

Targeting G-quadruplex Forming Sequences with Cas9

Hamza Balci,* Viktorija Globyte, and Chirlmin Joo*

Cite This: *ACS Chem. Biol.* 2021, 16, 596–603

Read Online

ACCESS |



Metrics & More

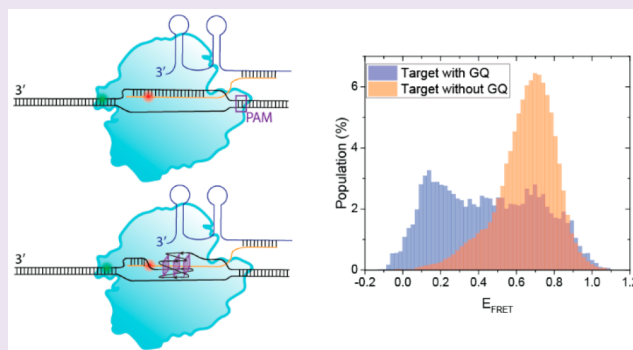


Article Recommendations



Supporting Information

ABSTRACT: Clustered regularly interspaced palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins, particularly Cas9, have provided unprecedented control on targeting and editing specific DNA sequences. If the target sequences are prone to folding into noncanonical secondary structures, such as G-quadruplex (GQ), the conformational states and activity of the CRISPR–Cas9 complex may be influenced, but the impact has not been assessed. Using single molecule FRET, we investigated structural characteristics of the complex formed by CRISPR–Cas9 and target DNA, which contains a potentially GQ forming sequence (PQS) in either the target or the nontarget strand (TS or NTS). We observed different conformational states and dynamics depending on the stability of the GQ and the position of PQS. When PQS was in NTS, we observed evidence for GQ formation for both weak and stable GQs. This is consistent with R-loop formation between TS and crRNA releasing NTS from Watson–Crick pairing and facilitating secondary structure formation in it. When PQS was in TS, R-loop formation was adequate to maintain a weak GQ in the unfolded state but not a GQ with moderate or high stability. The observed structural heterogeneity within the target dsDNA and the R-loop strongly depended on whether the PQS was in TS or NTS. We propose these variations in the complex structures to have functional implications for Cas9 activity.



INTRODUCTION

The capability to target desired DNA/RNA sequences or secondary structures with high specificity is crucial for many scientific, technological, and medical applications. Various approaches have been employed to achieve this goal including small molecules, nucleic acids, peptides, and proteins that recognize these sequences or structures. However, low specificity has consistently hindered widespread use of these approaches in complex settings, such as the mammalian cells where the large genome size or presence of similar structures demand high specificity. In this context, the repurposing of an adaptive prokaryotic immune system from *Streptococcus pyogenes* into a potent genome targeting and editing tool has been one of the most important scientific developments of recent decades. This immune system, clustered regularly interspaced short palindromic repeats (CRISPR), consists of an array of viral-derived DNA fragments (spacers) collected from previous attacks by various mobile genetic elements.^{1–3} CRISPR-associated (Cas) proteins, CRISPR-RNA (crRNA), which guides Cas proteins to the target sequence, and trans activating crRNA (tracrRNA) are other important agents of this system.

For target DNA to be cleaved, near-perfect complementarity between the spacer in crRNA and the target DNA is required. In addition, a protospacer adjacent motif (PAM) needs to be in the immediate vicinity of the target sequence.^{4,5} For Cas9 derived from *S. pyogenes*, the target sequence is 20-nt long and

the PAM sequence is NGG.⁶ Also, Cas9 effectively functions with a single guide RNA (sgRNA) which combines tracrRNA and crRNA.⁷ The capabilities of various Cas proteins and their interacting partners over a broad range of applications are still actively researched.^{8,9} However, Cas9 has been the most widely used system due to the simplicity of the CRISPR–Cas9 complex and its high specificity.¹⁰ In addition to wild-type Cas9 from different bacteria, its engineered mutants that promise higher specificity, tighter binding, and reduced or disabled cleavage activity, including endonuclease-dead Cas9 (dCas9) or enhanced nuclease activity, have been generated.^{11–13}

Despite the vast amount of information and know-how that have been accumulated on CRISPR–Cas9 in the last several years, the capabilities and limitations of this system in targeting DNA sequences that can form secondary structures have not been systematically investigated. Secondary structures such as G-quadruplex (GQ) structures, hairpins, and various loops (R-, D-, T-loops) are physiologically significant and have a high propensity to form when the double stranded DNA (dsDNA)

Received: August 25, 2020

Accepted: March 18, 2021

Published: March 26, 2021



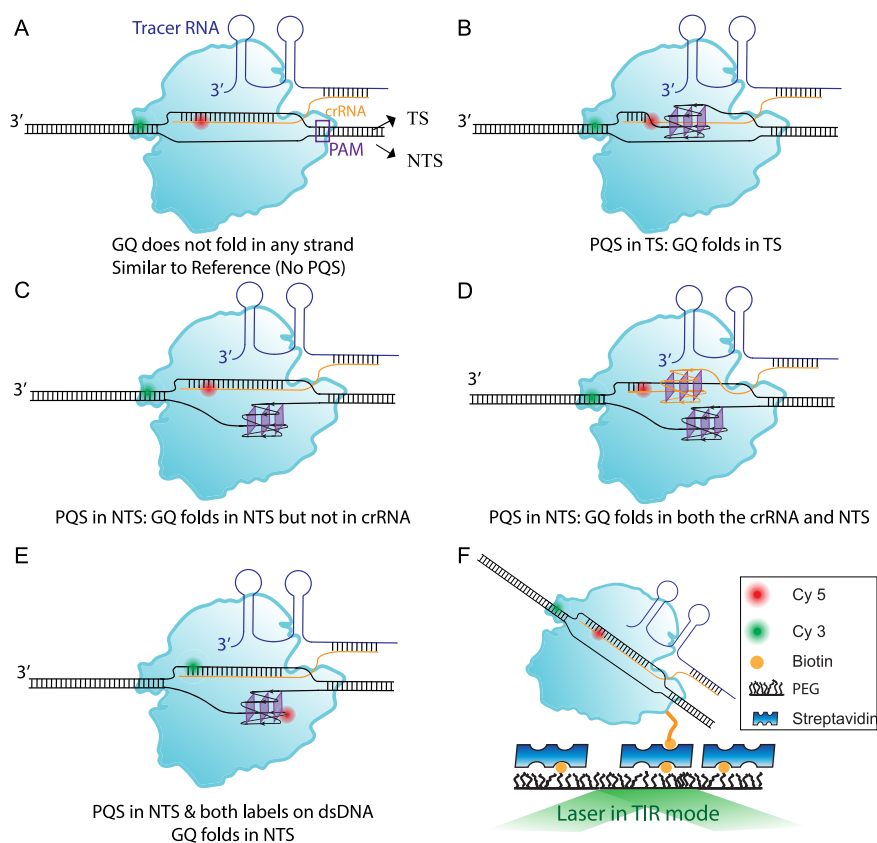


Figure 1. Schematics demonstrating different potential conformations of CRISPR–Cas9 and dsDNA complex. The PQS is in either TS or NTS. In A–D, the donor fluorophore (Cy3-green) is on TS and acceptor fluorophore (Cy5-red) is on crRNA. In E, the donor is on TS and acceptor on NTS, on different sides of PQS. When PQS is in NTS, it also must be in crRNA, which raises the possibility of GQ formation in crRNA as shown in D. For all these different conformations, GQ may or may not fold or transition between different states. (F) A schematic of slide surface and laser excitation in TIR mode for smFRET measurements. A schematic of the steps of the FRET assay are shown in Supporting Information Figure S1.

is unwound during transcription, replication, or repair. CRISPR–Cas9 also targets one of the strands of dsDNA (target strand, named TS) with crRNA, which results in an R-loop between TS and crRNA,¹⁴ while the other strand (nontarget strand, named NTS) is released and may fold into alternative secondary structures. In addition, the complementarity between TS and NTS needs to be broken for crRNA and TS to hybridize. During this transition, TS with an appropriate sequence might also transiently fold into alternative structures, especially if crRNA and TS are not perfectly complementary. Such transient or persistent secondary structures could influence target recognition,¹⁵ binding stability, conformational dynamics, and cleavage activity of CRISPR–Cas9. Considering the abundance of sequences that could form secondary structures, it is critical to understand how they influence CRISPR–Cas9 complex structure and function.

GQs form in guanine-rich regions of the genome¹⁶ and are characterized by stacked G-tetrad layers, which contain a guanine (G) in each corner. G-tetrads are stabilized by Hoogsteen hydrogen bonds between the guanines and monovalent cations that intercalate between them. In addition to numerous demonstrations of their formation,^{17,18} both DNA and RNA GQs have been visualized in human cells and were shown to be modulated during the cell cycle.^{19,20} GQs are generally thermally more stable than the corresponding dsDNA formed by Watson–Crick base pairing²¹ and require protein activity to be destabilized and unfolded.^{22–25} The

inability to unfold or destabilize GQs impedes replication machinery and results in elevated levels of DNA breaks and genomic instability.^{26–28} The most prominent sites for potentially GQ forming sequences (PQS) are the telomeric overhangs and genomic regions involved in transcription or translation level gene expression regulation. The 3' telomeric overhang in human cells has ~200-nucleotide (nt)-long GGGTTA repeats, which form multiple 3-layered GQs. Unless unfolded, these GQs prevent telomere elongation by inhibiting telomerase,²⁹ making them prominent drug targets in cancer therapy.³⁰ In addition to telomeres, genome-wide computational studies and high-throughput sequencing have identified several hundred thousand PQSs in the human genome.^{31,32} The PQSs are significantly enriched in the human genome compared to *S. cerevisiae*, ~40-fold when normalized with respect to genome size, suggesting a functional significance in higher organisms.³³ While the coding regions are in general poor in PQS, promoters, especially the immediate vicinity of transcription start site (TSS), are rich in PQS, suggesting a role in transcription level regulation of gene expression.³⁴ About 50% of the human genes contain a PQS within 1000 nts upstream of TSS,³² and interestingly PQSs are more prevalent in promoters of oncogenes and regulatory genes, such as transcription factors, compared to housekeeping genes.³⁵

The thermodynamic stability of GQs can be greatly modulated while keeping the overall length of the sequence similar and within the ~20 nt target sequence of Cas9 by modulating the loop length, loop organization, or the number

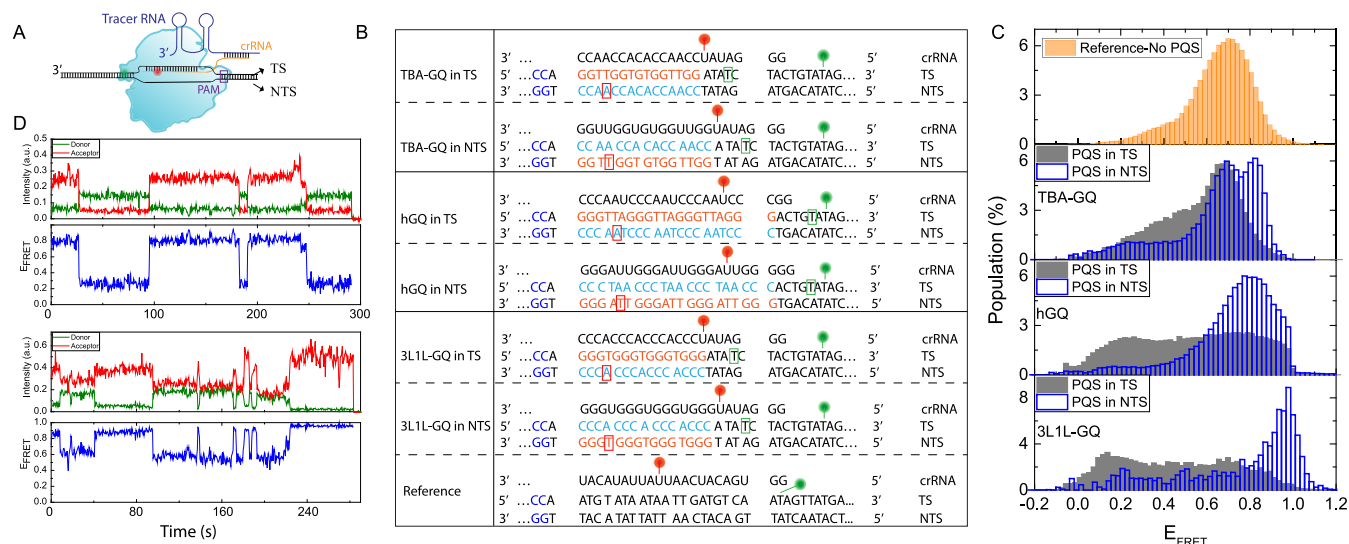


Figure 2. (A) A schematic of the complex. Donor Cy3 (green ball) is on TS, and acceptor Cy5 is on crRNA. (B) Schematics of TS, NTS, and crRNA sequences and positions of Cy3 and Cy5 (green and red balls). The PAM sequence is indicated with blue fonts, PQS with orange fonts, and complementary C-rich sequence with cyan fonts. The labeling positions for Figures 3 and 4 are indicated with green (Cy3) and red (Cy5) rectangles. (C) smFRET histograms for the reference sample that does not contain a PQS (top panel), TBS-GQ (second panel), hGQ (third panel), and 3L1L-GQ (bottom panel). All experiments were performed in KCl. [PQS in TS] data are shown with gray filled columns, while [PQS in NTS] data are shown with blue empty columns. The contrast between [PQS in TS] and [PQS in NTS] cases is particularly prominent for the 3L1L-GQ construct (bottom). The numbers of molecules in each histogram are $N = 602$ for [TBA-GQ in TS] and $N = 603$ for [TBA-GQ in NTS]; $N = 1237$ for [hGQ in TS] and $N = 539$ for [hGQ in NTS]; $N = 582$ for [3L1L-GQ in TS] and $N = 29$ for [3L1L-GQ in NTS]; and $N = 461$ for the Reference-No PQS sample. (D) Example smFRET traces demonstrating dynamics in the [TBA-GQ in NTS] construct.

of G-tetrad layers.^{21,36} In general, longer loops reduce GQ stability while more G-tetrad layers increase it. To illustrate, the PQS commonly used as a thrombin binding aptamer (TBA-GQ: GGTGGTGTGGTTGG)³⁷ and the PQS that is used as an HIV integrase inhibitor (3L1L-GQ: GGGTGGGTGGGTGGG)^{38,39} are both only 15-nt long but have thermal melting temperatures (T_m) that differ by >45 °C under physiological ion and pH conditions ($T_m = 51$ °C for TBA-GQ while $T_m > 95$ °C for 3L1L-GQ). TBA-GQ has only two G-tetrad layers and relatively long loops while 3L1L-GQ has three G-tetrad layers and 1-nt loops. We and others have demonstrated that such variations in structure influence the stability of GQs against ssDNA binding proteins and helicases.^{18,24,40,41} TBA-GQ and 3L1L-GQ will be used in this study as representatives of low and high stability GQ, respectively. In addition, the GQ formed by human telomeric sequence, hGQ, GGGTTAGGGTTAGGGTTAGGG, will serve as the GQ with intermediate stability ($T_m = 68$ °C).⁴⁰ Using smFRET, we investigated complex formation and dynamic interactions between CRISPR–Cas9 and target dsDNA that contains one of these PQSs in either TS or NTS.

MATERIALS AND METHODS

DNA/RNA Constructs. Table S1 lists sequences of the DNA/RNA constructs and locations of Cy3/Cy5. The DNA oligonucleotides were purchased from Ella Biotech GmbH (Germany). RNA oligos were purchased from Dharmacon-USA (now part of Horizon Discovery, UK) or IBA GmbH (Germany). All oligonucleotides were HPLC purified by vendors. Labeling with Cy3/Cy5 was performed in the lab using a previously published protocol.⁴² The double stranded DNA constructs were created by annealing TS and NTS at 95 °C for 5 min followed by slow cooling to RT at 0.2 °C/min.

Figure 1 shows schematics of constructs that have a PQS in either TS or NTS. To clarify the writing, these cases will be referred to as “[PQS in TS]” or “[PQS in NTS],” and in certain cases PQS will be replaced with the name of GQ construct, such as “[hGQ in TS]” to

describe having hGQ sequence in the target strand. In one of the labeling schemes, donor (Cy3) was placed on TS while the acceptor (Cy5) was on crRNA (Figure 1A–D). The fluorophore positions are kept at consistent separations for all DNA constructs with PQS and the reference construct that does not include a PQS. This arrangement enabled probing the complex between TS and crRNA. Following a smFRET assay (Supporting Information Figure S1), several different sites on TS were tested to identify a location for Cy3 such that the main FRET peak is in $E_{\text{FRET}} \approx 0.6$ – 0.7 range, which should make it sensitive to conformational changes in the complex. Sites several nts up or downstream of the optimal location selected for this study resulted in either a very high or very low FRET where this sensitivity was lost (Supporting Information Figure S2). Depending on whether the PQS is in TS or NTS and whether the GQ forms in either strand or the crRNA, different complex conformations are possible (Figure 1A–D), which should result in different FRET levels. In the second labeling scheme, the Cy3 is placed on TS while Cy5 is placed on NTS within the loop region of PQS (Figure 1E).

SmFRET Assay and Setup. The protocols for purification and biotinylation of *Streptococcus pyogenes* Cas9 (SpCas9), labeling of nucleic acids with Cy3/Cy5, sample preparation for smFRET assay, and data acquisition and analysis have been detailed in earlier studies by Globyte et al.^{43,44} Briefly, quartz slides and glass coverslips were cleaned and coated with a larger (5000 Da, 97.5% PEG + 2.5% biotin-PEG) and smaller (333 Da) polyethylene glycol (PEG) molecules following a published double PEGylation protocol. After forming the microfluidic channels, the surfaces were also treated with 5% Tween-20 to reduce nonspecific binding. To activate the CRISPR–Cas9 complex, biotin-SpCas9 (1 nM), crRNA (2 nM), and tracrRNA (12 nM) were mixed in Buffer A (100 mM KCl, 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, and 1 mM DTT) and incubated at 37 °C for 20 min. During this incubation, 0.1 mg mL⁻¹ streptavidin was added to the microfluidic channel. After 2 min of incubation, the excess streptavidin was washed away. At the end of the 20 min activation, the CRISPR–Cas9 complex was diluted two times in Buffer A and introduced to the microfluidic channel. After 2 min of incubation, the channel was washed with imaging buffer (50 mM Tris-HCl (pH 7.5), 150 mM KCl, 0.8% w/v glucose, 2 mM MgCl₂, 1 mM Trolox, 1 mg

mL^{-1} glucose oxidase [Sigma], 170 $\mu\text{g}/\text{mL}$ catalase [Merck]). In the case of constructs with the first labeling scheme where Cy5 is attached to crRNA, Cy5 molecules were excited with a red laser, and the surface density of bound CRISPR–Cas9 complexes was confirmed. This initial red-excitation step was not performed for constructs that did not have a fluorophore on crRNA. Then, target dsDNA (8 nM) was introduced to the channel and image acquisition started. A schematic summarizing these steps of the smFRET assay is shown in Supporting Information Figure S1. For measurements performed in LiCl, KCl was replaced with equimolar LiCl in Buffer A, the imaging buffer, and all buffers used to dilute biotin–SpCas9 to ensure that oligos containing PQS are not exposed to any significant concentration of KCl as this facilitates GQ formation.

A custom-built prism-type TIRF instrument was used to collect smFRET data. Short (15 frames) and long (500–2000 frames) movies were collected at 100–300 ms integration time. An Olympus IX-71 microscope equipped with a 60 \times (NA 1.20) water-immersion objective (Olympus) and an EMCCD camera (Andor Ixon Ultra) formed the main components of the instrument. The donor fluorophores were excited with a 532 nm diode laser. A 635 nm laser was used to directly excite the acceptor molecules. All histograms were created by trace-by-trace analysis where data on each molecule were inspected and background subtracted. The numbers of molecules are given in figure captions and were on average a few hundred per histogram, except the [3L1L-GQ in NTS] construct (Figure 2), where complications in complex formation resulted in a significantly smaller number of molecules. This might be due to the inability to prevent formation of a very stable GQ in crRNA before the activation step.

RESULTS AND DISCUSSION

Figure 2 shows smFRET data on GQ constructs that have a TBA-GQ, hGQ, or 3L1L-GQ in either TS or NTS, in addition to a reference construct that does not contain a PQS in either strand. The donor was on TS and the acceptor on crRNA as shown in Figure 2A, where GQ formation is not shown for simplicity. The separation between donor and acceptor was the same for all constructs except for the [hGQ in TS] case, which was one bp different from the others. The corresponding separation in the reference sample was also similar to that in PQS constructs, so FRET histograms can be directly compared. Figure 2B shows a schematic of labeling positions.

The data on [TBA-GQ in NTS] are different from those for [TBA-GQ in TS] or the reference sample (Figure 2C). If a GQ does not form in NTS, the FRET distributions for [PQS in TS], [PQS in NTS], and the reference sample should be very similar since fluorophore positions are the same. Therefore, the observed difference can be attributed to GQ formation. This suggests that even TBA-GQ, the weakest of the GQs, can fold and modify the complex structure if it is in NTS. Since NTS is freed from Watson–Crick pairing by R-loop formation between TS and crRNA, GQ may more readily form in NTS. The data for [TBA-GQ in NTS] show two prominent peaks, which might be due to different folding states of GQ. The higher FRET peak would be consistent with the folded GQ state since this peak is not observed in the reference construct. The data on [hGQ in NTS] and [3L1L-GQ in NTS] support this interpretation as they show a systematic transition to higher FRET levels as the GQ gets more stable (blue histograms in Figure 2C). While FRET distribution for [hGQ in NTS] is broad, suggesting structural heterogeneity, the higher FRET states are clearly more populated compared to [TBA-GQ in NTS]. The data on [3L1L-GQ in NTS] show a single high-FRET peak, in agreement with folding of this very stable GQ in NTS. In all the studied cases, the complexes

demonstrate significant dynamics, as exemplified in single molecule time traces in Figure 2C.

The [TBA-GQ in TS] data show a similar FRET distribution to that of the reference sample (top two panels in Figure 2C). This suggests that R-loop formation between TS and crRNA prevents GQ formation in TS for this weak GQ. The [hGQ in TS] and [3L1L-GQ in TS] data (Figure 2C, third and fourth panels) are significantly different from those on reference and TBA-GQ constructs. The distributions for both hGQ and 3L1L-GQ are much broader than all other distributions. This suggests that the stability of the CRISPR–Cas9 complex is significantly lower when a moderate to high stability GQ is placed in TS, where R-loop formation must compete with folding of the GQ. Further supporting this observation, the lower FRET states become more populated as the stability of GQ is increased for [PQS in TS] cases, an opposite trend to [PQS in NTS] cases. Considering the fluorophores are on TS and crRNA, the transition to lower FRET states might be due to displacement of crRNA from the complex or an overall distortion in the CRISPR–Cas9 complex because of GQ formation in TS. These data demonstrate that there are limitations on the type and stability of secondary structures that can be maintained in an unfolded state by CRISPR–Cas9 and the R loop.

It is known that different monovalent cations stabilize the GQ at different levels: while K^+ is very effective, Li^+ is a weak stabilizer.²¹ This property allows modulating the stability of GQ without changing the overall ionic strength of the environment. For the reference sample which does not contain a PQS, LiCl and KCl give rise to very similar smFRET distributions (Supporting Information Figure S3). However, the distributions for the high stability 3L1L-GQ construct are significantly different in KCl vs LiCl (Supporting Information Figure S4). For the [3L1L-GQ in TS] case, the smFRET distribution is very broad in the presence of KCl (Figure 2C), which was interpreted as the GQ preventing a stable R-loop formation. In the presence of LiCl, the distribution is more concentrated at higher FRET levels (Supporting Information Figure S4), suggesting that R-loop formation is inhibited to a lesser extent by a lower stability GQ. For the [3L1L-GQ in NTS] case, the distribution was dominated by a single high-FRET peak in the presence of KCl (Figure 2C), which was interpreted as the conformation space being dominated by GQ formation in NTS. In the presence of LiCl, the distribution is not dominated by a single peak, but multiple conformations are present (Supporting Information Figure S4), as would be expected from a lower stability GQ. To illustrate, the distribution for [3L1L-GQ in NTS] in the presence of LiCl resembles the distribution for [TBA-GQ in NTS] in the presence of KCl.

Previous studies have demonstrated that the HNH domain, an endonuclease domain that cleaves TS (Supporting Information Figure S5), samples multiple conformations before docking into the cleavage-active state and that divalent cations, such as Mg^{2+} , are required for Cas9 to remain in this conformation.⁴⁵ To test the potential impact of such conformational changes, we performed studies in the absence and presence (2 mM) of MgCl_2 (Supporting Information Figure S5), which demonstrated very similar FRET distributions, suggesting HNH domain conformations are not the dominant factor for the broad distributions.

The hGQ and 3L1L-GQ constructs contain repeating sequences of GGGTTA and GGGT, respectively. This

symmetry raises the question of whether hybridization of crRNA with 1–3 of such repeats, instead of full hybridization with four repeats, could be the cause of some of the broadening in the FRET histograms. Several arguments can be made against this possibility. Some of the sharpest histograms were observed for [hGQ in NTS] and [3L1L-GQ in NTS]. Partial hybridization between TS and crRNA should have resulted in broad histograms for these constructs as well. Second, the symmetric PQS in the 3L1L-GQ construct is 15-nt long, and the last 5 nt of the target sequence in PAM-distal region break the symmetry between the ends. If the first GGGT repeat from the PAM-proximal region is skipped, the crRNA will not be able to hybridize with these 5-nt at the PAM-distal region either, resulting in an 11-bp-long R-loop. Skipping two GGGT repeats results in a 7-bp R-loop. Both 7-bp- and 11-bp-long R-loops would be significantly less stable than the 20-bp R-loop of full hybridization case. Following the same arguments for hGQ constructs, skipping one or two GGGTTA repeats will result in 14-bp- and 8-bp-long R-loops. Given these, we would have expected the competition between partial and full hybridization should have been more dominant in hGQ compared to 3L1L-GQ; however, the histograms of the latter are broader than the former, arguing against this scenario.

To monitor the GQ folding state and dynamics, we moved the fluorophores to TS and NTS on different sides of PQS, as shown in Figure 3A. We initially tested placing both the donor and acceptor outside of PQS; however, this resulted in distributions peaked at very low FRET levels (Supporting Information Figure S6), which is not ideal for detecting different structural features. Therefore, the acceptor was moved within the last loop of each PQS (Figure 2B and Table S1), which should maintain structural symmetry between different PQSs. However, as hGQ (21 nt) is longer than TBA-GQ and 3L1L-GQ (both 15 nt), the separation between donor–acceptor fluorophores was greater for hGQ compared to the others (22 bp vs 16 bp). To establish the FRET levels in these dual-labeled dsDNA constructs before they are targeted by CRISPR–Cas9, we created biotinylated versions of these dual labeled DNA constructs and immobilized them on the surface. These measurements showed low FRET peaks in both KCl and LiCl consistent with fully formed duplex DNA, i.e., unfolded GQ, before it is targeted by CRISPR–Cas9 (Supporting Information Figure S7). The peak positions for each construct were very similar in KCl and LiCl, further supporting the absence of GQ before the construct is targeted by CRISPR–Cas9.

Figure 3 demonstrates comparative studies in KCl and LiCl for this fluorophore arrangement after the dsDNA is targeted by CRISPR–Cas9 (Figure 3). Similar to the measurements in Figure 2, the CRISPR complex with biotinylated Cas9 was immobilized on the surface, and dual-labeled target dsDNA (not biotinylated) was introduced in the chamber. While the distributions for KCl and LiCl are similar for [3L1L-GQ in TS] (Figure 3B), there are significant differences for [3L1L-GQ in NTS] (Figure 3C). In Figure 3C, a major high FRET peak ($E_{\text{FRET}} \approx 0.9$) is present in KCl but not LiCl, while the population of a low FRET peak ($E_{\text{FRET}} \approx 0.4$) is significantly higher in LiCl. More prominent GQ formation in KCl compared to LiCl would suggest that the high FRET peak is due to GQ formation, while the low FRET peak is due to unfolding of the GQ.

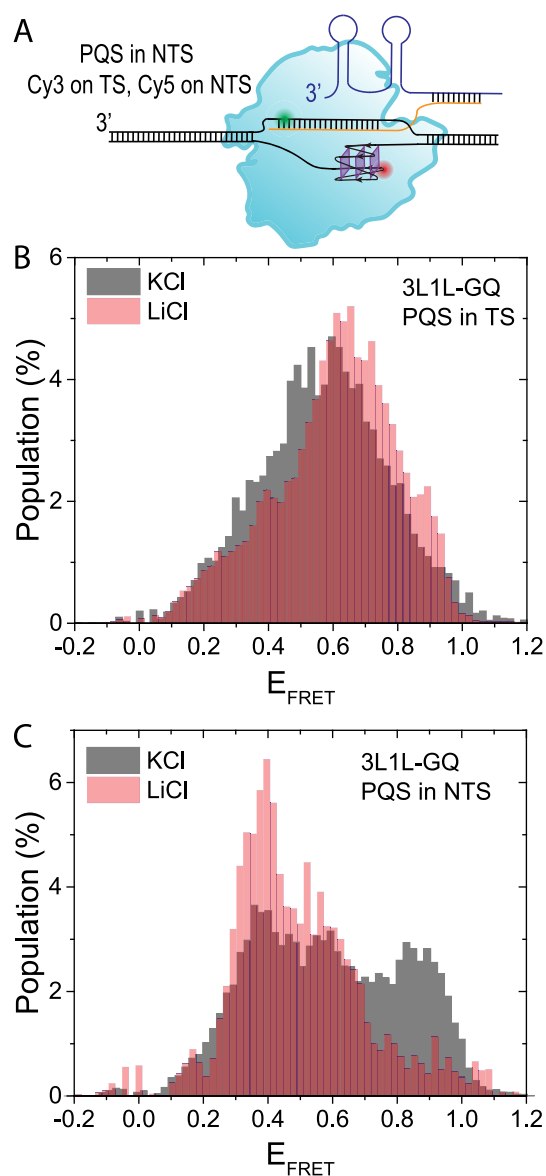


Figure 3. 3L1L-GQ is targeted by biotin-Cas9 while the GQ stability is modulated by maintaining KCl or LiCl in the environment. (A) Schematic showing the labeling scheme where the donor is on TS and the acceptor on NTS, on opposite sides of PQS, making FRET sensitive to GQ conformational states. Linearized DNA and crRNA constructs showing fluorophore positions are shown in Figure 2B (red and green rectangles). (B) For [PQS in TS], the distributions for KCl and LiCl are similar. (C) For [PQS in NTS], a high FRET peak ($E_{\text{FRET}} = 0.9$) that appears in KCl is absent in LiCl data, while the population of the low FRET peak ($E_{\text{FRET}} = 0.4$) is significantly higher. These data are consistent with the formation of a more stable GQ in KCl, resulting in the higher FRET peak, while unfolded GQ gives rise to the lower FRET peak. The number of molecules in each histogram are as follows: $N = 179$ for [PQS in TS] and $N = 189$ for [PQS in NTS] for KCl data, while $N = 301$ for [PQS in TS] and $N = 36$ for [PQS in NTS] for LiCl data.

The similarity of distributions in KCl and LiCl for [3L1L-GQ in TS] suggests that the variations in GQ stability are not adequate to result in significantly different structures for this case. Interestingly, the distribution for [3L1L-GQ in TS] is significantly narrower when both fluorophores are on the dsDNA (Figure 3B) compared to the case when one is on TS and the other on crRNA (Figure 2C, bottom panel). This

would suggest that in the presence of a very stable GQ in TS, the R-loop between crRNA and TS is very dynamic and explores many conformations while the target region in dsDNA is to a certain extent immune to these dynamics and explores fewer conformations.

Figure 4 shows comparative data in KCl on all three GQ constructs using the arrangement of fluorophores where donor and acceptor are on TS and NTS, respectively (see Figure 3A for a schematic), and CRISPR–Cas9 is immobilized on the surface. For all studied constructs, the FRET levels were significantly higher than those observed before the dsDNA was targeted by CRISPR–Cas9 (Supporting Information Figure S7). Since their donor–acceptor separations are the same (16 bp), the distributions for TBA-GQ (top panel of Figure 4A) and 3L1L-GQ (bottom panel of Figure 4A) can be directly compared. Despite some variation in their spread, the distributions for [3L1L-GQ in TS] and [TBA-GQ in TS] are surprisingly similar. This contrasts with the significant difference between the two cases in Figure 2 where the fluorophores are on TS and crRNA and FRET is sensitive to conformational variations between these two strands. The same conclusions are valid for the hGQ construct, i.e., the distribution for [hGQ in TS] is significantly narrower in Figure 4 compared to that in Figure 2. These suggest that having a moderate to high stability GQ in TS primarily results in structural instability between crRNA and TS, possibly within the R-loop, while the target dsDNA maintains a more stable structure.

For [TBA-GQ in NTS], two clearly distinguishable peaks are observed, which might be representing different folding states of GQ. The distribution for [3L1L-GQ in NTS] shows multiple peaks and a broader distribution, suggesting a structurally more heterogeneous system. This is again in contrast to the distribution in Figure 2, where a single high-FRET peak was observed for [3L1L-GQ in NTS]. The same conclusion is also valid for [hGQ in NTS] where the distribution in Figure 4 is significantly broader than that for the same case in Figure 2. These suggest that having a moderate to high stability GQ in NTS results in significant structural heterogeneity within the target dsDNA while the complex between TS and crRNA remains relatively stable. These conclusions are also supported by the broader distributions observed for all [PQS in NTS] cases compared to [PQS in TS] cases in Figure 4, i.e., having a PQS in NTS results in greater structural heterogeneity within the target dsDNA compared to having the same PQS in TS. In all cases we studied, the smFRET traces are dynamic and demonstrate frequent transitions between different FRET levels (Figure 4B).

CONCLUSIONS

We demonstrate that the position of PQS and stability of the GQ influence the conformations and structural heterogeneities experienced by the CRISPR–Cas9 complex. In the first set of measurements, the donor and acceptor were placed on TS and crRNA, which made FRET sensitive to conformational changes between TS and crRNA, i.e., the R-loop. The [PQS in TS] case for a weak GQ results in similar conformations to a reference construct that does not contain a PQS, which suggests that this weak GQ can be destabilized by R-loop formation. However, when the same PQS is placed in NTS, the resulting conformations are significantly different from those observed for the reference construct. This suggests that when

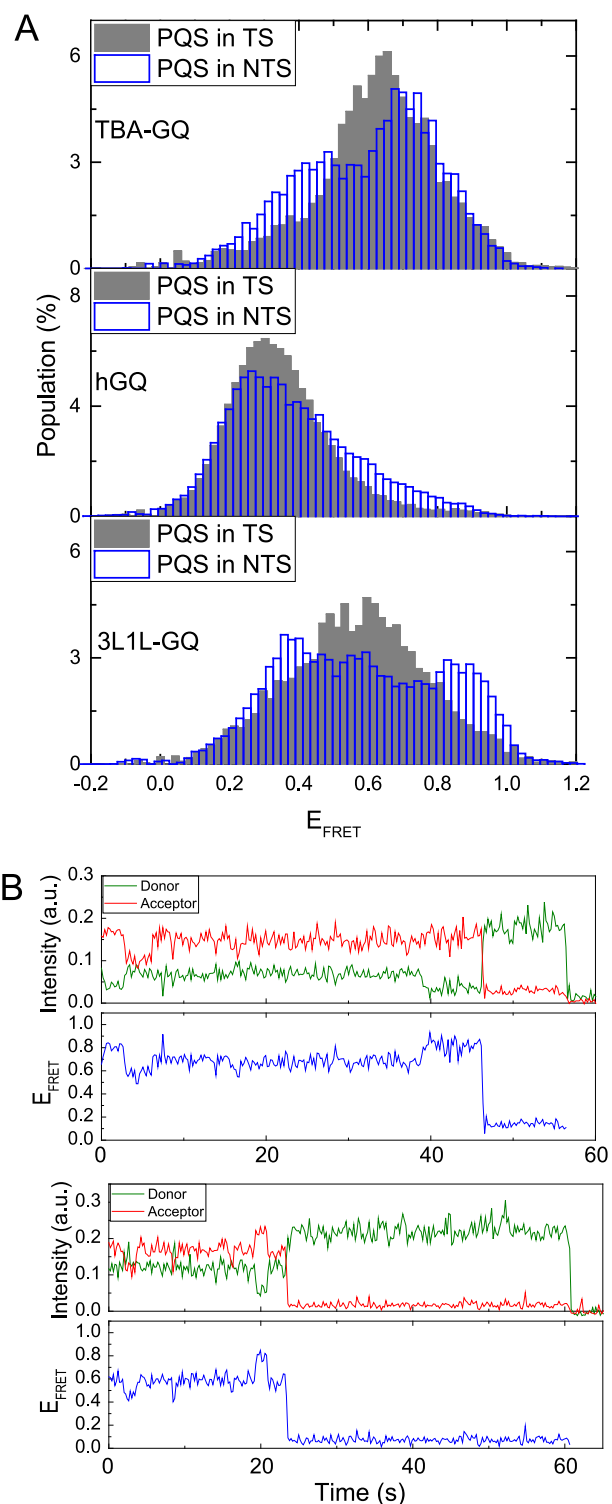


Figure 4. (A) SmFRET histograms for TBA-GQ (top), hGQ (middle), and 3L1L-GQ (bottom) when PQS is in TS (gray filled bins) or NTS (blue empty bins). All experiments were performed in KCl. Cy3 was on TS and Cy5 on NTS, and the fluorophore positions are shown by red and green rectangles in Figure 2B. The numbers of molecules in histograms were as follows: $N = 201$ for [TBA-GQ in TS] and $N = 263$ for [TBA-GQ in NTS]; $N = 511$ for [hGQ in TS] and $N = 453$ for [hGQ in NTS]; $N = 179$ for [3L1L-GQ in TS] and $N = 189$ for [3L1L-GQ in NTS]. (B) Example smFRET traces demonstrating dynamics in [TBA-GQ in TS] (top) and [3L1L-GQ in TS] (bottom) constructs.

in NTS, even a weak GQ can fold within the CRISPR–Cas9 complex and gives rise to variations in the complex structure. A GQ with higher stability creates a significant disturbance in the complex structure even when in TS, suggesting R-loop formation is not adequate to maintain such structures in the unfolded state. The broader FRET histograms for [hGQ in TS] and [3L1L-GQ in TS] cases shown in Figure 2, suggesting more heterogeneous complex structures, are a clear manifestation of this. We also observed persistent dynamics within the CRISPR–Cas9 complex during these interactions.

In the second set of measurements, the donor was on TS and the acceptor on NTS, on opposite sides of PQS, which made FRET more sensitive to conformational changes due to GQ folding dynamics. In these cases, [PQS in NTS] cases showed more heterogeneous structures compared to [PQS in TS] cases for all constructs. This suggests that [PQS in NTS] creates a more significant disturbance for the target dsDNA structure compared to [PQS in TS], which may be justified by the latter having to compete with the R-loop while the former is relatively less inhibited to attain alternative secondary structures.

These observations were made possible by optimizing the positions of donor/acceptor fluorophores on TS, NTS, and crRNA. Having established these structural and dynamic variations introduced by PQS, it will be critical to understand how they impact CRISPR–Cas9 activity in terms of target recognition, R-loop progression and stability, and target dsDNA cleavage. The understanding attained for GQs will likely have implications for other secondary structures that might form within the sequences targeted by CRISPR–Cas9.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acschembio.0c00687>.

Sequences of DNA and RNA constructs, schematic protocol for FRET assay, data for different labeling positions, data testing impact of MgCl₂ or LiCl (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Hamza Balci – Department of Physics, Kent State University, Kent, Ohio 44242, United States; Kavli Institute of NanoScience and Department of BioNanoScience, Delft University of Technology, Delft 2629HZ, The Netherlands; orcid.org/0000-0002-0273-4871; Email: hbaldi@kent.edu

Chirlmin Joo – Kavli Institute of NanoScience and Department of BioNanoScience, Delft University of Technology, Delft 2629HZ, The Netherlands; orcid.org/0000-0003-2803-0335; Email: c.joo@tudelft.nl

Author

Viktorija Globyte – Kavli Institute of NanoScience and Department of BioNanoScience, Delft University of Technology, Delft 2629HZ, The Netherlands

Complete contact information is available at: <https://pubs.acs.org/10.1021/acschembio.0c00687>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

H.B. was supported by a Marie-Curie Fellowship from the European Commission (795567). C.J. was supported by Vidi (864.14.002) of The Netherlands Organization for Scientific research and an ERC Consolidator Grant (819299) of the European Research Council.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank L. Loeff, T. Cui, M. Filius, and other members of the Joo lab for their help at different stages of the work.

■ REFERENCES

- (1) Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315 (5819), 1709–1712.
- (2) Brouns, S. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuys, R. J.; Snijders, A. P.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321 (5891), 960–964.
- (3) Mojica, F. J. M.; Juez, G.; and Rodriguezvalera, F. (1993) Transcription at Different Salinities of Haloferax-Mediterranei Sequences Adjacent to Partially Modified PstI Sites. *Mol. Microbiol.* 9 (3), 613–621.
- (4) Deveau, H.; Barrangou, R.; Garneau, J. E.; Labonte, J.; Fremaux, C.; Boyaval, P.; Romero, D. A.; Horvath, P.; and Moineau, S. (2008) Phage response to CRISPR-Encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190 (4), 1390–1400.
- (5) Horvath, P.; Romero, D. A.; Coute-Monvoisin, A. C.; Richards, M.; Deveau, H.; Moineau, S.; Boyaval, P.; Fremaux, C.; and Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190 (4), 1401–1412.
- (6) Anders, C.; Niewoehner, O.; Duerst, A.; and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513 (7519), 569.
- (7) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337 (6096), 816–821.
- (8) Knott, G. J.; and Doudna, J. A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science* 361 (6405), 866–869.
- (9) Zhang, F. (2019) Development of CRISPR-Cas systems for genome editing and beyond. *Q. Rev. Biophys.* 52, e6.
- (10) Jiang, F. G.; and Doudna, J. A. (2017) CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46 (46), 505–529.
- (11) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; and Doudna, J. A. (2017) Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* 550 (7676), 407–410.
- (12) Dominguez, A. A.; Lim, W. A.; and Qi, L. S. (2016) Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* 17 (1), 5–15.
- (13) Konermann, S.; Brigham, M. D.; Trevino, A. E.; Joung, J.; Abudayyeh, O. O.; Barcena, C.; Hsu, P. D.; Habib, N.; Gootenberg, J. S.; Nishimasu, H.; Nureki, O.; and Zhang, F. (2015) Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* 517 (7536), 583–588.
- (14) Jiang, F.; Taylor, D. W.; Chen, J. S.; Kornfeld, J. E.; Zhou, K.; Thompson, A. J.; Nogales, E.; and Doudna, J. A. (2016) Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* 351 (6275), 867–871.

- (15) Santiago-Frangos, A., Wiegand, T., and Wiedenheft, B. (2019) Cas9 slide-and-seek for phage defense and genome engineering. *EMBO J.* 38 (4), e101474.
- (16) Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat. Med.* 12 (10), 1133–1138.
- (17) Chaires, J. B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J.* 277 (5), 1098–1106.
- (18) Ray, S., Qureshi, M. H., Malcolm, D. W., Budhathoki, J. B., Celik, U., and Balci, H. (2013) RPA-mediated unfolding of systematically varying G-quadruplex structures. *Biophys. J.* 104 (10), 2235–2245.
- (19) Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* 5 (3), 182–186.
- (20) Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6 (1), 75–80.
- (21) Tran, P. L., Mergny, J. L., and Alberti, P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res.* 39 (8), 3282–3294.
- (22) Paeschke, K., Capra, J. A., and Zakian, V. A. (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* 145 (5), 678–691.
- (23) Gray, L. T., Vallur, A. C., Eddy, J., and Maizels, N. (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* 10 (4), 313–318.
- (24) Budhathoki, J. B., Maleki, P., Roy, W. A., Janscak, P., Yodh, J. G., and Balci, H. (2016) A Comparative Study of G-Quadruplex Unfolding and DNA Reeling Activities of Human RECQ5 Helicase. *Biophys. J.* 110 (12), 2585–2596.
- (25) Budhathoki, J. B., Stafford, E. J., Yodh, J. G., and Balci, H. (2015) ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity. *Nucleic Acids Res.* 43 (12), 5961–5970.
- (26) Ribeyre, C., Lopes, J., Boule, J. B., Piazza, A., Guedin, 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 (5), e1000475.
- (27) Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S., and Johnson, F. B. (2010) Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* 38 (4), 1114–1122.
- (28) Maizels, N. (2015) G4-associated human diseases. *EMBO Rep.* 16 (8), 910–922.
- (29) Fletcher, T. M., Sun, D., Salazar, M., and Hurley, L. H. (1998) Effect of DNA secondary structure on human telomerase activity. *Biochemistry* 37 (16), 5536–5541.
- (30) Balasubramanian, S., and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* 13 (3), 345–353.
- (31) Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33 (8), 877–881.
- (32) Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35 (2), 406–413.
- (33) Maizels, N., and Gray, L. T. (2013) The G4 genome. *PLoS Genet.* 9 (4), e1003468.
- (34) Qin, Y., and Hurley, L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* 90 (8), 1149–1171.
- (35) Eddy, J., and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 34 (14), 3887–3896.
- (36) Stegle, O., Payet, L., Mergny, J.-L., MacKay, D. J. C., and Huppert, J. L. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics* 25 (12), i374–382.
- (37) Nagatoishi, S., Isono, N., Tsumoto, K., and Sugimoto, N. (2011) Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* 93 (8), 1231–1238.
- (38) Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B., and Phan, A. T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.* 39 (21), 9448–9457.
- (39) Magbanua, E., Zivkovic, T., Hansen, B., Beschoner, N., Meyer, C., Lorenzen, I., Grotzinger, J., Hauber, J., Torda, A. E., Mayer, G., Rose-John, S., and Hahn, U. (2013) d(GGGT)4 and r(GGGU)4 are both HIV-1 inhibitors and interleukin-6 receptor aptamers. *RNA Biol.* 10 (2), 216–227.
- (40) Qureshi, M. H., Ray, S., Sewell, A. L., Basu, S., and Balci, H. (2012) Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency. *J. Phys. Chem. B* 116 (19), 5588–5594.
- (41) Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M., Jr, and Mergny, J. L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.* 44 (5), 1989–2006.
- (42) Joo, C., and Ha, T. (2012) Labeling DNA (or RNA) for single-molecule FRET. *Cold Spring Harb Protoc* 2012 (9), 1005–1008.
- (43) Globyte, V., Lee, S. H., Bae, T., Kim, J. S., and Joo, C. (2019) CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* 38 (4), e99466.
- (44) Globyte, V., and Joo, C. (2019) Single-molecule FRET studies of Cas9 endonuclease. *Methods Enzymol.* 616, 313–335.
- (45) Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A., and Yildiz, A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR–Cas9. *Sci. Adv.* 3 (8), ea00027.