMYC-Pu27 and ori  $\beta^A$  sequences (Figure 2B and Supplementary Table S3). We measured circular dichroism spectra of oligonucleotides containing the cMYC-Pu27 and ori  $\beta^A$  sequences in LiCl buffer, which exhibited the characteristic profile of parallel G4 structures (Supplementary Figure S6). The ori  $\beta^A$  sequence also exhibited G4 formation in 100 mM KCl without blocking oligonucleotide (Figure 2B and Supplementary Figure S5). Therefore, our assay can be used to characterize unambiguously rare G4 folding events.

### Comparative folding and unfolding dynamics of intramolecular G4s

Next, we measured the folding and persistence times of the G4 structures in our assay. To gain insights into the folding and unfolding dynamics, we recorded multiple events of G4 formation over several beads for each G4 forming sequence (recording more than a hundred G4 events per sequence, each experiment lasting up to 16 h). We were able



**Figure 2.** Extension traces from single molecules containing cKIT-2, hTelo 21-TTA, cMYC-Pu27 or ori  $\beta^A$  G4 forming sequences. (A) From top to bottom, cKIT-2, hTelo 21-TTA, cMYC and ori  $\beta^A$  G4 forming sequences in 100 mM KCl and blocking oligonucleotide. (B) ori  $\beta^A$  G4 forming sequence in KCl without blocking oligonucleotide (top) and in LiCl plus blocking oligonucleotide (middle). cMYC-Pu27 in LiCl plus oligonucleotide (bottom). Formation of a G4 structure is detected by a blockage at 7 pN (red), as described in Figure 1. Single G4 events are underlined (red).

to monitor a field of multiple molecules at the same time to obtain statistically significant measurements of the G4 folding and persistence times under various conditions. For each sequence studied, we compiled the experimental folding times ( $T_f$ ) as all the times where the blocking oligonucleotide delayed hairpin refolding at  $F_{\text{hold}}$  (extension of 0.75  $\mu\text{m}$  at 7 pN) before observing G4 folding (extension of 0.45  $\mu\text{m}$  at 7 pN). Similarly, we defined persistence times ( $T_p$ ) as

the time spent by a molecule in a G4 conformation before the structure unfolded. The average folding time ( $\overline{T_f}$ ) of a given sequence was therefore obtained by dividing the sum of  $T_f$  over all beads and over the entire acquisition, by the number of times distinct G4 events were observed. Thus, the inverse of  $\overline{T_f}$  ( $k_f = 1/\overline{T_f}$ ) is a direct measure of the probability of folding a G4 structure per unit of time spent at 7 pN. A longer average folding time indicates a smaller fold-

ing probability while a shorter average folding time indicates a larger folding probability. Briefly, the average persistence time ( $\overline{T}_p$ ) is acquired by dividing the sum of  $T_p$  over all beads and over the entire acquisition, by the number of times a G4 is unfolded (corresponding to the first absence of blockage at both 0.75 and 0.45  $\mu\text{m}$  extension following detection of a G4 structure). Therefore  $\overline{T}_p$  represents the stability of a given G4 structure in a context alternating between ssDNA and dsDNA. The inverse of  $\overline{T}_p$  ( $k_u = 1/\overline{T}_p$ ) indicates the unfolding rate of the G4 structure. Details on  $\overline{T}_f$  and  $\overline{T}_p$  calculation is given Supplementary Figure S7.

Our assay involves a periodic testing of the presence of a G4 in the substrate at every cycle. During a fraction of this cycle, the presence of the G4 cannot be detected, either because the hairpin is open, because the G4 is embedded in double-stranded DNA or the hairpin is blocked by an oligonucleotide. Thus, there is a probability that a G4 structure unfolds and then refolds during this time. We could then in theory observe two successive blockages at the G4 position that we would consider as a unique structure while in reality, two successive folding and unfolding events happened. This probability depends on the dynamic of the structure and can be calculated. Details on calculation of this probability are given in Supplementary Figure S8. This probability is low for most of G4 studied here except for ori  $\beta^A$ . However, this probability was taken into account in the error estimation of  $\overline{T}_f$  and  $\overline{T}_p$ .

The dynamic properties of all G4 structures tested in this work are shown in Table 1 and Figure 3A. The four studied G4 sequences (hTelo 21-TTA, cKIT-2, cMYC-Pu27 and ori  $\beta^A$ ) displayed large difference on their  $\overline{T}_f$  (from  $\sim 30$  to  $\sim 12\,000$  s) and  $\overline{T}_p$  (from  $\sim 500$  to  $\sim 11\,000$  s) at 25°C in 100 mM KCl plus blocking oligonucleotide. Some G4 sequences, such as cMYC-Pu27 and ori  $\beta^A$ , had a very short folding time and very long persistence time. In comparison, cKIT-2 exhibited a longer folding time and a shorter persistence time compared to cMYC-Pu27 and ori  $\beta^A$ . Finally, hTelo 21-TTA presented the most dynamic structures, with relatively short folding and short persistence times.

The distribution of folding ( $T_f$ ) and persistence ( $T_p$ ) times followed in most of the cases an exponential law, as exemplified for the cKIT-2 G4 in Figure 3C and D. This suggests folding and unfolding of the G4 in a single step. However the persistence time of the hTelo 21-TTA G4 is best fitted by a power law (Figure 3E), which would indicate a mixture of indistinguishable conformations with different ( $\overline{T}_p$ ), corroborating previous results pointing to structural diversity of human telomeric G4 (43).

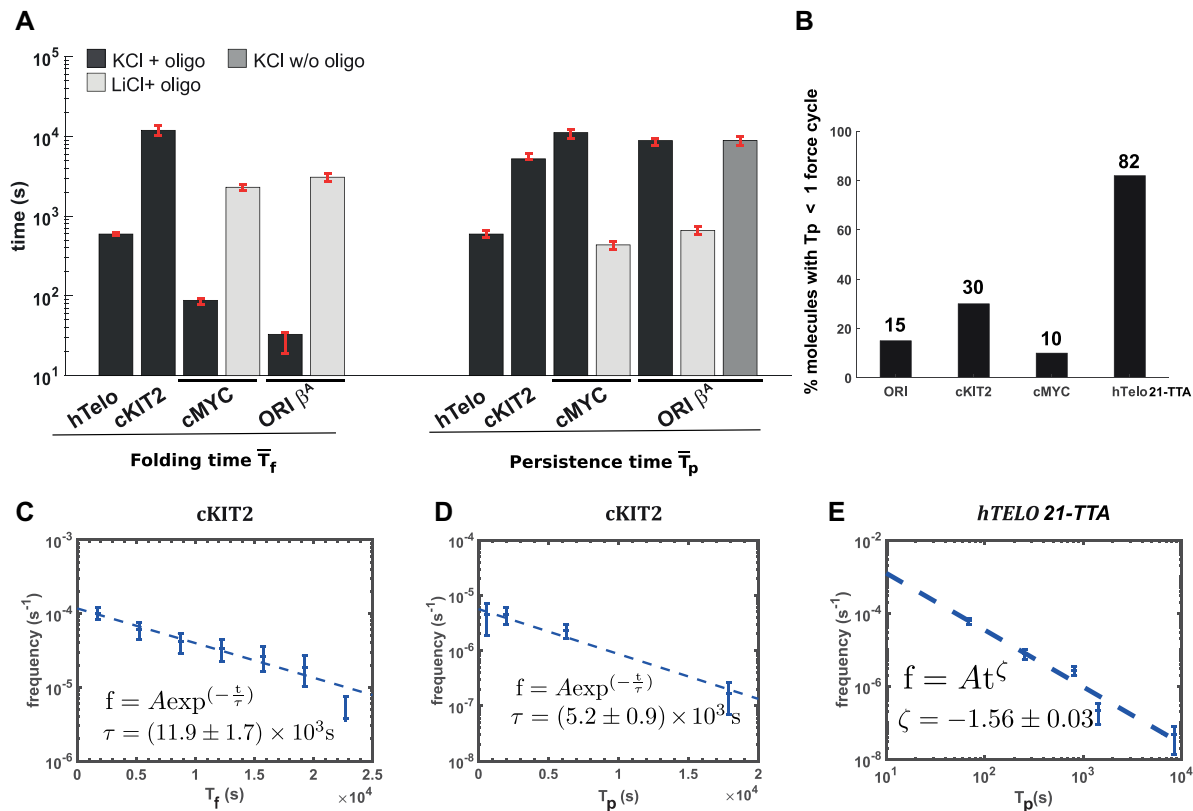
By design, one force cycle corresponds to the lower-resolution limit of our assay. Therefore, structures that were stable less than one cycle could represent unstable structures either with a short persistence time or structures that are mechanically unfolded during the pulling step at 20 pN. These ambiguous events were thus not considered in the calculation of  $\overline{T}_p$ . Such unstable structures were frequently observed for the telomeric G4 (21-TTA), with 82% of the structures observed exhibiting a persistence time shorter than one cycle. For cMYC-Pu27, cKIT-2 and ori  $\beta^A$  G4s, the structures lasted less than one cycle represented respectively 10%, 30% and 15% of the events observed (Figure

3B). Interestingly, as noted above, we could also observe for the ori  $\beta^A$  and the cMYC-Pu27 G4s a blockage at 0.45  $\mu\text{m}$  extension in LiCl in presence of a blocking oligonucleotide (Figure 3 and Supplementary Figure S4 and S5). However, the average folding time of the structures was longer ( $\sim 26$ -fold increase for cMYC-Pu27 and 100-fold increase for ori  $\beta^A$ ), and their average persistence time was significantly shorter (respectively 27-fold and 14-fold for cMYC-Pu27 and ori  $\beta^A$ ) than the ones obtained in KCl plus blocking oligonucleotide. The lower persistence of the structures observed in LiCl compared to KCl confirms that the blockage at 0.45  $\mu\text{m}$  is due to G4 structures. We cannot completely rule out the contribution of a structure forming on the opposite C-rich strand, namely an i-motif, although the neutral pH buffer conditions used in our assay should disfavor its formation (58). In the absence of the blocking oligonucleotide and in 100 mM KCl, ori  $\beta^A$  could form a structure with similar persistence (9000 s) than the one in KCl plus oligonucleotide. This result suggests two additional possibilities: the ori  $\beta^A$  sequence might not require the opening of dsDNA to form a G4 or it may fold in a ssDNA context even at high force (20 pN).

### Folding time and persistence time of ori $\beta^A$ mutants

Sequences composed of long guanine stretches (such as the ori  $\beta^A$  sequence) form preferentially intermolecular G4s in bulk experiments, preventing from assessing the effect of intramolecular G4 folding. Our single-molecule assay helps to overcome this issue. In order to test the sensitivity of the ori  $\beta^A$  sequence to mutation in our assay, we measured the dynamics of mutated sequences of ori  $\beta^A$  containing an adenine instead of a guanine at position 12 (m12) or at position 16 (m16) (Table 1 and Supplementary Figure S10C). These point mutations were previously reported to lower the ability of these sequences to form G4 structures based on prediction from sequence features and were shown to correlate with lower replication origin activity *in vivo* (10). In our assay, the two ori  $\beta^A$  mutants could still fold into a G4 structure, although they exhibited a much longer folding time compared to wild-type (wt) ori  $\beta^A$  (46-fold and 300-fold increase for m12 and m16 respectively). The m12 G4 structure exhibited a long persistence time, similar to the wt ori  $\beta^A$  structure, while m16 G4 had a much shorter  $\overline{T}_p$  (72-fold) compared to wt and to m12. In addition, m16 presented  $\sim 4.5$ -fold increase in the percentage of G4 structures with very short persistence time ( $\leq 1$  force cycle) compared to wt and to m12 (69% for m16 instead of 15% for wt and m12) (Supplementary Figure S10B). These results show that our assay can reveal intramolecular G4 dynamics and the effect of subtle sequence difference in G4 forming sequences made of long guanine stretches. In the case of the ori  $\beta^A$  mutants, we see that the guanine at position 12 only affects the folding rate of the G4, whereas the guanine at position 16 is critical for the efficient folding and persistence of the G4 structure in the dsDNA context. These results are consistent with G4 folding prediction made from the ori  $\beta^A$  sequence. The wt ori  $\beta^A$  sequence has the possibility to form at least ten different three-quartet G4 structures. However, the m12 mutant conserves only one possibility out of ten to form a three-quartet G4, which could explain its lower





**Figure 3.** Folding time and persistence time of G4 structures. (A) Average folding times and persistence times of G4 structures. (B) Percentage of G4 structures with a persistence time shorter than one force cycle. Frequency of folding times (C) and persistence times (D) of cKIT-2 G4. Both are fitted with a first order exponential law. (E). Frequency of persistence time of hTelo 21-TTA G4 and power law fit of the data.

**Table 1.** Average folding and persistence times of G4 structures in different conditions

Conditions	cMYC	cKIT-2	ori $\beta^A$ wt	ori $\beta^A$ m12	ori $\beta^A$ m16	21-TTA	45-TTA	21-CTA
$\bar{T}_f$ (s)								
LiCl	nd	nd	–	nd	nd	nd	nd	nd
LiCl + oligo	2300 $\pm$ 200	–	3000 $\pm$ 350	–	–	–	–	–
KCl	–	–	nd	–	–	–	–	–
KCl + oligo	90 $\pm$ 10	12 000 $\pm$ 1700	30 $\pm$ 13	1400 $\pm$ 70	9000 $\pm$ 1300	350 $\pm$ 30	150 $\pm$ 13	1700 $\pm$ 130
LiCl	nd	nd	–	nd	nd	nd	nd	nd
$\bar{T}_p$ (s)								
LiCl + oligo	400 $\pm$ 50	–	650 $\pm$ 8	–	–	–	–	–
KCl	–	–	9000 $\pm$ 1200	–	–	–	–	–
KCl + oligo	11 000 $\pm$ 1700	5000 $\pm$ 900	9000 $\pm$ 1300	6500 $\pm$ 1300	125 $\pm$ 20	600 $\pm$ 80	500 $\pm$ 50	350 $\pm$ 60

nd: not determined. –: no G4 formation observed. means  $\pm$  estimation error are indicated.

folding rate compared to the wt sequence. The m16 mutant can form two-quartet G4s or a three-quartet G4 with zero nucleotide-loops, which affects both folding and unfolding rates.

### Folding time and persistence time of dimeric human telomeric repeats and of the common CTAGGG variant

Human telomeres contain variant repeats interspersed with the consensus TTAGGG motif. The CTAGGG variant, when present as a contiguous array within the telomere, causes telomere instability in the male germ line and somatic cell (59). It has been shown that four successive CTAGGG repeats were prone to form a G4 *in vitro*, and adopted a different G4 conformation than the wt (TTAGGG)<sub>4</sub> repeats in KCl (60). In our assay, the forma-

tion of the 21-CTA variant in KCl was five times slower than the wt 21-TTA G4 ( $\bar{T}_f = 1700$  s versus 350 s). The persistence time of the variant 21-CTA G4 was also reduced by  $\sim 2$ -fold compared to the wild type ( $\bar{T}_p = 350$  s versus 600 s) (Supplementary Figure S11A, B and Table 1). Therefore, the variant cytosine affects the folding ability of the G4 and has a modest effect on its persistence time. Altogether, these results indicate that the CTAGGG variant has a slightly lower ability to fold into a G4 and to persist in DNA than the wt telomere repeat, thus suggesting that the increase of genomic instability at human telomeres due to this variant cannot be accounted by G4 formation on its own.

Since telomeres are composed of tandem repeats *in vivo*, we also measured the folding and unfolding dynamics of one versus two G4 units (45-TTA) (Supplementary Figure S11A, B and Table 1). Folding of this sequence into a G4

**Table 2.**  $\overline{T}_f$  and  $\overline{T}_p$  of hTelo 21-CTA at different concentration of ligand 360B or BG4 antibody

[360B]	100 nM	500 nM	1 $\mu$ M	1.5 $\mu$ M	2 $\mu$ M
$\overline{T}_f$ (s)	1350 $\pm$ 110	750 $\pm$ 110	300 $\pm$ 35	130 $\pm$ 15	70 $\pm$ 10
$\overline{T}_p$ (s)	300 $\pm$ 30	450 $\pm$ 60	350 $\pm$ 50	1050 $\pm$ 250	1350 $\pm$ 300
[BG4]	20 nM	100 nM	200 nM		
$\overline{T}_f$ (s)	1050 $\pm$ 130	400 $\pm$ 30	150 $\pm$ 15		
$\overline{T}_p$ (s)	300 $\pm$ 70	1150 $\pm$ 150	1250 $\pm$ 250		

Means  $\pm$  estimation error are indicated.

took in average half the time ( $\overline{T}_f = 150$  s) compared to the 21-TTA sequence in KCl. This difference in mean folding time between one and two G4 units is compatible with independent folding of the two G4 structures, suggesting no or little interaction of the tandem G4s in our system. The 45-TTA G4 displayed a persistence time ( $\overline{T}_p = 500$  s) similar to the one of 21-TTA G4 ( $\overline{T}_p = 600$  s), indicating again a lack of a cooperative effect on the persistence time of the structures. The decrease in folding times suggests a higher probability to form the G4 structure in the case of the repeated sequence 45-TTA. This is compatible with the larger number of G4 nucleation sites permitted by the repeat of G4 motifs (Supplementary Figure S11C). Among these five nucleation sites, three of them (position 2/3/4) prevent the formation of two consecutive G4 structures. Thus, the probability that the 45-TTA sequence folds into a unique G4 structure is higher than the probability to form into two consecutive ones. However, our assay does not allow distinguishing effectively between the formation of one or two structures in this case.

### G4 stabilizing ligands affect both G4 folding and unfolding dynamics

Finally, we wanted to assess the effect of G4 ligands on the kinetics of G4 folding and persistence in our assay. The potential to manipulate G4 formation at telomeres or oncogene promoters have made them an attractive target for antitumoral drug design (61–64). Small molecules and antibodies have been developed for targeting G4 *in vitro* and have been widely used to evidence G4 formation *in vivo* and potentialize their biological effect (47,65,66). We tested how such molecules affect folding and unfolding dynamics of G4 structures in the context of our single molecule assay, using the 21-CTA telomere variant as a model sequence (Table 2). As demonstrated above, this sequence has the benefit of exhibiting a long folding time ( $\overline{T}_f = 1700$  s) and a short persistence time ( $\overline{T}_p = 350$  s), therefore allowing any effect of the ligand on these two values to be easily measurable.

We investigated the effect of 360B, a selective G4 ligand of the pyridine derivative series, on both folding and unfolding dynamics of the CTA variant G4. The pyridine dicarboxamide derivative 360B is a more water-soluble formulation of the commonly used ligand 360A (Supplementary Figure S12A) (46,67). We measured the folding and persistence times of the G4 in 100 mM KCl and blocking oligonucleotides with increasing concentrations of 360B (Figure 4). First, we observed that the percentage of structures lasting less than one open/close cycle decreases from

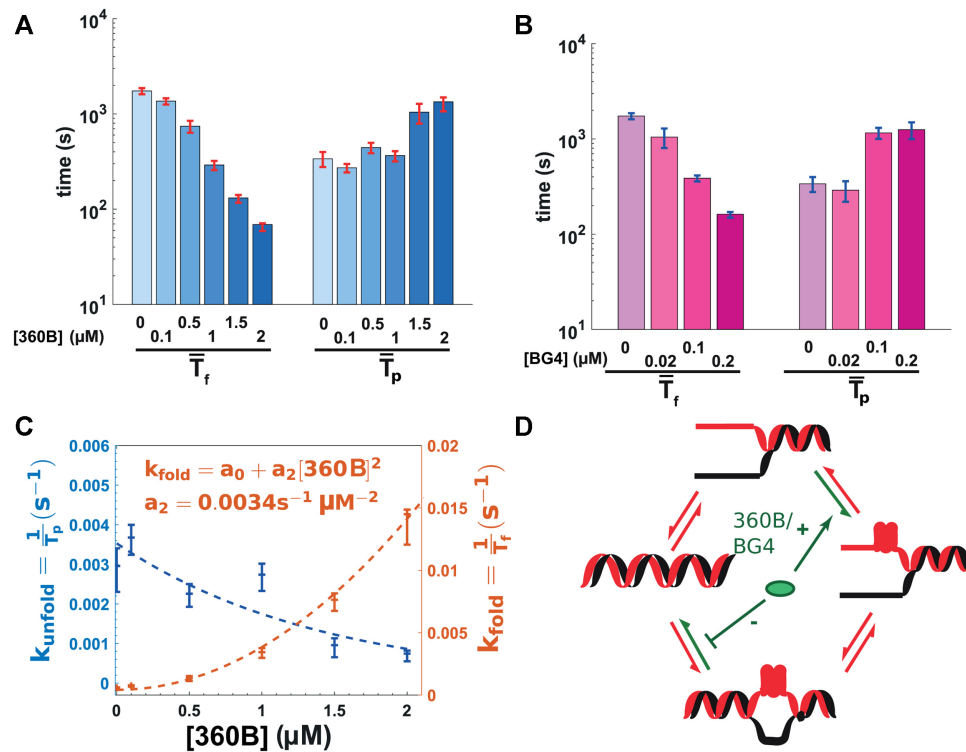
88% to 45% when the ligand concentration increases from 0 to 2  $\mu$ M, confirming the stabilization of the structures (Supplementary Figure S12B). Second, increasing concentration of 360B affected both the folding time and the persistence time of the structure in a dose-dependent manner. With increasing concentrations of 360B, we observed higher G4 folding rate (decreased  $\overline{T}_f$ ) and more persistent G4s (increased  $\overline{T}_p$ ), suggesting that the ligand favors the formation of the structure and increases its persistence time (Figure 4A). Interestingly,  $k_{\text{fold}} = 1/\overline{T}_f$  increased non-linearly with the concentration of ligand (Figure 4C). The data fit to a quadratic polynomial equation ( $R^2 = 0.9957$ ), which would suggest that two ligand molecules bind simultaneously on one G4 structure. This is reminiscent of what has been reported previously with the interaction between human telomeric G4 and the ligand CuII-tolyterpyridine (68). The ligand also affected the persistence time of the structure. Indeed,  $k_u = 1/\overline{T}_p$  decreased with increasing concentration of 360B, with the data best fitted by a first order exponential decay. We also tested the effect of 360B on the cKIT-2 G4, a structure which exhibits long  $\overline{T}_f$  and long  $\overline{T}_p$ . Similarly to the 21-CTA G4, we observed a higher folding rate and a longer persistence time of the cKIT-2 G4 in the presence of ligand (Supplementary Figure S13). 100 nM of 360B was sufficient to reduce  $\overline{T}_f$  by 4-fold and increase  $\overline{T}_p$  by almost 20-fold. Because of its already long persistence time, most of the cKIT-2 G4 ( $\sim 60\%$ ) were not unfolded during the recorded time of the experiment ( $\geq 90\,000$  s) at a concentration of 100 nM of 360B.

Finally, we tested the effect of the anti-G4 single chain BG4 antibody (47) on the 21-CTA sequence. Similarly to our results with 360B, BG4 elicited a dose-dependent effect on both  $T_f$  and  $T_p$  (Figure 4B and Supplementary Figure S12). Therefore, the two ligands behave in our assay by favoring G4 formation in ssDNA and slowing down its unfolding, as schematized in Figure 4D.

## DISCUSSION

In this work, using a single molecule DNA manipulation assay, we compared the dynamics of G4 folding and unfolding for several G4 forming sequences in a context where ssDNA exists in equilibrium with dsDNA. Here, we studied well-known G4 motifs, including human telomeres, human oncogene promoters of cMYC and cKIT genes, and a chicken replication origin.

Overall, our results are in agreement with previously determined parameters from bulk studies (Table 3). However, our assay allows accessing folding and persistent rates, that are likely to be more reflective of their dynamic *in vivo*. Indeed, determining both folding and unfolding rates leads to interesting observations even for stable G4s. For example, although the cKIT-2 G4 exhibits a high melting temperature ( $T_m = 70^\circ\text{C}$ ), our assay reveals that it has a low folding rate. In comparison, the telomeric G4 hTelo 21-TTA exhibits a similar  $T_m$  of  $68^\circ\text{C}$  but shows much faster dynamics. Our textbook view of G4 formation in genomic DNA is that the folding of such structure needs a transient ssDNA intermediate, which could likely occur during replication, transcription or DNA repair (71). Indeed, to form in our assay,



**Figure 4.** Effect of G4 ligand 360B and BG4 antibody on G4 folding and persistence times. (A) Folding and persistence times of 21-CTA telomeric G4 with increasing concentration of the small ligand 360B. (B) Folding and persistence times of 21-CTA G4 with increasing concentration of BG4 antibody. (C) Folding and unfolding rates of the 21-CTA G4. Fitting equations of the experimental data are displayed. (D) Schematic diagram showing the effect of the persistence of G4. G4 ligand and antibody (green ellipse) may favor the folding as well as the persistence of G4.

most structures required a step in a single strand conformation at low force. Interestingly, in the case of the ori  $\beta^A$  sequence, G4 can form in absence of this step at low force. This suggest that ori  $\beta^A$  can form a G4 in ssDNA at high force, or that it can form spontaneously in dsDNA. In addition, this is the first time that an intramolecular G4 formed by a long guanine stretches like the ori  $\beta^A$  sequence was experimentally observed, since they tend to fold into intermolecular G4 structures (or a mixture of intra- and intermolecular G4s) at the DNA concentrations required for common *in vitro* spectroscopic studies (10). Finally, although our label-free assay is a simplistic model with regard to an actual *in vivo* situation, the observed dynamics of the G4 structures are consistent with their purported role *in vivo*. In the case of a replication origin G4, fast folding and long persistence time in a dsDNA should be advantageous if the structures serve as anchors for protein complexes prior to replication licensing as previously proposed (38). Although not forming as easily as the ori  $\beta^A$  G4, the two promoter G4s, cMYC and cKIT, also displayed long persistence in dsDNA, which is consistent with a sustained effect on transcription activity (72).

At the other end of the spectrum, human telomeric sequences were the weakest of the tested structures. They were rapid to form, but showed limited mechanical stability under a 20 pN force and exhibited relatively short persistence time, compared to other sequences. The short persistence (less than one force cycle) observed with human telomeric

**Table 3.** Comparison of  $\overline{T}_p$  and  $\overline{T}_f$  with melting temperatures ( $T_m$ ) and unfolding forces determined in identical buffer conditions (10 mM Tris-HCl pH 7.5, 100 mM KCl), ranked by decreasing  $T_m$

G4	$\overline{T}_p \pm \text{error}$	$\overline{T}_f \pm \text{error}$	$T_p \leq T_c$	$T_m$ (°C)	Unfolding force (pN)
$\beta^A$ ori	9000 $\pm$ 1300	30 $\pm$ 13	15%	80 (10)	nd
cMYC	11 000 $\pm$ 1700	90 $\pm$ 10	10%	78 (69)	22, 50 (42)
cKIT-2	5000 $\pm$ 900	12 000 $\pm$ 1700	30%	70 (70)	nd
21-TTA	600 $\pm$ 80	350 $\pm$ 30	82%	68 (68)	11, 22 29, 36 (43)
21-CTA	350 $\pm$ 60	1700 $\pm$ 130	88%	67 (70)	nd
45-TTA	500 $\pm$ 5	150 $\pm$ 13	83%	59 (70)	nd
$\beta^A$ -m12	6500 $\pm$ 1300	1400 $\pm$ 170	15%	55 (10)	nd
$\beta^A$ -m16	125 $\pm$ 20	9000 $\pm$ 1300	69%	45 (10)	nd

nd: not determined.  $T_c$ : cycle time. references in parentheses.

G4s (Supplementary Figure S11) can be attributed either to a low mechanical stability at high force (20 pN) or to a low persistence within a dsDNA bubble. Although we cannot distinguish between these two possibilities, previous single-molecule assays have shown that the human telomeric G4 repeat could fold into four different G4 conformations, two of them being unstable around 20 pN (Table 3). The distribution of persistence times measured with our assay was best fitted by a power law, also suggesting the contribution of several structures. Our results corroborate other former single molecule studies (43) and point out

that a single human telomeric G4 is a dynamic structure, forming promptly in ssDNA but also dissolved rapidly under mechanical forces or in the presence of a complementary strand. In the human genome, the telomeric sequence TTAGGG is mainly embedded in a large tandemly repeated dsDNA regions which may form multiple consecutive G4s. This raises the possibility that the impediment to molecular motors when progressing through telomeric repeats owes little to the stability of individual G4 structures alone, at least in the dsDNA portion of telomeres, but probably to other factors, like DNA-bound proteins.

Our results using G4 stabilizing ligands showed that the 360B ligand and the BG4 antibody significantly increased the formation rate of the G4s, as well as their persistence time, proving that they favor G4 formation in addition to stabilizing the structure, as previously noted (66,73). In this regard, our assay could help characterize ligands or antibodies which perturb less (or more) folding rates and would be a useful addition for *in vivo* physiological studies of G4 dynamics.

## CONCLUSION

Contrarily to previous single-molecule measurements that were performed at constant forces and without the presence of a complementary strand (39–45), our assay allows *in vitro* measurement of G4 folding and persistence times in a context that simulates the competitive environment of G4 structures formation *in vivo*. This assay could be easily adapted to study various biochemical processes (in first place replication and transcription), by tuning the parameters of the force cycles and by studying the effect of proteins involved in G4 processing. It is noteworthy that such modifications in the force cycles significantly affect the kinetics of formation but not the persistence time of a G4 structure (Supplementary Figure S14). Therefore, our assay allows assessing a wide range of G4 folding dynamics while the DNA molecule is subjected to different mechanical stresses or G4-interacting ligands. Finally, gathering folding rates and persistence times of a wider array of sequences in this assay could help improve the predictive power of computational methods developed to find G4 forming sequence in genomic data and score their ability to form a stable structure in a dsDNA context (74).

## DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding authors.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank Patrick Mailliet for providing the 360B ligand and Kevin D. Raney for providing us with the BG4 antibody. The authors thank Bertrand Ducos and Jessica Valle Orero for discussions, Jean-François Riou, Jean-Louis Mergny, Nicolas Desprat and the anonymous reviewers for fruitful comments on the manuscript.

## FUNDING

National Research Agency [MuSeq, ANR-15-CE12-0015, G4-crash - 19-CE11-0021-01]; CNRS; INSERM; Museum National d'Histoire Naturelle to the Genome Structure and Instability unit; P.L.T.T. was supported by a Marie Skłodowska Curie individual fellowship (2019–2021); work in the group of V.C. is part of 'Institut Pierre-Gilles de Gennes' ['Investissements d'Avenir' program ANR-10-IDEX-0001-02 PSL, ANR-10-LABX-31]; Qlife Institute of Convergence (PSL Univesité). Funding for open access charge: ANR.

*Conflict of interest statement.* V.C. is cofounder of PicoTwist. V.C and J.F.A. are academic founders and shareholders in Depixus SAS.

## REFERENCES

- Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
- Lightfoot, H.L., Hagen, T., Tatum, N.J. and Hall, J. (2019) The diverse structural landscape of quadruplexes. *FEBS Lett.*, **593**, 2083–2102.
- Spiegel, J., Adhikari, S. and Balasubramanian, S. (2020) The structure and function of DNA G-quadruplexes. *Trends Chem.*, **2**, 123–136.
- Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
- Rawal, P., Kumarasetti, V.B.R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K. and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
- Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
- 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. Biotech.*, **33**, 877–881.
- Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Di Antonio, M. and Balasubramanian, S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.
- Fouk, M.S., Urban, J.M., Casella, C. and Gerbi, S.A. (2015) Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res.*, **125**, 725–735.
- Valton, A.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., Riou, J.F. and Prioleau, M.N. (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.*, **33**, 732–746.
- Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., Guédin, A., Mergny, J.L., Damaschke, J., Schepers, A. *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. *Nature Commun.*, **10**, 3274.
- Brooks, T.A., Kendrick, S. and Hurley, L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.
- Wang, W., Hu, S., Gu, Y., Yan, Y., Stovall, D.B., Li, D. and Sui, G. (2020) Human MYC G-quadruplex: From discovery to a cancer therapeutic target. *Biochim. Biophys. Acta Rev. Cancer*, **1874**, 188410.
- Cogoi, S., Rapozzi, V., Cauci, S. and Xodo, L.E. (2017) Critical role of hnRNP A1 in activating KRAS transcription in pancreatic cancer cells: a molecular mechanism involving G4 DNA. *Biochim. Biophys. Acta Gen. Sub.*, **1861**, 1389–1398.
- Gomez, D., Guédin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P. and Calsou, P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.

16. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5'UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218.
17. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
18. Jimeno, S., Camarillo, R., Mejias-Navarro, F., Fernandez-Avila, M.J., Soria-Bretones, I., Prados-Carvajal, R. and Huertas, P. (2018) The helicase PIF1 facilitates resection over sequences prone to forming G4 structures. *Cell Rep.*, **24**, 3262–3273.
19. Bryan, T.M. (2019) Mechanisms of DNA replication and repair: insights from the study of G-quadruplexes. *Molecules*, **24**, 3439.
20. Bryan, T.M. (2020) G-quadruplexes at telomeres: friend or foe? *Molecules*, **25**, 3686.
21. Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M.-N., Balasubramanian, S. and Sale, J.E. (2014) Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J.*, **33**, 2507–2520.
22. Mao, S.-Q., Ghanbarian, A.T., Spiegel, J., Martinez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D. and Balasubramanian, S. (2018) DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.*, **25**, 951–957.
23. Mukherjee, A.K., Sharma, S. and Chowdhury, S. (2019) Non-duplex G-quadruplex structures emerge as mediators of epigenetic modifications. *Trends Genet.*, **35**, 129–144.
24. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.
25. Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V.A., Mergny, J.-L. and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.*, **5**, e1000475.
26. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R.M.J. and Mergny, J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.
27. Lerner, L.K. and Sale, J.E. (2019) Replication of G quadruplex DNA. *Genes*, **10**, 95.
28. Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C. and Zakian, V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
29. Tarsounas, M. and Tijsterman, M. (2013) Genomes and G-quadruplexes: for better or for worse. *J. Mol. Biol.*, **425**, 4782–4789.
30. Giraldo, R. and Rhodes, D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.*, **13**, 2411–2420.
31. Lopez, C.R., Singh, S., Hambarde, S., Griffin, W.C., Gao, J., Chib, S., Yu, Y., Ira, G., Raney, K.D. and Kim, N. (2017) Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA. *Nucleic Acids Res.*, **45**, 5850–5862.
32. Muniyappa, K., Anuradha, S. and Byers, B. (2000) Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation. *Mol. Cell Biol.*, **20**, 1361–1369.
33. Zheng, K.W., He, Y.D., Liu, H.H., Li, X.M., Hao, Y.H. and Tan, Z. (2017) Superhelicity constrains a localized and R-loop-dependent formation of G-quadruplexes at the upstream region of transcription. *ACS Chem. Biol.*, **12**, 2609–2618.
34. Selvam, S., Koirala, D., Yu, Z. and Mao, H. (2014) Quantification of topological coupling between DNA superhelicity and G-quadruplex formation. *J. Am. Chem. Soc.*, **136**, 13967–13970.
35. Chalikian, T.V., Liu, L. and Macgregor, R.B. (2020) Duplex-tetraplex equilibria in guanine- and cytosine-rich DNA. *Biophys. Chem.*, **267**, 106473.
36. De Magis, A., Götz, S., Hajikazemi, M., Fekete-Szücs, E., Caterino, M., Juranek, S. and Paeschke, K. (2020) Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nature Commun.*, **11**, 3907.
37. Lam, E.Y.N., Beraldi, D., Tannahill, D. and Balasubramanian, S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.
38. Prioleau, M.-N. (2017) G-Quadruplexes and DNA replication origins. *Adv. Exp. Med. Biol.*, **1042**, 273–286.
39. You, H., Zeng, X., Xu, Y., Lim, C.J., Efremov, A.K., Phan, A.T. and Yan, J. (2019) Dynamics and stability of polymorphic human telomeric G-quadruplex under tension. *Nucleic Acids Res.*, **42**, 8789–8795.
40. Cheng, Y., Tang, Q., Li, Y., Zhang, Y., Zhao, C., Yan, J. and You, H. (2019) Folding/unfolding kinetics of G-quadruplexes upstream of the P1 promoter of the human BCL-2 oncogene. *J. Biol. Chem.*, **294**, 5890–5895.
41. Yu, Z., Schonhofs, J.D., Dhakal, S., Bajracharya, R., Hegde, R., Basu, S. and Mao, H. (2009) ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities. *J. Am. Chem. Soc.*, **131**, 1876–1882.
42. You, H., Wu, J., Shao, F. and Yan, J. (2015) Stability and kinetics of c-MYC promoter G-quadruplexes studied by single-molecule manipulation. *J. Am. Chem. Soc.*, **137**, 2424–2427.
43. Mitra, J., Makurath, M.A., Ngo, T.T., Troitskaia, A., Chemla, Y.R. and Ha, T. (2019) Extreme mechanical diversity of human telomeric DNA revealed by fluorescence-force spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8350–8359.
44. 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.
45. Dhakal, S., Yu, Z., Konik, R., Cui, Y., Koirala, D. and Mao, H. (2012) G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys. J.*, **102**, 2575–2584.
46. Jeitany, M., Pineda, J.R., Liu, Q., Porreca, R.M., Hoffschir, F., Desmaze, C., Silvestre, D.C., Mailliet, P., Junier, M.P., Londono-Vallejo, A. *et al.* (2015) A preclinical mouse model of glioma with an alternative mechanism of telomere maintenance (ALT). *Int. J. Cancer*, **136**, 1546–1558.
47. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
48. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. and Balasubramanian, S. (2018) Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature Protoc.*, **13**, 551–564.
49. Pennarun, G., Granotier, C., Gauthier, L.R., Gomez, D., Hoffschir, F., Mandine, E., Riou, J.-F., Mergny, J.-L., Mailliet, P. and Boussin, F.D. (2005) Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene*, **24**, 2917–2928.
50. Byrd, A.K., Zybailov, B.L., Maddukuri, L., Gao, J., Marecki, J.C., Jaiswal, M., Bell, M.R., Griffin, W.C., Reed, M.R., Chib, S. *et al.* (2016) Evidence that G-quadruplex DNA accumulates in the cytoplasm and participates in stress granule assembly in response to oxidative stress. *J. Biol. Chem.*, **291**, 18041–18057.
51. Raj, S., Bagchi, D., Valle Orero, J., Banroques, J., Tanner, N.K. and Croquette, V. (2019) Mechanistic characterization of the DEAD-box RNA helicase Ded1 from yeast as revealed by a novel technique using single-molecule magnetic tweezers. *Nucleic Acids Res.*, **47**, 3699–3710.
52. Bhattacharyya, D., Mirihana Arachchilage, G. and Basu, S. (2016) Metal cations in G-quadruplex folding and stability. *Frontiers Chem.*, **4**, 38.
53. Gosse, C. and Croquette, V. (2002) Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys. J.*, **82**, 3314–3329.
54. Strick, T.R., Allemand, J.-F., Bensimon, D., Bensimon, A. and Croquette, V. (1996) The elasticity of a single supercoiled DNA molecule. *Science*, **271**, 1835–1837.
55. 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.
56. Ding, F., Manosas, M., Spiering, M.M., Benkovic, S.J., Bensimon, D., Allemand, J.-F. and Croquette, V. (2012) Single-molecule mechanical identification and sequencing. *Nat. Methods*, **9**, 367372.
57. Allshire, R.C., Dempster, M. and Hastie, N.D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res.*, **17**, 4611–4627.

58. Gabelica, V., Baker, E.S., Teulade-Fichou, M.P., De Pauw, E. and Bowers, M.T. (2007) Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: The role of central cations and small planar ligands. *J. Am. Chem. Soc.*, **129**, 895–904.
59. Mendez-Bermudez, A., Hills, M., Pickett, H.A., Phan, A.T., Mergny, J.-L., Riou, J.-F. and Royle, N.J. (2009) Human telomeres that contain (CTAGGG)<sub>n</sub> repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res.*, **37**, 6225–6238.
60. Lim, K.W., Alberti, P., Guédin, A., Lacroix, L., Riou, J.-F., Royle, N.J., Mergny, J.-L. and Phan, A.T. (2009) Sequence variant (CTAGGG)<sub>n</sub> in the human telomere favors a G-quadruplex structure containing a GCGC tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.
61. Asamitsu, S., Bata, S., Yu, Z., Bando, T. and Sugiyama, H. (2019) Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules*, **24**, 429.
62. 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.
63. Monsen, R.C. and Trent, J.O. (2018) G-quadruplex virtual drug screening: a review. *Biochimie*, **152**, 134–148.
64. Wong, H.M., Payet, L. and Huppert, J.L. (2009) Function and targeting of G-quadruplexes. *Curr. Opin. Mol. Ther.*, **11**, 146–155.
65. Piazza, A., Boulé, J.-B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M.-P. and Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **38**, 4337–4348.
66. Di Antonio, M., Ponjavic, A., Radzevicius, A., Ranasinghe, R.T., Catalano, M., Zhang, X., Shen, J., Needham, L.M., Lee, S.F., Klenerman, D. *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.*, **12**, 832–837.
67. Granotier, C., Pennarun, G., Riou, L., Hoffschir, F., Gauthier, L.R., De Cian, A., Gomez, D., Mandine, E., Riou, J.-F., Mergny, J.-L. *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.
68. Marchand, A., Strzelecka, D. and Gabelica, V. (2016) Selective and cooperative ligand binding to antiparallel human telomeric DNA G-quadruplexes. *Chemistry*, **22**, 9551–9555.
69. Bhat, J., Mondal, S., Sengupta, P. and Chatterjee, S. (2017) *In silico* screening and binding characterization of small molecules toward a G-quadruplex structure formed in the promoter region of c-MYC oncogene. *ACS Omega*, **2**, 4382–4397.
70. Tran, P.L.T., Largy, E., Hamon, F., Teulade-Fichou, M.-P. and Mergny, J.-L. (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie*, **93**, 1288–1296.
71. Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
72. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
73. Koirala, D., Dhakal, S., Ashbridge, B., Sannohe, Y., Rodriguez, R., Sugiyama, H., Balasubramanian, S. and Mao, H. (2011) A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nat. Chem.*, **3**, 782–787.
74. Puig Lombardi, E. and Londono-Vallejo, A. (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.*, **48**, 1–15.