Folding kinetics of G-quadruplexes : duplex stem loops drive and accelerate G-quadruplex folding

Nguyen, Thi Quynh Ngoc; Lim, Kah Wai; Phan, Anh Tuấn

2020

Nguyen, T. Q. N., Lim, K. W. & Phan, A. T. (2020). Folding kinetics of G-quadruplexes : duplex stem loops drive and accelerate G-quadruplex folding. The Journal of Physical Chemistry B, 124(25), 5122-5130. <https://dx.doi.org/10.1021/acs.jpcb.0c02548>

<https://hdl.handle.net/10356/150057>

<https://doi.org/10.1021/acs.jpcb.0c02548>

This document is the Accepted Manuscript version of a Published Work that appeared in final form in The Journal of Physical Chemistry B, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://doi.org/10.1021/acs.jpcb.0c02548>

Downloaded on 26 Apr 2023 18:10:45 SGT

Folding Kinetics of G-quadruplexes: Duplex Stem Loops Drive and Accelerate G-quadruplex Folding

Thi Quynh Ngoc Nguyen,[†] Kah Wai Lim[†] & Anh Tuan Phan^{*,†,‡}

[†]School of Physical and Mathematical Sciences, Nanyang Technological University,
Singapore 637371, Singapore

[‡]NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921,
Singapore

*Corresponding author: phantuan@ntu.edu.sg

Thi Quynh Ngoc Nguyen's present address: Department of Pharmacy, National University of
Singapore, Singapore

Key words: folding kinetics; G-quadruplex; loop structure

ABSTRACT

The folding kinetics is an important parameter affecting the relevance of G-quadruplex (G4) structures in biological processes. Previous studies of G4 folding kinetics mainly depended on fast measurement techniques. In previous available examples of G4s with loops up to three residues, the folding kinetics spanning several orders of magnitude has been reported, ranging from milliseconds to over 100 s. It is difficult to systematically and fundamentally understand the effect of multiple parameters, especially the loop properties, on the G4 folding kinetics, as the G4 fold is often altered when the sequence is varied. In this study, judicious choices of multiple parameters allowed us to bring various systems into the measurable window of a simple UV absorption technique. In conjunction with a well-controlled parallel-stranded G4 fold, we were able to systematically investigate the effect of 7 different parameters (environment and loop properties) on G4 folding kinetics. We found that structured loops can drive up the G4 folding rate: for a long loop, the fast folding of a stem loop can guide the G4 folding and accelerate its folding kinetics by several orders of magnitude compared to an unstructured loop counterpart.

INTRODUCTION

Guanine-rich nucleic acid sequences have a propensity to adopt a non-canonical structure known as the G-quadruplex (G4)¹⁻³ in the presence of cations such as potassium and sodium. The G4 is a four-stranded helical structure built from multiple G•G•G•G tetrads⁴ stacked on top of each other. G4-forming sequences are enriched in important regions of the human genome such as promoters,⁵ 5'-UTR RNA³ and telomeres,⁶ indicating the regulatory role of G4 structures at the DNA and RNA level. There has been increasing evidence of G4 formation and function in biological processes,⁷⁻⁸ making G4 a promising target for diseases.⁹⁻¹⁰

The G4 structure was shown to be compatible with other structural motifs of nucleic acids. Previous studies have shown that a single duplex stem could be connected to a G4 core in different ways: coaxial or orthogonal orientation.¹¹ Multiple duplex stems could form within a single G4 loop.¹² While unstructured long loops were observed to destabilize G4s, duplex stems within the loops could have positive effect on the thermal stability of G4s as compared to non-structured loops.¹³

While the thermodynamic behaviors of G4s have been extensively studied, much less knowledge about the G4 folding kinetics was reported in the literature. In cells, G4 formation is competitive with the binding of the complementary strand in the case of a G-rich sequence located within a duplex, or DNA binding protein in the case of a telomeric single strand located at the end of the chromosome.¹⁴ The folding rate of a G4 is an important parameter when its formation takes place in cells during various time scales of biological processes such as replication, transcription and translation. G-rich strands could be liberated during transcription or replication when the duplex is separated by a helicase, or the DNA binding protein is dissociated from single stranded telomeric DNA, allowing the formation of G4s. During these biological processes, potential G4 formation in cells could be governed by the kinetic rather than thermodynamic properties.

Previous studies proposed that the folding pathway of G4 involves multiple steps and intermediates. Random-coiled DNA could initially fold into a G-hairpin, before folding into a G-triplex, and then finally into G4.¹⁵⁻¹⁸ Some studies proposed that unfolded DNAs experience four-stranded misfolded conformation before finally rearranging into the correct fold.¹⁹⁻²³ While the final fold of 3-layer G4s often requires two or three monovalent cations for stabilization, the G-hairpin and G-triplex can be stabilized by one cation.^{15-16, 18}

Folding kinetics of G4 formation was mainly studied on human telomeric DNA and RNA sequences.^{16-17, 23-24} Folding kinetics of G4s was observed by monitoring the change in UV,^{16, 23} single molecule FRET,¹⁷ CD,^{17, 22} fluorescence,²⁴⁻²⁵ and NMR^{20, 26}. The folding kinetics was reported to be dependent on the temperature and the nature/concentration of cations.^{16, 23} The folding kinetics of DNA and RNA telomeric G4s were observed to be similar.¹⁶ G4-forming sequences containing G₃ tracts (potentially forming 3-layer G4) were found to fold much faster than those containing G₂ tracts (potentially forming 2-layer G4).¹⁶ Some examples of limited loop size variation suggested that sequences with short loops fold faster than those with long loops.^{16, 27} Xue *et al.* have shown that while the parallel G4 conformation of human telomeric DNA is more abundant than the (3+1) hybrid G4 form in crowding condition, the latter could fold much faster.²⁴ While the question on how different G4 conformations affect the folding kinetics is very interesting,²⁴ for fundamental understanding of G4 folding it is important to study the effect of a particular parameter (ranging from the external environment to the DNA/RNA properties) on the folding kinetics of the same G4 fold. In previous available examples of G4s with loops up to 3 residues, the folding kinetics spanning 5 orders of magnitude has been reported (ranging from the milliseconds comparable to the transcription/replication time scale to over 100 s exceeding the lifetimes of some regulatory RNAs).¹⁶ It is difficult to systematically and fundamentally understand the effect of multiple parameters on the G4 folding kinetics, as sequence variation can often lead to a change in the G4 fold.

Here we systematically investigate the effect of 7 different parameters (environment and loop properties) on the folding kinetics of a well-controlled parallel G4. Judicious choices of multiple parameters allowed us to bring various systems into the measurable window of a simple UV absorption technique. Our systematic and multi-parameter study revealed some simple rules governing the folding kinetics of G4s. Particularly, we found that the fast folding of the stem loop can guide the G4 folding and accelerate its folding kinetics by several orders of magnitude compared to an unstructured loop.

METHODS

DNA Sample Preparation

Oligonucleotides were synthesized on ABI 394 DNA/RNA synthesizer using product from Glen Research and Cambridge Isotope Laboratories. Samples were then purified following

Glen Research Protocol and dialyzed successively again water, 10 mM Tris pH 7.0, and water. Next, samples were dried in lyophilizer, and then dissolved in a buffer containing 10 mM Tris (pH 7.0).

UV Absorption

The UV measurement was conducted on a JASCO V-650 spectrophotometer using 1- cm path length quartz cuvettes, unless otherwise stated. UV measurements of 50 and 100 μ M DNA samples were conducted with 1- mm path length quartz cuvettes. The full UV spectra were recorded from 220 nm to 320 nm at a scanning rate of 400 nm/min, data intervals of 1 nm, and bandwidth of 2 nm. Unless otherwise specified, UV signal was recorded at 25 °C.

Circular Dichroism

Circular dichroism (CD) spectra were recorded over the range of 220–320 nm on a Jasco-815 spectropolarimeter at a scanning rate of 500 nm/min, data intervals of 1 nm and bandwidth of 2 nm. Experiments were performed with quartz cuvettes, 1-cm path-length for DNA samples of 4 μ M, or 1- mm path-length for DNA samples of 50 μ M. Unless otherwise specified, the CD spectra were recorded at 25 °C.

NMR Spectroscopy

NMR experiments were performed on 600 MHz Bruker spectrometers at 25 °C, unless otherwise specified. DNA samples were dissolved in a buffer contains 15 mM KCl and 10 mM Tris (pH 7). Standard concentration of NMR samples was typically 0.2–1.0 mM.

Kinetic Measurements

Oligonucleotide and KCl were dissolved separately in 10 mM Tris buffer (pH 7.0). G4 folding was induced by manual adding KCl in 10 mM Tris (pH 7.0) to oligonucleotide sample in 10 mM Tris (pH 7.0) buffer to give a final reaction of 4 μ M oligonucleotide. The folding kinetics of oligonucleotide was monitor via UV absorbance changes at 295 nm as a function of time (path length 1 cm, bandwidth 2 nm). The dead time of these manual mixing experiments, from addition of a cation to initiation of data collection, is 3-7s. To keep the temperature constant, cuvettes of oligonucleotide and KCl solution were placed in UV chamber 30 minutes prior to mixing. Each experiment was repeated three times and data were averaged for analysis. Plots of raw data were fitted to exponential relaxation using Sigma Plot software.

RESULTS AND DISCUSSION

Measuring G4 folding kinetics

As the primary goal of this study was to investigate the effect of various properties of a loop on the folding kinetics of a G4, we first wanted to make sure that all studied sequences containing varying middle loops folded into the same folding topology. The Circular Dichroism (CD) spectra of all studied sequences (Table 1) show a positive peak at ~260 nm and a negative trough at ~240 nm (Figure S1), consistent with the formation of a parallel-stranded G4 structure.²⁸ The Nuclear Magnetic Resonance (NMR) spectra of sequences with varying size of the middle loop (Figure S2A) show imino protons at 11-12 ppm with similar spectral patterns (the spectral pattern gradually shifts with the increase of the size of the central loop), suggesting the formation of the same G4 fold as the reference propeller-type parallel G4 named *T95-2T*²⁹ where each propeller loop is a single thymine residue (Figure S3). In the case of *K-5T* where the central loop consists of five thymines, unambiguous spectral assignments were performed for some specific guanines in the sequence, supporting the formation of the same folding topology (Figure S3). For four other sequences (*K-3bp6T*, *K-6bp6T*, *K-10bp6T* and *K-18bp6T*, Table 1), a duplex stem of 3, 6, 10 and 18 base pairs was inserted in the middle loop; the formation of additional Watson-Crick base pairs in the duplex stem was indicated by the observation of imino protons at 12-14 ppm (Figure S2B).¹³

Table 1. Sequences used in this study

Name	Sequences			
<i>T95-2T</i>	TT GGG T GGG	T	GGG T GGG T	
<i>K-5T</i>	TT GGG T GGG	T ₅	GGG T GGG T	
<i>K-10T</i>	TT GGG T GGG	T ₁₀	GGG T GGG T	
<i>K-15T</i>	TT GGG T GGG	T ₁₅	GGG T GGG T	
<i>K-20T</i>	TT GGG T GGG	T ₂₀	GGG T GGG T	
<i>K-25T</i>	TT GGG T GGG	T ₂₅	GGG T GGG T	
<i>K-5T5A</i>	TT GGG T GGG	TAATTAAATT	GGG T GGG T	
<i>K-3bp6T</i>	TT GGG T GGG	T ₃ CGC GCA GCGT ₃	GGG T GGG T	
<i>K-6bp6T</i>	TT GGG T GGG	T ₃ CGCGAA GCA TTCGCGT ₃	GGG T GGG T	
<i>K-10bp6T</i>	TT GGG T GGG	T ₃ CGCGAATCAT GCA ATGATTTCGCGT ₃	GGG T GGG T	
<i>K-18bp6T</i>	TT GGG T GGG	T ₃ CGCGAATCTGTGATCGAT GCA ATCGATCACAGATTTCGCGT ₃	GGG T GGG T	

The progress of each folding reaction was followed over time by recording the UV absorbance signal at 295 nm (*Abs*), which is a measure of stacking of G-tetrads and therefore G4 structure formation.³⁰ The G4-forming DNA oligonucleotides were allowed to fold until the absorbance signal approached a constant value, signifying that the G4 folding reaction had reached

equilibrium. The observed data were well described in terms of single exponential relaxation as a function of time: $\Delta Abs = A_1 + A_2 \exp(-kt)$, where A_1 and A_2 are constants, k is the folding rate, and t is time (Figure S4A). Although fitting to a double-exponential relaxation (Figure S4B) resulted in similar R value and residual plot, the relaxation rate of the major component was similar to that obtained in a single-exponential fitting and the weightage of the second exponential is small (only about 3 %) compared to that of the first exponential. We concluded that our data best fit to single exponential relaxation and the G4 folding time can be derived as: $\tau = 1/k$ in which k is folding rate (as mentioned above).

Folding data for *K-20T* (Table 1) – a G4-forming sequence with a central loop of 20 thymines – are shown in Figure 1A. In contrast to K^+ , Li^+ does not favor G4 formation;³¹⁻³² a control experiment of mixing the same oligonucleotide in a buffer containing LiCl showed no significant change of absorbance at 295 nm over time. Folding of *K-20T* in 15 mM KCl was also observed by CD and NMR spectroscopy. The G4 folding process followed by CD at 264 nm is shown in Figure 1B: after adding KCl, the molar ellipticity at 264 nm gradually increased and plateaued after ~4000 s. Initially, one-dimensional (1D) imino proton NMR spectrum of *K-20T* in Tris buffer (pH 7.0) showed no peaks in the region of 10-12.5 ppm, characteristic for G4 formation. After adding KCl, 12 peaks appeared and increased with time in the region of 10-12.5 ppm, indicating the formation of a G4 (Figure 1C). Superimposition of folding curves monitored by UV, CD and NMR gives similar results on the folding kinetics (Figure 1D), cross-validating different methods of measurements. Folding kinetics of a G4 may involve multiple steps¹⁵⁻²³ in which the last step is usually the longest one^{16, 22-23}. During 3-7 s dead-time of our manual mixing measurement, some fast events at the beginning could be not observed. However, the evolution of imino protons in NMR spectra (Figure 1C) indicated that our measurement reflects a main folding event involving the formation of Hoogsteen hydrogen bonds between guanines to bring them into the final G4 fold. Furthermore, little difference in the folding kinetics was observed for DNA concentrations in the range of 4-100 μ M (Figure S5). In the following sections, the G4 folding is monitored by UV absorption at 295 nm for a sample containing 4 μ M DNA, unless otherwise stated. In these sections, we will first investigate the effect of the experimental conditions and environment on the folding kinetics of G4. Next, we will investigate the effect of various properties of a loop including loop length, sequence and structure on the folding kinetics of G4.

Figure 1. Monitoring of G4 formation of *K-20T* using different spectroscopy: (A) UV, (B) CD, (C) NMR. G4 formation was induced by adding KCl into oligonucleotide in 10 mM Tris buffer (pH 7) at 25°C, final oligonucleotide and KCl concentrations were 50 μ M and 15 mM, respectively (A) left: UV different spectra of *K-20T*, right: Changing in UV absorbance at 295 nm of *K-20T* induced by adding KCl: experimental data (blue); fitting to single exponential (red), by adding LiCl (grey). (B) left: CD spectra of *K-20T* at different time points, right: Folding curve of *K-20T* monitored by change in CD absorption at 264 nm. (C) left: 1D imino proton NMR spectra of *K-20T* before and after adding 15 mM KCl at different time points, right: Normalized fraction folded curve. (D) Superposition of folding curves observed by different methods of measurement: UV (blue), CD (black) and NMR (pink).

Effect of temperature on G4 folding kinetics

The effect of temperature on the folding kinetics of *K-20T* in buffer containing 15 mM KCl (Figure S6) is presented in Figure 2. As temperature increased from 15 to 35°C, the folding time of *K-20T* increased by about 4 folds. Increasing of the folding time with temperature indicates that the folding activation energy is negative. A similar effect of temperature on folding kinetics of an intramolecular G4 was observed in a previous study for human telomeric DNA, which showed that the folding rate of human telomere in K^+ decreased 5-fold between 15 and 45°C.²³ The slope of the plot in Figure 2 corresponds to an apparent activation energy of -48.8 kJ/mol. Negative activation energies were previously observed in the folding kinetics of other biomolecules, such as DNA hairpin,³³ β -hairpin-forming peptide³⁴ and protein,³⁵ with possible involvement of multiple hidden elementary steps³⁶.

Figure 2. Effect of temperature on the folding time of *K-20T* in a buffer containing 15 mM KCl. Final oligonucleotide concentration was 4 μ M. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Effect of K^+ concentration on G4 folding kinetics

The presence of cations in the G-tetrad core is an important factor for stabilizing a G4 structure. In a G-tetrad, monovalent cations such as Na^+ and K^+ coordinate with guanine O6 atoms to neutralize the partial negative charges of these oxygen atoms. Therefore, it is of interest to study the role of K^+ in the G4 folding kinetics. Folding reactions of *K-20T* (4 μ M) was performed in solutions containing KCl at the concentration of 5, 15 and 30 mM to examine the dependence of the G4 folding kinetics on the potassium concentration (Figure S7). The folding rate (inverse of folding time) decreased significantly as the KCl concentration decreased from 30 to 5 mM (Figure 3). A previous study³⁷ has shown that the exchange of a cation bound to a

G4 with those in solution occurs at a rate greater than 10^3 s^{-1} which is much higher than the G4 folding rate observed in this study, therefore the dynamics of cation exchange would not influence the G4 folding rate. To clarify whether the folding reaction depends on the KCl concentration itself or the ionic strength, we carried the folding reaction of *K-20T* in solutions of 5 mM KCl + 25 mM LiCl and 15 mM KCl + 15 mM LiCl to keep the ionic strength ($\text{K}^+ + \text{Li}^+$) constant. In the solution of the same ionic strength, the folding time of *K-20T* was also observed to increase with a decrease of the K^+ concentration but with a lower slope than in the previous experiment. This indicates that the folding time of G4 depends on both the ionic strength and the K^+ concentration. A previous study on folding kinetic of antiparallel and (3+1) hybrid G4s adopted by a human telomere sequence also showed that adding Li^+ together with Na^+ or K^+ cation also leads to increased folding rate of these G4 forms.³⁸ The effect of Li^+ on accelerating folding rate of G4 could implicate the potential role of Li^+ in stabilization of intermediates in G4 folding pathway.

Figure 3. Dependence of the folding time of *K-20T* at 25°C on the concentration of K^+ (filled circle), and the concentration of K^+ and Li^+ where total concentration of K^+ and Li^+ was kept constantly at 30 mM (open circle). Final oligonucleotide concentration was 4 μM . Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Effect of a co-solute, PEG200, on G4 folding kinetics

It was reported that the folding rate of G4s increased with increasing amount of co-solutes including methanol and ethanol.³⁹ Co-solutes such as ethanol and PEG were shown to deplete water and interact favorably at the exposed G-tetrad planar surface of a parallel-stranded G4, hence, favoring it over other human telomeric G4 conformations.⁴⁰⁻⁴³ Unlike methanol or ethanol, PEG is highly viscous which could lead to a deceleration of G4 folding kinetics.⁴⁴ Here, we investigated the effect of addition of various amounts of PEG200 on the folding

kinetics of *K-15T*, an all-parallel-stranded G4 (Figure S8). In spite of high viscosity of PEG200, their effect on G4 folding kinetics was similar to other reported less viscous co-solutes,³⁹ the folding time of *K-15T* was observed to decrease with the increase of the PEG200 concentration from 20 to 40% (Figure 4). This indicates that the effect of high viscosity of PEG200, if exists, is overcome by its G4 interaction and water depletion effects. Small-molecules G4 ligands could have a similar effect as PEG200 on accelerating the folding of G4s. The bioactivity of G4 binding ligands could arise from their effect on G4 folding kinetics in addition to the effect on thermal and mechanical stabilization.

Figure 4. Dependence of folding time of *K-15T* on percentage of added PEG200. G4 formation was induced by adding oligonucleotide to buffer containing 10 mM Tris (pH 7), KCl and different amount of PEG, final concentrations of oligonucleotide and KCl were 4 μ M and 0.5 mM, respectively. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Effect of loop length on G4 folding kinetics

Effect of loop length on the thermal stability of G4s has been systematically studied,⁴⁵ showing that an increase in length of non-structured loop leads to a decrease in stability of G4s. While the effect of loop size on folding kinetics of G4s has not been systematically studied, some reported examples suggested that sequences with short loops fold faster than those with long loops.^{16,27} Here, we systematically investigated the effect of loop length on folding kinetics of a parallel G4 structure. The folding kinetics of five sequences with the middle propeller loop being 10T, 15T, 20T, 25T were studied in buffer containing 15 mM KCl (Figure S9). The folding time of these sequences was observed to increase with increasing loop length (Figure 5). A similar trend was observed for RNA G4s in a previous study which showed that the folding rate increased significantly when the loop length was decreased.¹⁶ This observation is also similar to the folding kinetic of hairpin.⁴⁶ A longer loop reduces the likelihood of G-tracts

coming close together in the context of a random coil, and consequently lead to increased folding time of G4s.

A study on genome instability in *S. cerevisiae* showed that shortened loops (≤ 4 nt) but not long-loop G4 motifs are able to induce genomic instability by acting as a stable roadblock for the replicative polymerase.⁴⁷ The authors concluded that high thermodynamic stability of short loop length G4s is correlated with genome variant instability. With our finding on the dependence of folding kinetics on loop length of G4, high risk of short loop length for genomic instability may be explained by the fact that a short loop length sequence could fold into a G4 faster and therefore could be formed within the replication time scale while long loop sequences do not. A recent study showed that altering the loop of *pilG4*, a G4 structure required for pilin antigenic variation by recombination, decrease or abrogate antigenic variation.⁴⁸ The antigenic variation is mediated by homologous recombination which involves the binding of RecA recombinase to *pilG4*. *In vitro* experimental data showed that 2 loop variants, T:TT:AT and T:TT:T, had a similar binding affinity to RecA but these mutants displayed significant antigenic variation efficiency *in vivo*, suggesting that the G4 folding kinetics of these loop mutants could determine their biological effect.

Figure 5. Dependence of folding time on different loop size. G4 formation was induced by adding KCl to oligonucleotide in 10 mM Tris buffer (pH 7) at 25°C, final oligonucleotide and KCl concentrations were 4 μ M and 15 mM, respectively. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Effect of loop sequence

The effect of loop sequence on folding kinetic of duplex hairpin was investigated previously which showed that the folding rates of duplexes with loops of the same size but different sequences were remarkably different.⁴⁶ The effect of loop sequence on folding kinetics of a 1-2-nt double-chain-reversal (or propeller) loop has been reported.²⁵ Here, we studied the effect

of different loop sequences with the same size (10 nt) on the G4 folding kinetics (Figure S10). In agreement with the previous findings on loop dynamics in hairpins⁴⁶ the folding time of *K-10T* which contains only T in the loop was observed to be longer than that of *K-5T5A* which has the same loop size but contains both thymines and adenines in the loop (Figure 6).

Figure 6. Dependence of folding time on different temperature of *K-5T5A* (filled circle) and *K-10T* (open circle). G4 formation was induced by adding KCl to oligonucleotide in 10 mM Tris buffer (pH 7), final oligonucleotide and KCl concentrations were 4 μ M and 5 mM, respectively. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Effect of loop secondary structure: A duplex stem loop accelerates G4 folding rate

Previous thermodynamic study on G4s showed that increases in a non-structured loop length reduced the stability of a G4 while a reverse trend was observed for a structured loop.¹⁹ However, effect of a structured loop on the folding kinetics of G4 remains unknown. Here, we compared the folding rate of a G4 structure containing one duplex hairpin motif of 15 bases in the middle loop (*K-3bp6T*, Figure 7A) and another G4 structure containing 15 thymines in a non-structured middle loop (*K-15T*, Figure 7B). Interestingly, our data showed that structured loop could reduce folding time significantly as compared to non-structured loop. In 0.5 mM KCl, folding time of *K-3bp6T* was 167.8 s while that of *K-15T* was 4487.9 s (Figure 7C). It was shown that a duplex can be formed in microsecond to millisecond time scale^{23,25,26}. When cation is introduced, duplex hairpin would quickly form and act as a zip to bring distant G-tracts closer together. The random-coil nature of non-structured loop would reduce the possibility of bringing G-tracts close to each other. Enhancement of structured loop on G4 folding rate implies the significance of such G4-forming motifs in human genome, since they may have higher probability to form G4 in the time scale of replication or transcription.

Figure 7. Effect of loop structure on G4 folding kinetics: (A) Schematic of stem loop *K-3bp6T*, (B) schematic of unstructured loop *K-15T* and (C) folding curve of *K-3bp6T* and *K-15T*. G4 formation was induced by adding KCl to oligonucleotide in 10 mM Tris buffer (pH 7) and LiCl at 25°C, the final oligonucleotide, LiCl and KCl concentrations were 4 μ M, 20 mM and 0.5 mM, respectively.

Folding kinetics of stem-loop G4s with various stem length

Folding of four sequences containing one stem in the middle loop with a length of 3, 6, 10 and 18 base pairs (*K-3bp6T*, *K-6bp6T*, *K-10bp6T* and *K-18bp6T*, Table 1) in 0.5 mM KCl were investigated to examine the effect of stem length on the G4 folding kinetics (Figures S11-12). In the case of non-structured loop, it was shown in previous section that folding rate was strongly dependent on loop size. In the case of stem loop G4, the effect of stem loop size on folding rate of G4 is totally different. As shown in Figure 8, folding rate of *K-3bp6T*, *K-6bp6T*, *K-10bp6T* and *K-18bp6T* are almost similar although the length of the central loop in these oligonucleotides ranges from 15 bases to 45 bases. This is in contrast to the dependency of the G4 folding time on the length of the central unstructured loop (plotted as dashed line in Figure 8). Note that in order to compare in the same measurable window, experiments for unstructured loop G4s were performed in 15 mM K^+ instead of 0.5 mM K^+ for stem loop G4s. Based on the salt-dependent and loop-length-dependent studies presented above, the folding time for stem loop G4 and unstructured loop G4 can be different by several orders of magnitude at the same salt concentration. This observation suggests that folding kinetics of G4 containing duplex stem in loop is mostly independent on stem length. The reason for this independence of folding kinetics on loop length could be explained by similar context at the junction between duplex and G4 regardless of loop length. In addition, the folding time of stem loop G4 containing 29

and 45 nucleotides in the middle loop (*K-10bp6T* and *K-18bp6T*) was lower than that of unstructured loop G4, *K-5T* which contains only 5 residues in the middle loop (Figure S11).

Figure 8. Dependence of folding time on size of stem loop. G4 formation was induced by adding KCl to oligonucleotide in 10 mM Tris buffer (pH 7) at 25°C, the final oligonucleotide and KCl concentrations were 4 μ M and 0.5 mM, respectively. For comparison, the folding times of G4 of unstructured loop of different sizes are plotted in the same graph and connected by a dash line. Note that data for stem loops were measured for 0.5 mM KCl, while those for unstructured loops were measured for 15 mM KCl. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

Fast folding kinetics of stem loop G4 does not require pre-arrangement of the loop.

Our data has shown that a sequence with a structured loop could fold into G4 at a significantly higher rate than non-structured loop sequences. In experiments in the previous section, all DNA were dissolved in 10 mM Tris pH 7.0 buffer. In this buffer, the stem loop could be folded before adding KCl (Figure 9A) as indicated by peaks in region of 12.5-14 ppm of imino proton NMR spectra (Figure S13A). A question arises as to whether pre-formation of a structured loop leads to higher folding rate of G4. To address this question, we dissolved *K-10bp6T* into KOH 0.5 M (final pH ~13.5) to prevent formation of the duplex stem with Watson-Crick base-pairs (Figure 9B). NMR spectrum of this sequence in KOH 0.5 M pH~13.5 shows no peak in the region of 10-14 ppm, indicating that no Watson-Crick base pair and G4 was formed (Figure S13B). The folding kinetics of *K-10bp6T* initially in pH 13.5 was then investigated by adding 499.5 μ l of 10 mM Tris pH 7.0 buffers to 0.5 μ l of DNA in KOH, with a final K^+ concentration of 0.5 mM. After adding 10 mM Tris, the pH of the oligonucleotide solution quickly dropped close to a neutral value (as verified by a pH meter), allowing the formation of both duplex stem and G4 (Figure 9C, Figure S13C). The result (Figure 9D) shows the high similarity in folding rate of *K-10bp6T* with and without pre-dissolving in KOH 0.5 M. This observation suggests

that the pre-formation of the stem does not affect the effect of stem loop on driving up the folding rate of G4. Previous studies of folding kinetics of duplex hairpin showed that duplex hairpin could fold within microsecond to millisecond time scale^{23,26} which is much shorter than the G4 folding time presented in this study. Therefore, when the folding of oligonucleotide is induced by changing pH, the duplex stem loop would form first and bring the G-tracts close to each other. The condition of experiment in which G4 folding was induced by a pH jump from 13.5 to 7.0 in K^+ solution is close to the context in cells where the single strands are released from complementaries, allowing duplex stem and G4 to start to fold simultaneously.

Figure 9. Schematic showing structure of *K-10bp6T* in (A) 10 mM Tris pH 7.0, (B) 0.5 M KOH pH 13.5, and (C) 10 mM Tris pH 7.0 and 0.5 mM KCl. (D) Folding time of *K-10bp6T* as a function of temperature in two folding experiments (i) inducing by adding KCl to oligonucleotide in 10 mM Tris buffer, pH7 (gray) and (ii) inducing by adding Tris buffer, pH 7.0 into 4 μ M oligonucleotide in 0.5 M KOH (black). Final oligonucleotide and KCl concentrations were 4 μ M and 0.5 mM, respectively. Error bars indicate the standard deviation of three repeated measurements for each condition. Logarithmic scale is used for the vertical axis.

CONCLUSION

In summary, we systematically studied the effect of different parameters on the folding kinetics of G4s by monitoring the change in UV absorbance at 295 nm over time. While adopting the same folding topology, the folding kinetics of a G4 greatly depends on temperature, K^+ concentration, ionic strength, co-solute, loop length, loop sequence and loop structure. Similar to duplex hairpin, the folding kinetics of a G4 structure is favored at high K^+ concentration, high ionic strength and short loop. Especially, our study showed that a stem loop (at the central position) accelerates the folding rate of a G4 significantly as compared to a non-structured loop, highlighting the possible biological relevance of G4s containing structured loops within the time scale of cellular processes. In contrast to a non-structured loop G4, the folding kinetics of a G4 involving a structured loop does not depend on the length of the loop.

ACKNOWLEDGEMENTS

This work was supported by the Singapore National Research Foundation Investigatorship (NRF-NRFI2017-09) and Singapore Ministry of Education Academic Research Fund Tier 3 (MOE2012-T3-1-001) and grants from Nanyang Technological University to A.T.P.

SUPPORTING INFORMATION

CD and NMR spectra of oligonucleotides used. Folding curves of different DNA sequences in different conditions observed by the change in UV absorbance at 295 nm. This information is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Davis, J. T., G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* **2004**, *43* (6), 668-98.
2. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **2006**, *34* (19), 5402-15.
3. Patel, D. J.; Phan, A. T.; Kuryavyi, V., Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res* **2007**, *35* (22), 7429-55.
4. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix formation by guanylic acid. *Proc Natl Acad Sci U S A* **1962**, *48*, 2013-8.
5. Huppert, J. L.; Balasubramanian, S., G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **2007**, *35* (2), 406-13.
6. Allshire, R. C.; Dempster, M.; Hastie, N. D., Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res* **1989**, *17* (12), 4611-27.
7. Rhodes, D.; Lipps, H. J., G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **2015**, *43* (18), 8627-37.

8. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **2013**, *5* (3), 182-6.
9. Balasubramanian, S.; Neidle, S., G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol* **2009**, *13* (3), 345-53.
10. Balasubramanian, S.; Hurley, L. H.; Neidle, S., Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov* **2011**, *10* (4), 261-75.
11. Lim, K. W.; Phan, A. T., Structural basis of DNA quadruplex-duplex junction formation. *Angew Chem Int Ed Engl* **2013**, *52* (33), 8566-9.
12. Lim, K. W.; Nguyen, T. Q.; Phan, A. T., Joining of multiple duplex stems at a single quadruplex loop. *J Am Chem Soc* **2014**, *136* (52), 17969-73.
13. Lim, K. W.; Khong, Z. J.; Phan, A. T., Thermal stability of DNA quadruplex-duplex hybrids. *Biochemistry* **2014**, *53* (1), 247-57.
14. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L., Targeting telomeres and telomerase. *Biochimie* **2008**, *90* (1), 131-155.
15. Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H., Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures. *J Am Chem Soc* **2010**, *132* (42), 14910-8.
16. Zhang, A. Y.; Balasubramanian, S., The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J Am Chem Soc* **2012**, *134* (46), 19297-308.
17. Gray, R. D.; Trent, J. O.; Chaires, J. B., Folding and unfolding pathways of the human telomeric G-quadruplex. *J Mol Biol* **2014**, *426* (8), 1629-50.
18. Marchand, A.; Gabelica, V., Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res* **2016**, *44* (22), 10999-11012.
19. Koirala, D.; Ghimire, C.; Bohrer, C.; Sannohe, Y.; Sugiyama, H.; Mao, H., Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J Am Chem Soc* **2013**, *135* (6), 2235-41.
20. Ceru, S.; Sket, P.; Prislán, I.; Lah, J.; Plavec, J., A new pathway of DNA G-quadruplex formation. *Angew Chem Int Ed Engl* **2014**, *53* (19), 4881-4.
21. Stadlbauer, P.; Kuhrova, P.; Vicherek, L.; Banas, P.; Otyepka, M.; Trantirek, L.; Sponer, J., Parallel G-triplexes and G-hairpins as potential transitory ensembles in the folding of parallel-stranded DNA G-Quadruplexes. *Nucleic Acids Res* **2019**, *47* (14), 7276-7293.
22. Gray, R. D.; Trent, J. O.; Arumugam, S.; Chaires, J. B., Folding landscape of a parallel G-quadruplex. *J Phys Chem Lett* **2019**, *10* (5), 1146-1151.
23. Gray, R. D.; Chaires, J. B., Kinetics and mechanism of K⁺- and Na⁺-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res* **2008**, *36* (12), 4191-203.
24. Xue, Y.; Liu, J. Q.; Zheng, K. W.; Kan, Z. Y.; Hao, Y. H.; Tan, Z., Kinetic and thermodynamic control of G-quadruplex folding. *Angew Chem Int Ed Engl* **2011**, *50* (35), 8046-50.
25. Gray, R. D.; Petraccone, L.; Trent, J. O.; Chaires, J. B., Characterization of a K⁺-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. *Biochemistry* **2010**, *49* (1), 179-94.
26. Bessi, I.; Jonker, H. R.; Richter, C.; Schwalbe, H., Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angew Chem Int Ed Engl* **2015**, *54* (29), 8444-8.
27. Mergny, J. L.; De Cian, A.; Amrane, S.; Webba da Silva, M., Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res* **2006**, *34* (8), 2386-97.
28. Gray, D. M.; Wen, J. D.; Gray, C. W.; Repges, R.; Repges, C.; Raabe, G.; Fleischhauer, J., Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **2008**, *20* (3-4), 431-40.
29. Do, N. Q.; Phan, A. T., Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry* **2012**, *18* (46), 14752-9.

30. Mergny, J. L.; Phan, A. T.; Lacroix, L., Following G-quartet formation by UV-spectroscopy. *FEBS Lett* **1998**, *435* (1), 74-8.
31. Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C., Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG₃GCG). *Biochemistry* **1992**, *31* (3), 833-41.
32. Sen, D.; Gilbert, W., A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **1990**, *344* (6265), 410-4.
33. Wallace, M. I.; Ying, L.; Balasubramanian, S.; Klenerman, D., Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc Natl Acad Sci U S A* **2001**, *98* (10), 5584-9.
34. Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., A statistical mechanical model for beta-hairpin kinetics. *Proc Natl Acad Sci U S A* **1998**, *95* (11), 5872-9.
35. Oliveberg, M.; Tan, Y. J.; Fersht, A. R., Negative activation enthalpies in the kinetics of protein folding. *Proc Natl Acad Sci U S A* **1995**, *92* (19), 8926-9.
36. Revell, L. E.; Williamson, B. E., Why are some reactions slower at higher temperatures? *J Chem Educ* **2013**, *90* (8), 1024-1027.
37. Hud, N. V.; Smith, F. W.; Anet, F. A.; Feigon, J., The selectivity for K⁺ versus Na⁺ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR. *Biochemistry* **1996**, *35* (48), 15383-90.
38. You, J.; Li, H.; Lu, X. M.; Li, W.; Wang, P. Y.; Dou, S. X.; Xi, X. G., Effects of monovalent cations on folding kinetics of G-quadruplexes. *Biosci Rep* **2017**, *37* (4), BSR20170771.
39. Marchand, A.; Ferreira, R.; Tateishi-Karimata, H.; Miyoshi, D.; Sugimoto, N.; Gabelica, V., Sequence and solvent effects on telomeric DNA bimolecular G-quadruplex folding kinetics. *J Phys Chem B* **2013**, *117* (41), 12391-401.
40. Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z., Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K⁺ solution under molecular crowding condition. *J Am Chem Soc* **2007**, *129* (36), 11185-91.
41. Heddi, B.; Phan, A. T., Structure of human telomeric DNA in crowded solution. *J Am Chem Soc* **2011**, *133* (25), 9824-33.
42. Buscaglia, R.; Miller, M. C.; Dean, W. L.; Gray, R. D.; Lane, A. N.; Trent, J. O.; Chaires, J. B., Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res* **2013**, *41* (16), 7934-46.
43. Trajkovski, M.; Endoh, T.; Tateishi-Karimata, H.; Ohyama, T.; Tanaka, S.; Plavec, J.; Sugimoto, N., Pursuing origins of (poly)ethylene glycol-induced G-quadruplex structural modulations. *Nucleic Acids Res* **2018**, *46* (8), 4301-4315.
44. Lannan, F. M.; Mamajanov, I.; Hud, N. V., Human telomere sequence DNA in water-free and high-viscosity solvents: G-quadruplex folding governed by Kramers rate theory. *J Am Chem Soc* **2012**, *134* (37), 15324-30.
45. Guedin, A.; Gros, J.; Alberti, P.; Mergny, J. L., How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res* **2010**, *38* (21), 7858-68.
46. Bonnet, G.; Krichevsky, O.; Libchaber, A., Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci U S A* **1998**, *95* (15), 8602-6.
47. Piazza, A.; Adrian, M.; Samazan, F.; Heddi, B.; Hamon, F.; Serero, A.; Lopes, J.; Teulade-Fichou, M. P.; Phan, A. T.; Nicolas, A., Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J* **2015**, *34* (12), 1718-34.
48. Prister, L. L.; Yin, S.; Cahoon, L. A.; Seifert, H. S., Altering the *Neisseria gonorrhoeae* pilE guanine quadruplex loop bases affects pilin antigenic variation. *Biochemistry* **2020**, *59* (10), 1104-1112.

TOC Illustration