Article

Real-time detection reveals responsive co-transcriptional formation of persistent intramolecular DNA and intermolecular DNA:RNA hybrid G-quadruplexes stabilized by R-loop

YANG ZHAO, Jia-yu Zhang, Zong-yu Zhang, Tan-jun Tong, Yu-hua Hao, and Zheng Tan

Just Accepted

Real-time detection reveals responsive co-transcriptional formation of persistent intramolecular DNA and intermolecular DNA:RNA hybrid G-quadruplexes stabilized by R-loop

Yang Zhao[1], Jia-yu Zhang[2], Zong-yu Zhang[1], Tan-jun Tong[1*], Yu-hua Hao[2] and Zheng Tan[2*]

[1]Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Peking University Health Science Center, Peking University Research Center on Aging, Beijing 100083, China

[2]State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, P. R. China

[*]Corresponding authors: Tan-jun Tong (Email: ttj@bjmu.edu.cn) or Zheng Tan (Email: z.tan@ioz.ac.cn)

ABSTRACT: G-quadruplex (GQ) structures are implicated in important physiological and pathological processes. Millions of GQ-forming motifs are enriched near transcription start sites (TSSs) of animal genes. Transcription can induce formation of GQs which in turn regulates transcription. The kinetics of the formation and persistence of GQs in transcription is crucial for the role they play, but has not yet been explored. We established a method based on the fluorescence resonance energy transfer (FRET) technique to monitor in real-time the co-transcriptional formation and post-transcriptional persistence of GQs in DNA. Using a T7 transcription model, we demonstrate that a representative intramolecular DNA GQ and DNA:RNA hybrid GQ promptly form in proportion to transcription activity and, once formed, maintain for hours or longer at physiological temperature even after transcription is stopped. Both their formation and persistence strongly depend on R-loop, a DNA:RNA hybrid duplex formed during transcription. Enzymatic removal of R-loop dramatically slows their formation and accelerates their unfolding. These results suggest that a transcription event is promptly read-out by GQ-forming motifs and the GQ formed can either perform a regulation in fast response to transcription and/or memorized in DNA to mediate time-delayed regulation under the control of RNA metabolism and GQ-resolving activity. Alternatively, GQs need to be timely resolved to warrant success of translocating activities such as replication. The kinetic characteristics of GQs and its connection with R-loop may have implications in transcription regulation, signal transduction, G-quadruplex processing, and genome stability.

Guanine-rich nucleic acids can fold into a four-stranded G-quadruplex (GQ) structure when four guanine-tracts (G-tract) are clustered in proximity.[1] Millions of GQ-forming motifs are present in the human genome.[2-4] The existence of GQs in both eukaryotic[5-9] and prokaryotic[10] cells has only been detected in several recent studies. GQ-forming motifs are significantly enriched near transcription start sites (TSSs),[3,11-14] and transcription has recently been shown to induce formation of GQs in duplex DNA under both in vitro[3,4,15-18] and in vivo[10] conditions in the transcribed region downstream of TSS. These GQs fall into two types, i.e. intramolecular DNA GQs (DQ) consisting of G-tracts all from the non-template DNA strand and DNA:RNA hybrid GQs (HQ) involving G-tracts from both the non-template DNA strand and RNA transcript.[4,16] GQs play role in a large variety of cellular pathways.[19] In particular, GQs formed in transcription can in turn regulate transcription[4,14] and may also be involved in replication initiation.[17,20,21]

GQ formation in genomic DNA occurs in a unique environment where DNAs are long base-paired duplex. Unlike in a single-stranded form, a GQ-forming motif in a duplex DNA is constrained at the two ends by flanking sequences and the formation of GQ, on the other hand, has to compete against the annealing with the complementary DNA strands during and after transcription. This situation does not favor GQ for-

mation. In addition, a DNA being transcribed is also in interaction with proteins such as RNA polymerase. In such an environment, the kinetics of the formation of a GQ and its persistence in physiological processes is an essential property that governs the functional role of the GQ. More specifically, how fast a GQ forms and how long it maintain determine respectively the responsiveness and duration of the effect of the GQ. Previously, detection and quantitation of transcriptional formation of GQ in DNA solely relied on end-point measurements[3,4,10,15-18] which is difficult to provide the desired kinetic information.

Here we describe a real-time method based on the FRET technique to investigate the kinetics of co-transcriptional formation and post-transcriptional persistence of GQs in duplex DNA. We studied a representative DQ and HQ, respectively, in duplex DNA transcribed by T7 phage RNA polymerases (RNAP). Both the DQ and HQ displayed fast formation and slow collapsing in transcription. Quantitative analyses allowed us to establish the dependence of the rate constant of GQ accumulation on transcription activity. We further studied how these processes were affected by RNA transcripts, either free in solution or in the form of R-loop, a transcription by-product in which a RNA transcript hybridizes with the template DNA strand.[22] Our data showed that R-loop contributed a fundamen-

tal role in the stabilization of DQ and particularly HQ. In the presence of R-loop, the two GQs promptly formed and maintained for hours or much longer at the physiological temperature. However, a removal of R-loop by RNase H significantly slowed down the formation and accelerated the collapse of both GQs. Our results also show that free RNA transcripts had a minor effect on the formation and collapse of DQ and HQ. All these results suggested that the formation and persistence of both types of GQs could be affected by RNA metabolism, even for the DQ whose assembly does not involve a participation of RNA. The prompt formation of long-lasting GQs and its connection with RNA metabolism set a physical basis for the functionality of transcriptionally formed GQs which may have implications in a number of cellular processes.

EXPERIMENTAL SECTION

Oligonucleotides. Dual-labeled oligonucleotides were purchased from Integrated DNA Technologies (IDT, U.S.A). Other oligonucleotides were purchased from Sangon Biotech (Shanghai, China). Their concentrations were determined by the absorbance at 260 nm using the OligoAnalyzer 3.1 from IDT (http://www.idtdna.com). Stock solution of double-stranded DNA (dsDNA) was prepared by heating two complementary oligonucleotides (1 μM each) at 95 °C for 5 min and then cooling down slowly to room temperature in a buffer containing 10 mM Tris-HCl (pH 7.9) and 50 mM LiCl.

In vitro transcription. DNA in Li^+ solution was diluted into transcription buffer containing 40 mM Tris-HCl (pH 7.9), 6.7 nM DNA (each strand), 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 2 mM NTPs (Fermentas, U.S.A.). Transcription was initiated by adding the indicated amount of T7 polymerase. For some experiments, transcription was terminated by an addition of competitive DNA (cpDNA, 5'-GAAATTAATACGACTCACTATA-3')[16] to a final concentration of 333 nM alone or together with 6.67 U/μl RNase T1 or 0.33 U/μl RNase H.

Fluorescence spectroscopy. Transcriptions were carried out at 37 °C in a 1.5 ml volume on a QuantaMaster 40 spectrofluorometer (PTI, USA). Emission spectra were recorded from 538 to 700 nm over time with an excitation at 520 nm.

RESULTS AND DISCUSSION

Identification of GQ structures. We prepared two duplex DNAs, 4G3 and G5G7, containing a $G_3(AG_3)_3$ and G_5AG_7 G-core respectively on the non-template strand. The $G_3(AG_3)_3$ G-core exists in the human PAX9 gene and has been shown to form an intramolecular G-quadruplex which enhances the splicing of the PAX9 gene.[23] The G_5AG_7 G-core is present in mitochondria genomes in the CSB II region. A GQ formed in the CSB II stimulates mitochondrial transcription termination and primer formation.[21] We recently found that the G_5AG_7 forms a HQ[16] which is involved in the transcription termination at the CSB II.[17]

We first tested GQ formation in the DNAs in K^+ solution using an exonuclease arrest assay in which exonuclease cleavage of DNA from the 3' end is arrested by a GQ.[24,25] Figure S1 shows a gel electrophoresis analysis of the hydrolysis products. As expected, the $G_3(AG_3)_3$ readily formed an intramolecular DNA GQ as demonstrated by a single termination band (lane 2) below the original DNA (lane 1). The G_5AG_7 also

formed a GQ when an RNA G_3 tract was supplied (lane 4). Without RNA, the G5G7 DNA was fully degraded (lane 5), indicating that the GQ formed by G_5AG_7 in the G5G7 with RNA was a HQ.

We then transcribed the two DNAs by T7 RNAP and analyzed GQ formation by the exonuclease arrest assay. The 3' end of the template DNA strand was protected by three phosphorothioate nucleotides such that only the non-template DNA strand was prone to the cleavage. In Figure S2, it can be found that without a prior transcription, the non-template strands in all DNAs were fully digested (lanes 2, 6, and 10). However, the digestion was arrested in the transcribed 4G3 and G5G7 (lanes 8 and 12, open arrow), but not in the random one (lane 4), indicating a formation of GQ in the 4G3 and G5G7 DNA. This assay also provided a rough estimation on the fraction of GQ formed which was approximately 77% in the 4G3 and 50% in the G5G7, respectively. It should be noted that the quantity derived could be underestimated because of the unfolding of the GQs during the digestion and the difficulty in avoiding over- or under-digestion with the enzyme.

The G_5AG_7 in the G5G7 is unable to form a DQ, but HQ.[16] While the transcriptional formation of HQ in the G_5AG_7 has been identified in our previous studies,[16,17] the structural nature of the GQ in 4G3 remained unclear. In theory, the $G_3(AG_3)_3$ in the 4G3 is capable of forming either a DQ or HQ because G-tracts from the RNA transcripts may jointly form a GQ with G-tracts from the DNA. To identify the structure, we transcribed the 4G3 DNA along with a 5G3 DNA and treated them by the exonucleases (Figure S2, right gel). With a $G_3(AG_3)_4$ G-core, the 5G3 DNA is able to form a DQ at two positions. The GQs in the two DNAs produced one and two termination bands, respectively (lanes 14 and 16, open arrows). The number of terminations correlated with the number of positions at which a DQ could form in the corresponding DNA (scheme under gel, black arrowhead). This result supports a formation of DQ as the major form of GQ. If HQ could effectively form, then we would expect at least two and three termination bands (half-white-half-black arrowheads), instead of one and two for the 4G3 and 5G3 DNA, respectively. For this reason, we assume that DQ was the major structure formed in the 4G3 DNA. We recently reported that the formation of a DQ on the non-template DNA strand could be triggered by an approaching RNAP without passing a G-core.[15] This fact suggests that the DNA G-core has a priority to form a DQ over a HQ before the G-core is transcribed.

The formation of GQs in the 4G3 and G5G7 DNA was verified using DMS footprinting which detects GQ assembly by a protection to the guanine residues in a GQ against chemical cleavage.[26] In Figure S3, we see that the G-tracts (brackets) in both 4G3 and G5G7 suffered much less cleavage in the transcribed than in the non-transcribed DNAs. This protection truly reflected GQ formation because it was not seen for the guanines flanking the G-cores (arrowheads).

Real-time monitoring of transcriptional GQ formation by FRET. The FRET technique has been well adopted in studying GQ in single-stranded nucleic acids.[27,28] To test its applicability in the transcription of duplex DNA, the $G_3(AG_3)_3$ and G_5AG_7 G-cores were both labeled with a Cy5 at the 5' side and a Cy3 at the 3' side as an acceptor and donor, respectively. GQ formation brings closer the two dyes and enhances

2

the energy transfer between them, simultaneously resulting in a decrease in the emission of the donor and an increase in the emission of the acceptor (Figure 1). The DNAs were transcribed in a K^+ solution by T7 RNAP. Time-based wavelength scans showed that the emission intensity of the acceptor increased over time while that of the donor decreased at the same time in both DNAs (Figures 2A and 2B). A clear isosbestic point near 630 nm supported a structural conversion in the DNAs. In contrast, little change in fluorescence was observed for a random duplex DNA without a G-core to form a GQ (Figure 2C). We used the proximity ratio, $F_A/(F_A+F_D)$, as a measure of FRET and plotted its change with time (Figure 2D). The increase in FRET in the 4G3 and G5G7 in comparison with that in the random DNA indicated that GQ formed during transcription in both the 4G3 and G5G7.
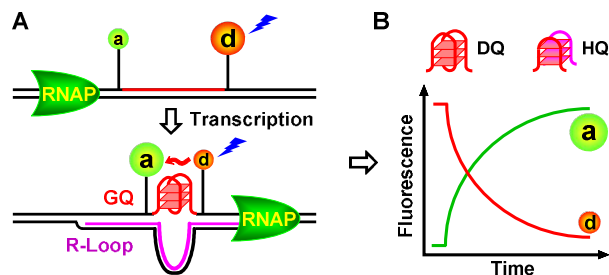
Figure 1. GQ formation in transcribed duplex DNA and detection by FRET. (A) A G-core motif (red line) is flanked by two fluorescent dyes serving as donor (d) and acceptor (a), respectively. A formation of G-quadruplex (GQ) reduces the distance between the two dyes to enhance energy transfer between them. A nascent RNA transcript may anneal with the template strand forming an R-loop. (B) The non-template DNA strand can form either an intramolecular DNA G-quadruplex (DQ) when the G-core bears four or more G-tracts or a DNA:RNA hybrid G-quadruplex (HQ) by recruiting G-tracts from RNA when the G-core has less than four G-tracts. This leads to an increase in the emission of the acceptor and a decrease in the emission of the donor.

FRET is a measure of distance between acceptor and donor. The two dyes were separated by 23 nucleotides in the DNAs. It was noted that the FRET of the G5G7 was much smaller than that of the 4G3. We previously showed that the G_7 tract in a G_5AG_7 tends to serve as two G_3 tracts with the guanine in the middle serving as a loop.[16] Theoretically, the formation of a HQ in G5G7 may involve 1-3 G-tracts from DNA and, in the meantime, 3-1 G-tracts from RNA. Therefore, the smaller FRET for the G5G7 could be explained by a contraction of fewer DNA G-tracts in the HQ than in the 4G3 which involves a contraction of four DNA G-tracts in a DQ. Due to the options of G-tract combinations, the observed FRET for the G5G7 would reflect the average of the FRET values of the corresponding HQs.

Kinetics of GQ formation. The changes in the fluorescence of both donor and acceptor allowed us to follow the kinetics of the formation or accumulation of GQ in transcription. For either the acceptor or donor in a FRET pair, the total fluorescence intensity (F) of a fluorophore was a sum of those from the folded and unfolded DNA species, which could be written as

$$F = \theta F_f + (1-\theta)F_u = \theta(F_f - F_u) + F_u \quad, \tag{1}$$

where θ was the fraction of folded species, F_f and F_u the mean fluorescence intensity when the DNAs were in the folded and unfolded form, respectively. The above equation was rewritten to give

$$\frac{F - F_u}{F_u} = \left(\frac{F_f - F_u}{F_u}\right) \cdot \theta \quad, \tag{2}$$

or simply

$$\Delta F = \Delta F_{max} \cdot \theta \quad, \tag{3}$$

where ΔF was the fractional change in fluorescence and ΔF_{max} the maximal value of ΔF. As an empirical approach, we fitted the time course of ΔF to a single exponential or dual exponential kinetics[29] and found that better results were obtained with the latter using the following equation,

$$\theta = 1 - C \cdot e^{-k_{obs1}t} - (1-C) \cdot e^{-k_{obs2}t} \quad, \tag{4}$$

where C was a fractional constant and k_{obs1} and k_{obs2} were two rate constants. For this reason, Eqs. (3) and (4) were chosen to fit our data which allowed us to obtain the constants ΔF_{max}, k_{obs1}, k_{obs2}, and C. The mean rate constant, k_{obs}, is given as

$$k_{obs} = C \cdot k_{obs1} + (1-C) \cdot k_{obs2} \quad. \tag{5}$$
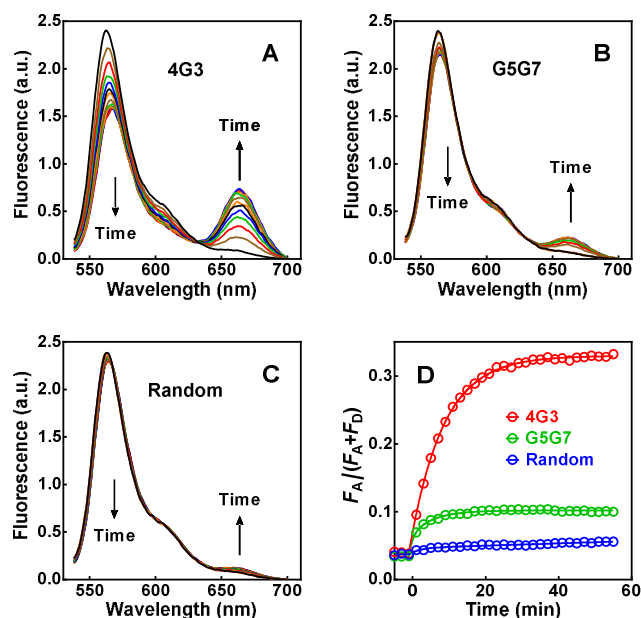
Figure 2. Detection of GQ formation in transcribed duplex DNA by FRET. (A-C) Wavelength scans over time during transcription of (A) 4G3 capable of forming a DQ; (B) G5G7 capable of forming an HQ; (C) Random DNA containing no GQ-forming sequence. (D) Proximity ratio as a function of transcription time for the DNAs in (A-C), where F_D and F_A are the fluorescence intensities of the donor (Cy3) and acceptor (Cy5), respectively.

We transcribed the 4G3 DNA at different concentrations of T7 RNAP; each produced two time courses representing respectively the changes in fluorescence of the acceptor and donor upon the formation of GQ (Figure 3A). The two time courses were simultaneously fitted to Eqs. (3) and (4) using

3

k_{obs1}, k_{obs2}, and C as global parameters. The dependence of k_{obs} on RNAP concentration was then obtained by these pair-wise fittings, which showed that the rate of DQ formation increased with the increases in RNAP concentration (Figure 3B).

Dependence of DQ and HQ formation on transcription activity. When rate constants under several transcription conditions were calculated, it would be more convenient to perform a single fitting using the fluorescence from one fluorophore instead of several pair-wise fittings. In this case, the acceptor was preferred because of its larger relative change in fluorescence (Figure 3A). Figure S4A show results of such an analysis in which all the fluorescence traces of the acceptor were simultaneously fitted to Eqs. (3) and (4) using ΔF_{max} as a global parameter. From Figure S4B, it is seen that the dependence of k_{obs} on RNAP concentration obtained was in agreement with that obtained from the pair-wise fittings (Figure S4B versus 3B). The successful extraction of the kinetic rate constant k_{obs} for the DQ in the 4G3 promoted us to perform the same analysis on the G5G7 DNA. We transcribed the G5G7 under identical conditions as the 4G3. The result (Figure 4) showed a similar RNAP concentration dependence.
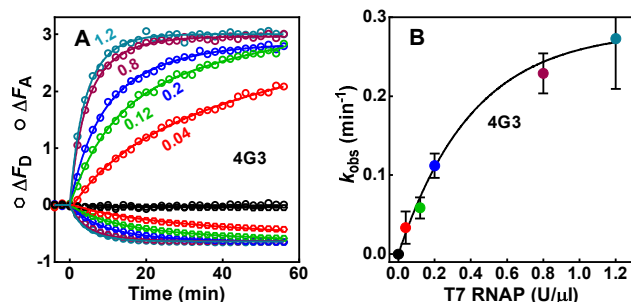
Figure 3. Kinetics of GQ formation in transcribed 4G3 DNA. (A) Fractional Changes in fluorescence intensity of acceptor (ΔF_A, open circle) and donor (ΔF_D, open diamond) during transcription. Each pair of fluorescence traces (symbols) were fitted (solid line) simultaneously to Eqs. (3) and (4) using k_{obs1}, k_{obs2}, and C as global parameters within the pair. Numbers by the curves indicate concentrations (U/µl) of T7 RNAP. (B) Observed rate constant k_{obs} (with SE) of GQ formation as a function of T7 RNAP concentration derived from (A).

Effect of R-loop on the formation kinetics of DQ and HQ. During transcription, a RNA transcript may anneal with the template DNA strand.[22] This RNA:DNA hybrid, called R-loop, forms more preferably in G/C-rich DNA due to the exceptional stability of the rG:dC base pairing.[30] R-loop has been detected in the transcription of the G5G7 DNA.[16] Apparently, R-loop resists the annealing between the two DNA strands in a duplex DNA and, as a result, is expected to stabilize both DQ and HQ. This expectation has not been studied because of the lack of proper method to monitor the formation and collapse of GQs in transcribed DNA.
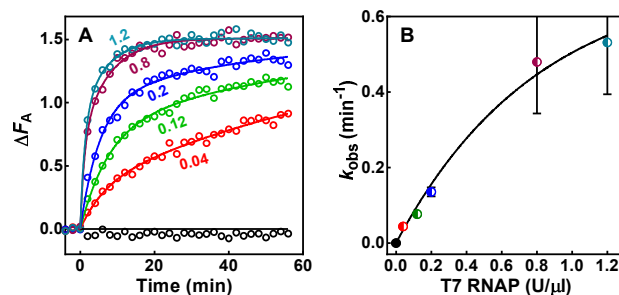
Figure 4. Kinetics of HQ formation in transcribed G5G7 DNA. (A) Changes in fluorescence intensity of acceptor (ΔF_A, circles) were simultaneously fitted (solid line) to Eq. (3) and (4) using ΔF_{max} as a global parameter. Numbers by the curves indicate concentrations (U/µl) of T7 RNAP. (B) Observed rate constant k_{obs} (with SE) of HQ formation as a function of T7 RNAP concentration derived from (A).

The FRET method allowed us to study how R-loop could affect GQ formation. We transcribed the 4G3 and G5G7 in the presence of RNase H, an enzyme that specifically degrades the RNA in R-loops, but not that in HQs.[16] The RNase H significantly reduced the rate of GQ formation in both DNAs (Figure 5, blue circles), especially in the G5G7, suggesting that the removal of R-loop destabilized both the DQ and HQ. A HQ forms via an R-loop → ssRNA → HQ cascade in which the RNA in an R-loop is displaced by the following round of transcription and this released RNA can jointly form a HQ with the non-template DNA strand.[16] Apparently, the RNase H inserted an additional effect on HQ by removing R-loop at the first step of HQ formation. This was likely why the formation of HQ in the G5G7 was reduced to a much lower level than the DQ was in the 4G3. Unlike RNase H, RNase T1, which hydrolyzes free ssRNA but not that in HQ,[16] only had a minor effect. Its inclusion in transcription slightly reduced the formation of GQ in both DNAs (Figure 5, green circles). We assume that its effect might be indirect and was possibly achieved by reducing the free RNA concentration to speed up the dissociation of R-loop that might be in slow equilibrium with ssRNA.
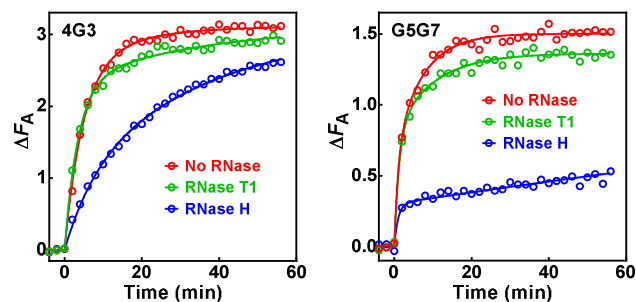
Figure 5. Kinetics of GQ formation in transcribed 4G3 and G5G7 DNA in the presence of RNA hydrolysis. Changes in fluorescence intensity of acceptor ΔF_A were plotted as a function of transcription time. GQ formation was catalyzed by T7 RNAP (0.8 U/µl) in the absence or presence of RNase T1 or H.

4

Post-transcriptional persistence of DQ and HQ and effect of R-loop. How long a GQ lasts is an important factor determining its effect in physiological processes. Many studies have been carried out regarding the stability of free GQs formed in single-stranded nucleic acids,[1,31,32] however, little has been done with those formed in transcribed duplex DNA. We examined the persistence of the DQ and HQ in the two DNAs by monitoring their collapse after transcription was terminated. In these assays, A DNA was first transcribed till the formation of GQ reached a plateau. Then excess competitive double-stranded (cpDNA) was added to terminate the transcription by capturing the RNAP molecules.[16] From this point, the persistence of GQs was monitored over time in the absence of transcription.

We found that the DQ formed in 4G3 (Figure 6A, red circles) was extremely stable that little collapse could be detected over one hour since the fluorescence remained flat after the termination of transcription. The HQ formed in the G5G7 was less stable with a slow collapse after transcription was terminated as indicated by a slow reduction in fluorescence (Figure 6B, red circles). To evaluate the effect of R-loop, we also added RNase H along with the cpDNA. Hydrolysis of R-loop accelerated the collapse of both the DQ and HQ (blue circles). Due to its lower stability than that of DQ, HQ collapsed at a much faster rate than the DQ did. Its fluorescence dropped by approximately 70% within a few minutes which was then followed by a slower decrease over time. In contrast, the DQ lasted for several tens of minutes. We also examined the effect of RNase T1. In this case, the collapse of DQ showed little change and that of HQ was moderately accelerated (green circles).
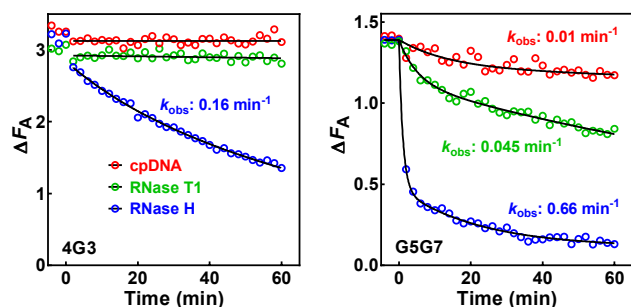
Figure 6. Post-transcription persistence of GQs in 4G3 and G5G7 DNA. Decays in fluorescence intensity of acceptor (ΔF_A, circles) were simultaneously fitted (solid line) to Eq. (3) and (4) using ΔF_{max} as a global parameter. GQ formation was catalyzed by T7 RNAP (0.8 U/μl) to reach plateau before transcription was terminated by cpDNA alone or accompanied by RNase T1 or H at which point (zero time) monitoring of fluorescence decay started. Numbers by the curves indicate rate constant k_{obs} of GQ collapse.

CONCLUSIONS

We established a method that monitors in real-time the formation or presence of GQ in transcription. Its application provides important kinetic insight into the transcriptional formation of GQs and their post-transcription persistence, as well as their connection with R-loop for the first time. The characteristics revealed of both DQ and HQ set a physical basis for their physiological function which may have implications in transcription regulation, signal transduction, G-quadruplex processing, and genome stability. As a regulatory element, the rate of GQ formation determines the responsiveness by which the element senses transcription activity to execute its function; the time a GQ remains determines how long the effect of a GQ lasts. For both the DQ and HQ we tested, their formation is fast and proportional to RNAP concentration, or in other words, dependent on transcription activity (Figures 3-4 and S4). This confers a prompt sensing of transcription and feedback for a GQ to perform regulation in response to transcription. Our data also show that the two GQs are stable along with R-loop. They are able to stay for hours if undisturbed. This fact suggests that a transcription event in cells may be memorized by stable GQs in a DNA for a period of time until they are unfolded, either spontaneously or catalyzed by GQ-resolving enzymes such as helicases.[33] In such a case, a transcription event may affect a cellular process that occurs later via a GQ at the same locus after the transcription itself has stopped, potentially creating a regulatory connection between two time points in genomic DNA.

On the other hand, the lifetime of the GQs exceeds certain processes on DNA like DNA replication. For example, the ~3,000,000 kbp human genome, which hold more than a million GQ-forming motifs (both DQs and HQs), is replicated in roughly 8 hours.[34] The 464 kbp E. coli genome, hosting nearly 3,000 GQs, is replicated in ~40 minutes.[35] We previously showed that GQs hinder the translocation of BLM helicase on DNA.[29] This is likely true for other translocating proteins. Therefore, GQs may create serious barriers for DNA replication and lead to pathological consequences if they are not timely resolved. In support of this notion, several disease-associated human helicases can bind and/or unwind different GQ structures,[36] implying an association of pathogenesis with GQ processing. One well-known example of such diseases is the Werner syndrome, a genetic disorder featuring genomic instability and premature aging, caused by deficiency in the WRN protein.[37] The WRN protein is a helicase that unwinds a number of structures, including GQs.[38,39] Its activity is required for proper replication of GQ-forming telomere DNA in human cells, indicating a necessity of active GQ-resolving for DNA replication.[40,41]

Our data revealed an essential contribution of R-loop in the reinforcement of DQ and HQ in both their formation and persistence (Figures 5 and 6). The dramatic effect of RNase H means that the two key features of GQs are subjected to a regulation by RNA metabolism. R-loop has been shown to form in the human genome[42,43] and defects in RNase H are associated with genome instability and pathological disorders.[44,45] According to our data, deficiency in RNase H is expected to dramatically enhance the formation and delay the collapse of GQs in cells. Therefore, GQ structures in transcribed genes may contribute to diseases associated with abnormal RNA metabolism.

Since our assays were conducted under in vitro conditions, the results may not precisely reflect the situation in cells. The concentration of cations, such as Mg^{2+} and K^+, is important for the stability of GQs.[46] Our assays were carried out with 10 mM Mg^{2+} which was in the same order of magnitude as that (15-25 mM) in both prokaryotic and mammalian cells.[47] The

5

50 mM K^+ was used according to the instruction of the manufacturer of the T7 RNAP to avoid inhibition at higher salt concentration. Although it was lower than the intracellular concentration of K^+ in animal (150 mM)[48] and bacteria (180-200 mM),[49] previous study showed that GQ formation in duplex DNA reaches a plateau at ~20 mM of K^+.[50] Besides the cations, molecular crowding, a reality of intracellular environment, can significantly enhance the stability of GQs,[50-53] which is expected to facilitate GQ formation. We intended to carry out measurements in the presence of PEG 200 that has been employed as a crowding agent. Unfortunately, the fluorescence was severely quenched in PEG solution although it was not in dilute solution (Figure S5). This prevented us to further investigate the effect of molecular crowding. Transcription by RNA polymerases can generate negative supercoiling in DNA in the transcribed region.[54] Negative supercoiling facilitates the formation of G-quadruplex in plasmid.[55,56] In comparison with the DNA in cells which is topologically constrained, the linear duplex is able to rotate at the time of transcription to relieve the generated superhelicity.[57] In view of all these factors, the formation of the DQ and HQ might be more robust in cells.

Recently, fluorescent GQ-binding ligands were used to indicate the formation of GQs in RNA in transcription.[58] The FRET method we described provides a simple, fast, and convenient tool to monitor GQs in DNA. It can not only follow the kinetics of co-transcriptional formation, but also the post-transcriptional persistence of GQs, together in a single assay. The analysis can be carried out in either a quantitative or qualitative manner, depending on the transcription conditions. When formation is much faster than collapse, the Eqs. (3) and (4) can be used to extract the rate constant of GQ formation/accumulation. Examples of this condition are the formation of DQ and HQ in the absence of RNases (Figures 3-4 and S4). When a formation of GQ is accompanied by a fast collapse of the structure, for instance in the presence of RNase H (Figure 5), the method can still be used to follow the changes in the amount of GQs. The collapse of GQs after the termination of transcription can always be analyzed quantitatively since it is a single-turnover reaction in the absence of other reactions (Figure 6).

For the DQ and HQ we analyzed, the fitting of GQ formation to a dual exponential kinetics may reflect a presence of multiple pathways of GQ formation. The transcriptional formation of GQs should be governed by the turnover of the enzymatic reaction as well as the response of the G-cores to fold into GQs. The rate constants we obtained for the DQ and HQ were much smaller than the turnover rate of T7 RNAP,[59] which seems to suggest the importance of the DNA sequences in determining the formation of the GQs. Sequences forming multiple GQ structures was found to fit more than one rate constants when unfolded by helicase.[29] Such heterogeneity could also be possible for the reverse reaction in which DNA molecules may fold into GQs of different conformations, leading to distinctive rate constants.

Our method used DNA internally labeled with FRET fluorophore at the two sides of a G-core motif. Internally labeled DNA with FRET fluorophores has been used in transcription with T7 and other RNA polymerase to analyze conformational changes in DNA duplex.[60-63] Such labeling did not significantly interfere with transcription.[60] Fluorophore labeling may also affect the formation of GQ. Attachment at the ends of a G-core constrains the freedom of the DNA and affects GQ folding equilibrium.[64] We expect such an effect, if any, should not be significant either in DNA duplex because the G-core is already constrained with two much bulkier flanking and one complementary DNA strands. A single molecular study revealed that GQ forms in sub-second time scale in transcribed DNA.[65] Gene expression in cells produces many copies of RNA molecules via multiple cycles of transcription events. In a single transcription event, a G-core may or may not form a GQ. Our ensemble FRET scheme follows the fractional formation of GQs in a population of DNA molecules at a larger time scale during the course of multiple cycles of transcription. The results obtained reflect the probability of formation or collapse of GQs in a gene.

ASSOCIATED CONTENT

Supporting Information

Additional experimental methods and figures S1-S5. These materials are available free of charge via the internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Email: ttj@bjmu.edu.cn or z.tan@ioz.ac.cn

REFERENCES

(1) Chen, Y.; Yang, D. *Curr Protoc Nucleic Acid Chem* **2012**, *Chapter 17*, Unit17 15.

(2) Bedrat, A.; Lacroix, L.; Mergny, J. L. *Nucleic Acids Res.* **2016**, *44*, 1746-1759.

(3) Li, X. M.; Zheng, K. W.; Zhang, J. Y.; Liu, H. H.; He, Y. D.; Yuan, B. F.; Hao, Y. H.; Tan, Z. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14581-14586.

(4) Zheng, K. W.; Xiao, S.; Liu, J. Q.; Zhang, J. Y.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 5533-5541.

(5) Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J. M.; Wong, J. M.; Monchaud, D. *J. Am. Chem. Soc.* **2015**, *137*, 8521-8525.

(6) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 75-80.

(7) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182-186.

(8) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8572-8577.

(9) Yuan, L.; Tian, T.; Chen, Y.; Yan, S.; Xing, X.; Zhang, Z.; Zhai, Q.; Xu, L.; Wang, S.; Weng, X.; Yuan, B.; Feng, Y.; Zhou, X. *Sci Rep* **2013**, *3*, 1811.

(10) Wu, R. Y.; Zheng, K. W.; Zhang, J. Y.; Hao, Y. H.; Tan, Z. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 2447-2451.

(11) Zhao, Y.; Du, Z.; Li, N. *FEBS Lett.* **2007**, *581*, 1951-1956.

(12) Eddy, J.; Vallur, A. C.; Varma, S.; Liu, H.; Reinhold, W. C.; Pommier, Y.; Maizels, N. *Nucleic Acids Res.* **2011**, *39*, 4975-4983.

(13) Zhang, C.; Liu, H. H.; Zheng, K. W.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 7144-7152.

(14) Xiao, S.; Zhang, J. Y.; Zheng, K. W.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 10379-10390.

6

(15) Liu, J. Q.; Xiao, S.; Hao, Y. H.; Tan, Z. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 8992-8996.

(16) Zhang, J. Y.; Zheng, K. W.; Xiao, S.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2014**, *136*, 1381-1390.

(17) Zheng, K. W.; Wu, R. Y.; He, Y. D.; Xiao, S.; Zhang, J. Y.; Liu, J. Q.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2014**, *42*, 10832-10844.

(18) Xiao, S.; Zhang, J. Y.; Wu, J.; Wu, R. Y.; Xia, Y.; Zheng, K. W.; Hao, Y. H.; Zhou, X.; Tan, Z. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 13110-13114.

(19) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627-8637.

(20) Cayrou, C.; Coulombe, P.; Puy, A.; Rialle, S.; Kaplan, N.; Segal, E.; Mechali, M. *Cell Cycle* **2012**, *11*, 658-667.

(21) Wanrooij, P. H.; Uhler, J. P.; Simonsson, T.; Falkenberg, M.; Gustafsson, C. M. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16072-16077.

(22) Aguilera, A.; Garcia-Muse, T. *Mol. Cell* **2012**, *46*, 115-124.

(23) Ribeiro, M. M.; Teixeira, G. S.; Martins, L.; Marques, M. R.; de Souza, A. P.; Line, S. R. *Hum. Genet.* **2015**, *134*, 37-44.

(24) Tang, J.; Kan, Z. Y.; Yao, Y.; Wang, Q.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2008**, *36*, 1200-1208.

(25) Yao, Y.; Wang, Q.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2007**, *35*, e68.

(26) Sun, D.; Hurley, L. H. *Methods Mol. Biol.* **2010**, *608*, 65-79.

(27) De Cian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teulade-Fichou, M. P.; Lacroix, L.; Mergny, J. L. *Methods* **2007**, *42*, 183-195.

(28) Juskowiak, B.; Takenaka, S. *Methods Mol. Biol.* **2006**, *335*, 311-341.

(29) Liu, J. Q.; Chen, C. Y.; Xue, Y.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2010**, *132*, 10521-10527.

(30) Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211-11216.

(31) Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat. Rev. Genet.* **2012**, *13*, 770-780.

(32) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482-5515.

(33) Mendoza, O.; Bourdoncle, A.; Boule, J. B.; Brosh, R. M., Jr.; Mergny, J. L. *Nucleic Acids Res.* **2016**, *44*, 1989-2006.

(34) Fangman, W. L.; Brewer, B. J. *Cell* **1992**, *71*, 363-366.

(35) Helmstetter, C. E. *J. Mol. Biol.* **1968**, *31*, 507-518.

(36) Maizels, N. *EMBO Rep.* **2015**, *16*, 910-922.

(37) Rossi, M. L.; Ghosh, A. K.; Bohr, V. A. *DNA Repair (Amst)* **2010**, *9*, 331-344.

(38) Mohaghegh, P.; Karow, J. K.; Brosh, R. M., Jr.; Bohr, V. A.; Hickson, I. D. *Nucleic Acids Res.* **2001**, *29*, 2843-2849.

(39) Tang, W.; Robles, A. I.; Beyer, R. P.; Gray, L. T.; Nguyen, G. H.; Oshima, J.; Maizels, N.; Harris, C. C.; Monnat, R. J., Jr. *Hum. Mol. Genet.* **2016**.

(40) Damerla, R. R.; Knickelbein, K. E.; Strutt, S.; Liu, F. J.; Wang, H.; Opresko, P. L. *Cell Cycle* **2012**, *11*, 3036-3044.

(41) Crabbe, L.; Verdun, R. E.; Haggblom, C. I.; Karlseder, J. *Science* **2004**, *306*, 1951-1953.

(42) Loomis, E. W.; Sanz, L. A.; Chedin, F.; Hagerman, P. J. *PLoS Genet.* **2014**, *10*, e1004294.

(43) Wahba, L.; Amon, J. D.; Koshland, D.; Vuica-Ross, M. *Mol. Cell* **2011**, *44*, 978-988.

(44) Chen, X.; Yang, J. R.; Zhang, J. *Genome Res.* **2016**, *26*, 50-59.

(45) Groh, M.; Gromak, N. *PLoS Genet.* **2014**, *10*, e1004630.

(46) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. *Front Chem* **2016**, *4*, 38.

(47) Moomaw, A. S.; Maguire, M. E. *Physiology (Bethesda)* **2008**, *23*, 275-285.

(48) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P.; Garland Science: New York, NY, 2002.

(49) Shabala, L.; Bowman, J.; Brown, J.; Ross, T.; McMeekin, T.; Shabala, S. *Environ. Microbiol.* **2009**, *11*, 137-148.

(50) Zheng, K. W.; Chen, Z.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2010**, *38*, 327-338.

(51) Zhou, J.; Tateishi-Karimata, H.; Mergny, J. L.; Cheng, M.; Feng, Z.; Miyoshi, D.; Sugimoto, N.; Li, C. *Biochimie* **2016**, *121*, 204-208.

(52) Jiang, H. X.; Cui, Y.; Zhao, T.; Fu, H. W.; Koirala, D.; Punnoose, J. A.; Kong, D. M.; Mao, H. *Sci Rep* **2015**, *5*, 9255.

(53) Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185-11191.

(54) Liu, L. F.; Wang, J. C. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7024-7027.

(55) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863-2874.

(56) Sun, D. *Methods Mol. Biol.* **2010**, *613*, 223-233.

(57) Harada, Y.; Ohara, O.; Takatsuki, A.; Itoh, H.; Shimamoto, N.; Kinosita, K., Jr. *Nature* **2001**, *409*, 113-115.

(58) Endoh, T.; Rode, A. B.; Takahashi, S.; Kataoka, Y.; Kuwahara, M.; Sugimoto, N. *Anal. Chem.* **2016**, *88*, 1984-1989.

(59) Skinner, G. M.; Baumann, C. G.; Quinn, D. M.; Molloy, J. E.; Hoggett, J. G. *J. Biol. Chem.* **2004**, *279*, 3239-3244.

(60) Cordes, T.; Santoso, Y.; Tomescu, A. I.; Gryte, K.; Hwang, L. C.; Camara, B.; Wigneshweraraj, S.; Kapanidis, A. N. *Biochemistry* **2010**, *49*, 9171-9180.

(61) Tang, G. Q.; Patel, S. S. *Biochemistry* **2006**, *45*, 4936-4946.

(62) Tang, G. Q.; Roy, R.; Bandwar, R. P.; Ha, T.; Patel, S. S. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 22175-22180.

(63) Tang, G. Q.; Patel, S. S. *Biochemistry* **2006**, *45*, 4947-4956.

(64) Wang, Q.; Ma, L.; Hao, Y. H.; Tan, Z. *Anal. Chem.* **2010**, *82*, 9469-9475.

(65) Shrestha, P.; Xiao, S.; Dhakal, S.; Tan, Z.; Mao, H. *Nucleic Acids Res.* **2014**, *42*, 7236-7246.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

for TOC only