JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

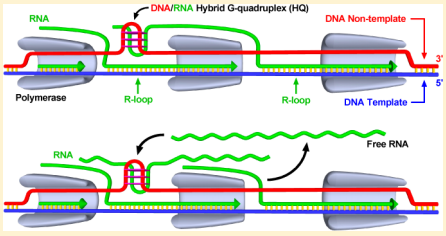
pubs.acs.org/JACS

Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA

Jia-yu Zhang, Ke-wei Zheng, Shan Xiao, Yu-hua Hao, and Zheng Tan*

State Key Laboratory of Biomembrane and Membrane Biotechnology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

ABSTRACT: We recently reported that a DNA:RNA hybrid G-quadruplex (HQ) forms during transcription of DNA that bears two or more tandem guanine tracts (G-tract) on the nontemplate strand. Putative HQ-forming sequences are enriched in the nearby 1000 nt region right downstream of transcription start sites in the nontemplate strand of warm-blooded animals, and HQ regulates transcription under both in vitro and in vivo conditions. Therefore, knowledge of the mechanism of HQ formation is important for understanding the biological function of HQ as well as for manipulating gene expression by targeting HQ. In this work, we studied the mechanism of HQ formation using an in vitro T7 transcription model. We show that RNA synthesis initially produces an R-loop, a DNA:RNA heteroduplex formed by a nascent RNA transcript and the template DNA strand. In the following round of transcription, the RNA in the R-loop is displaced, releasing the RNA in single-stranded form (ssRNA). Then the G-tracts in the RNA can jointly form HQ with those in the nontemplate DNA strand. We demonstrate that the structural cascade R-loop → ssRNA → HQ offers opportunities to intercept HQ formation, which may provide a potential method to manipulate gene expression.



INTRODUCTION

Nucleic acids carrying four tandem guanine tracts (G-tract) can fold into a four-stranded G-quadruplex structure.[1] Putative G-quadruplex-forming sequences (PQSs) are present in the promoter region of many proto-oncogenes, to mention a few examples, C-MYC,[2,3] C-KIT,[4] BCL-2,[5] NRAS,[6] and KRAS.[7] Surveys using computational methods revealed that PQS motifs are widely present in the genomes of various species, ranging from animals[8−16] to bacteria.[17,18] In humans, 376 000 PQS motifs have been identified.[9] They are found to exist in >40% of the human gene promoters.[12] More importantly, PQSs are not randomly distributed, but enriched near transcription start sites (TSSs),[10−17] suggesting that G-quadruplexes have functional roles in transcription. In support of this, experimental studies show that PQS motifs in the promoter regions of some proto-oncogenes, such as C-MYC,[3] C-KIT,[19] and KRAS,[7] form G-quadruplexes and affect transcription.

While previous investigations on genomic G-quadruplexes were almost exclusively focused on intramolecular structures formed by PQSs with four or more G-tracts, we recently reported a finding that transcription of double-stranded DNA (dsDNA) readily leads to formation of a DNA:RNA hybrid G-quadruplex (HQ) if the nontemplate DNA strand carries two or more G-tracts.[20] An HQ recruits G-tracts from both the nontemplate DNA strand and the nascent RNA transcript. Thus, it can form if a nontemplate DNA strand bears as few as two G-tracts. For this reason, putative hybrid G-quadruplex sequences (PHQSs) are far more abundant than PQSs in genomes near or within genes. In humans, for example, PHQSs are present in >97% of protein-coding genes, with an average of >73 PHQSs per gene. Similar to the PQSs, the PHQSs are concentrated immediately downstream of TSSs. Moreover, they are preferentially enriched in the nontemplate strand compared to the template strand and are mostly spliced out in mRNA.[21] These facts suggest a positive selection for PHQSs/HQ to function in transcription and transcription-related processes. Indeed, HQ was found to suppress transcription under both in vitro and in vivo conditions. The prevalence of PHQSs is not limited to humans. They have become constitutional in the genes of warm-blooded animals. It was suggested that HQ encodes an intrinsic cis control at the root level of transcription.

Given the universal presence of PHQSs in genes and the effect of HQ on transcription, the mechanism underlining HQ formation is of importance not only to transcription regulation, but also to other transcription-mediated processes. These processes may include immunogenesis,[22] recombination,[23] and genomic instability,[22−26] which involve transcription of DNA with guanine enrichment in the nontemplate strand. Using the T7 transcription model and the G-core (GGGGGAGGGGGGG) of the conserved sequence block II (CSB II) from human mitochondrial DNA as a PHQS motif, we studied the biochemical mechanism of HQ formation. Transcription of high G:C content DNA is often associated with the formation of an R-loop, a heteroduplex in which a RNA transcript remains annealed with its DNA template.[27] We show that a single transcription event originally forms an R-loop. In the following round of transcription, the RNA is displaced by the RNA being transcribed. The displaced RNA then has a chance to form HQ with the nontemplate DNA strand. We provide experimental evidence to show the R-loop → ssRNA → HQ structural cascade that an RNA transcript goes through during tran-

Received: August 21, 2013
Published: January 6, 2014

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

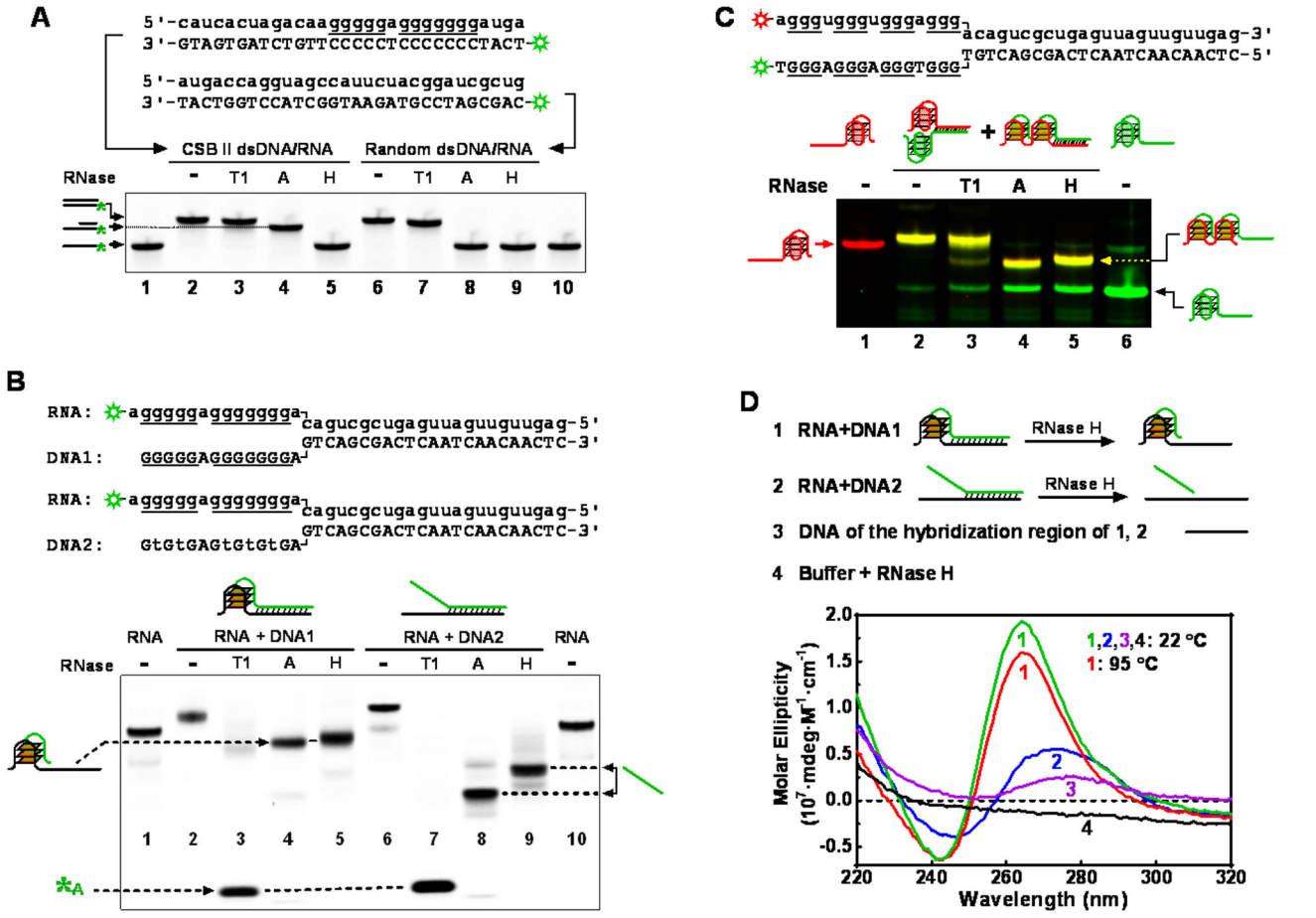


Figure 1. Susceptibility of the R-loop and HQ assembled from synthetic DNA and RNA to RNase T1, A, and H. (A) RNase A degraded the R-loop of the random sequence but not the U/C-free CSB II G-core. (B, C) The CSB II G-core and gggugggugggaggg formed HQ and were resistant to RNase T1, A, and H. Sequences (RNA in lowercase and DNA in uppercase) and constructs in (A)−(C) are shown above the gels. The construct was subjected to digestion with the RNase indicated. Cleavage products were resolved by native gel electrophoresis. The green asterisk in (A) and (B) indicates an FAM fluorescent dye covalently linked to the oligomer; green and red asterisks in (C) indicate Cy3 and Cy5 fluorescent dyes covalently linked to the oligomers, respectively. The minor band above the major one in (C), lane 6, might be a multimeric intermolecular G-quadruplex. (D) CD spectra of RNase H digested RNA + DNA1, RNA + DNA2, and control samples. Samples 1 and 2 were prepared as those in (B), lanes 5 and 9, respectively. The CD spectrum was obtained at 22 or 95 °C as indicated in the panel.

scription and demonstrate how HQ formation can be manipulated by intercepting RNA using a short oligonucleotide.

MATERIALS AND METHODS

Oligonucleotides and dsDNA. Synthetic oligonucleotides were obtained from Takara (Dalian, China) or Sangon Biotech (Shanghai, China). dsDNA was prepared by heating two complementary oligonucleotides at 95 °C for 5 min and then slowly cooling them to room temperature in a buffer of 10 mM $LiAsO_2(CH_3)_2$, pH 7.9, 50 mM LiCl. dsDNA used in Figures 4B and 5 was prepared by overlap polymerase chain reaction (PCR) using a biotinylated upstream primer and had an extended size downstream of the G-core to ensure that the G-core and its neighboring nucleotides were within the processive region of primer extension.

Assembly of the R-Loop and HQ and Digestion with RNases. The DNA:RNA heteroduplex was prepared by mixing synthetic complementary oligonucleotides, 1 μM each, in 10 mM Tris−HCl (pH 7.9) buffer containing 50 mM LiCl, followed by heating to 95 °C and then slow cooling to room temperature. The duplex was then diluted into 10 mM Tris−HCl (pH 7.9) buffer containing 40% PEG 200, 50 mM KCl, and 10 mM $MgCl_2$ to a final concentration of 50 nM. The samples were subjected to a digestion with the indicated RNase in the same way as the transcribed DNA for structural analysis.

Circular Dichroism (CD) Spectroscopy. DNA and RNA were annealed at equal molar ratio and diluted into $LiAsO_2(CH_3)_2$ buffer (pH 7.9), resulting in a final concentration of 4 μM oligonucleotide each, 40% (w/v) PEG 200, 50 mM KCl, and 8 mM $MgCl_2$. The samples were incubated at 37 °C for 1 h and then another 2.5 h with 22.5 U of RNase H. The digestion was stopped by adding 4 μL of 0.5 M EDTA. CD spectra were measured on a Chirascan-plus CD spectrometer (Applied Photophysics, United Kingdom) with a 0.5 mm path length cuvette at 22 or 95 °C.

Multiround Transcription. Transcription was carried out as we previously described[20] in 25 μL of transcription buffer containing 40 mM Tris−HCl, pH 7.9, 40% PEG 200, 50 mM KCl (Li^+ if indicated), 10 mM $MgCl_2$, 10 mM DDT, 2 mM spermidine, and 1 mM nucleoside triphosphates (NTPs) (Fermentas, Thermo Scientific, United States) at 37 °C for 1 h. The reaction was terminated by an addition of competitive dsDNA (5′-GAAATTAATACGACTCACTATA-3′)[28] to a final concentration of 2.5 μM.

Single- and Two-Round Transcription. Transcription was carried out in 25 μL of transcription buffer containing 0.05 μM DNA, 1 mM ATP, 1 mM GTP, 0.2 mM UTP, and 8 U/μL T7 RNA polymerase.[20] After incubation at 37 °C for 15 min, competitive DNA was added to a final concentration of 2.5 μM and CTP to 1 mM, followed by an incubation of 30 min.

Post-Transcription Digestion of DNA with RNase and Protease. Transcribed samples were mixed with an equal volume of

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

digestion buffer (10 mM Tris−HCl, pH 7.9, 40% PEG 200, 50 mM KCl) containing 0.2 U/μL RNase H, 2 U/μL RNase T1, or 0.4 μg/μL RNase A. The mixture was incubated at 37 °C for 30 min. The digestion buffer contained 40 mM EDTA when RNase A or T1 was used. When digestion with RNase H was followed by an RNase A treatment, a concentrate of RNase A and EDTA was added to final concentrations of 0.2 μg/μL and 20 mM, respectively. RNase-digested samples were further treated with 1 μg/μL proteinase K at 37 °C for 30 min.

Analysis of the RNA Transcript. A 2.5 pmol sample of DNA in 50 μL was transcribed with 2 mM ATP, CTP, GTP, or 7-deaza-GTP (dzGPT), 0.38 mM UTP, and 0.02 mM fluorescein-12-UTP (Roche). The samples were digested with 1 U of DNase I (Fermentas, Thermo Scientific) at 37 °C for 10 min. The reaction was stopped by adding 2 μL of 0.5 M EDTA. The samples were extracted with an equal volume of phenol/chloroform (1/1, v/v), dissolved in 50% deionized formamide, and resolved on a 6% denaturing gel.

DNA Dimethyl Sulfate (DMS) Footprinting. Transcribed DNA (100 μL) was mixed with an equal volume of Tris−HCl (pH 7.9) buffer containing 40% (w/v) PEG 200, 50 mM KCl, and 2 μg/μL proteinase K and incubated at 37 °C for 30 min. The DNA was then subjected to footprinting as described.[20,29,30]

RNA DMS Footprinting.[31,32] RNA in 200 μL of 0.05 μM transcribed DNA was mixed with 2 μL of 10% (v/v) DMS in ethanol and incubated at 37 °C for 2 min. After termination of the reaction by addition of 75 μL of water containing 20 μL of β-mercaptoethanol and 50 μg of fish sperm DNA, the samples were extracted with an equal volume of phenol/chloroform (1/1, v/v) and precipitated with 100% ethanol. The RNA was washed with 70% ethanol and reduced with sodium borohydride. After another precipitation with ethanol, aniline was added to induce cleavage at the modified nucleotides. The RNA was mixed with 2.5 μL of 4 μM primer (FAM-5′-TCTAGCAGCT-CGATGCAGATCG-3′) in water and 1 μL of 10 mM dNTPs, incubated at 65 °C for 5 min, and chilled on ice. Primer extension was initiated by supplying 4 μL of 5× RT buffer (250 mM Tris−HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT), 20 U of RiboLock RNase inhibitor (Fermentas, Thermo Scientific), and 200 U of Maxima reverse transcriptase (Fermentas, Thermo Scientific). After the reaction was stopped at 85 °C for 5 min, the extension product was precipitated with ethanol and resolved on a 10% denaturing gel.

S1 Nuclease Footprinting. Transcribed DNA (50 μL) was subjected to a post-transcription digestion with RNase A and H and protease K. Then 5 μL of 5 U/μL S1 nuclease (Takara, Dalian) and 12 μL of S1 nuclease buffer (300 mM CH_3COONa, pH 4.6, 2.8 M NaCl, and 10 mM $ZnSO_4$) were added, and the mixture was incubated at 37 °C for 5 min. The reaction was terminated by adding 83 μL of stop solution (4.4 μL of 0.5 M NaOH, 8 μL of 0.5 M EDTA, 10 μL of 10 mg/mL fish sperm DNA, and 60.6 μL of water). After phenol/chloroform (1/1, v/v) extraction and ethanol precipitation, the DNA was dissolved in 80% (v/v) deionized formamide, denatured at 95 °C for 5 min, and resolved by denaturing gel electrophoresis.

UV-Cross-Linking of RNA with DNA. Transcription and UV irradiation were carried out as described.[20] Then DNA was purified by the TIANquick mini purification kit (Tiangen, Beijing), followed by primer extension with 0.4 μM FAM-5′-TCTAGCAGCTCGATG-CAGATCG-3′ primer, 4 U of Deep VentR (exo−) (NEB, United States) in a 50 μL volume containing 75 mM Tris−HCl, pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20, 2 mM $MgCl_2$, 0.05 mM dNTP, and 5% (v/v) DMSO. G and T ladders were prepared in the same way using a nontranscribed and non-cross-linked DNA strand in the presence of Acy-CTP and Acy-TTP (NEB), respectively, in a molar ratio of 1/2 and 1/1 to dCTP and dTTP, respectively.

Probe RNA in HQ. DNA labeled at the 5′ end with a biotin was transcribed and digested with RNase A and H. An aliquot of 50 μL was mixed with 50 μL of a streptavidin MagneSphere paramagnetic particle suspension (Promega, United States). After three washes with 500 μL of 10 mM Tris−HCl buffer (pH 7.9) containing 40% PEG 200, 50 mM KCl, 10 ng/μL fish sperm DNA, and 20 mM EDTA, the immobilized DNA was dissolved in 10 μL of 20 mM EDTA solution containing 100 mM K^+ chelator 18-crown-6 (Sigma-Aldrich) and incubated at 37 °C for 10 min. The sample was then mixed with a 5′-Cy5-labeled probe (5′-GTCATCCCCCCCTCCCCCTTG-3′) in 50 μL of 10 mM Tris−HCl (pH 7.9) and 50 mM LiCl. The mixture was heated at 95 °C for 5 min, cooled to room temperature, and electrophoresed on a 12% native gel.

Polyacrylamide Gel Electrophoresis. Native gel electrophoresis was carried out at 4 °C and 8 V/cm on a polyacrylamide gel that contained 75 mM KCl and 40% (w/v) PEG 200 unless otherwise indicated in 1× TBE buffer that contained 75 mM KCl.[30] Denaturing gel electrophoresis was carried out on a polyacrylamide gel. DNA on the gel was recorded on a Typhoon 9400 phosphor imager (GE Healthcare, United States).

RESULTS

Identification of HQ and the R-Loop by RNase Digestion. We have used native gel electrophoresis to detect the formation of HQ in dsDNA.[20,21] This technique relies on the facts that the RNA fragment in HQ is resistant to cleavage of several RNases and that a dsDNA carrying an HQ migrates at a slower rate than the correspondent fully annealed dsDNA. Transcription of guanine-rich (G-rich) DNA produces an R-loop structure in which a nascent RNA transcript remains base-paired with the template DNA strand.[27] Previously, we used RNase A to digest RNA products that were in single-stranded form or hybridized with a DNA. RNase A cuts at the 3′ end of C's and U's in ssRNA as well as RNA in the DNA:RNA hybrid at low salt concentration of 0−100 mM.[20] To test the specificity of RNases in more detail, we analyzed the activity of RNase T1, A, and H on the RNA in the DNA:RNA duplex and HQ.

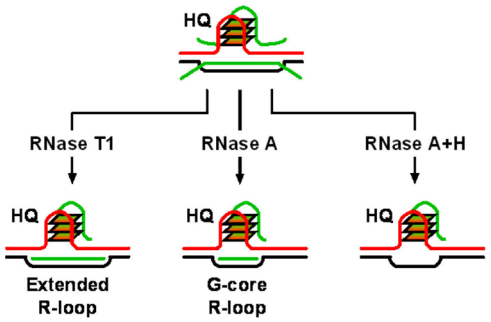


Figure 2. Structures remaining after RNase treatment (bottom, from left to right): HQ and the R-loop with residual flanking sequences beyond the G-rich region, HQ and the R-loop at the G-rich region, and HQ only. Red and green lines indicate the nontemplate DNA strand and RNA, respectively.

In Figure 1A, two DNA:RNA duplexes (i.e., R-loop), carrying either the mitochondrial CSB II G-core along with a random flanking sequence or a completely random sequence, were treated with RNase T1, A, and H, respectively. The removal of RNA from the duplex caused the DNA to change its mobility in native gel electrophoresis, which was detected by the fluorescent FAM dye covalently attached to its 5′ end. As expected, RNase T1, which cleaves ssRNA, had little effect on the duplexes (lanes 3 and 7); RNase H completely digested the RNA in the two substrates (lanes 5 and 9), restoring the original migration of the two DNA oligonucleotides (lanes 1 and 10). However, the activity of RNase A was sequence-dependent. The random RNA was completely degraded (lane 8), but the CSB II G-core-containing DNA only showed a slightly faster migration than the undigested one (lane 4). This

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

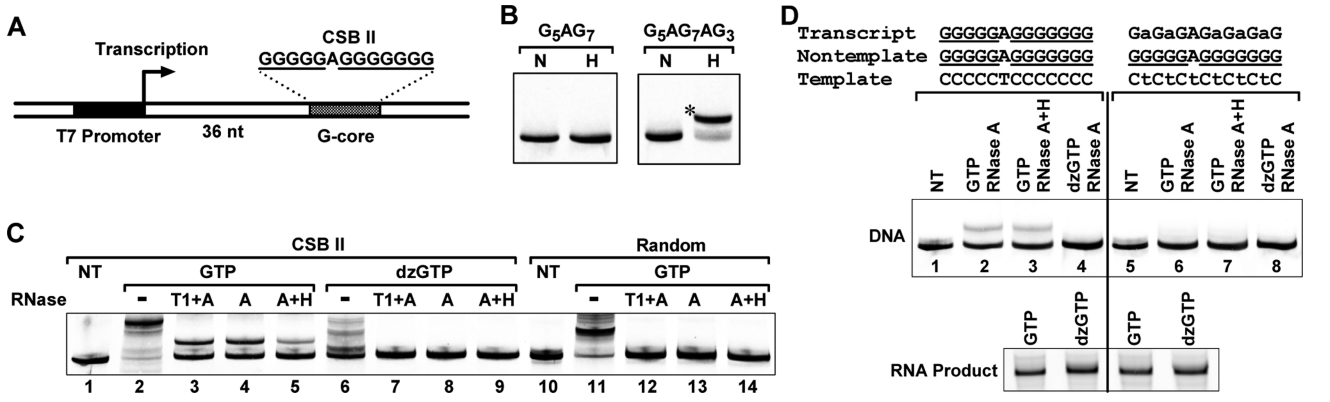


Figure 3. HQ and R-loop formation in multiple-round transcription. (A) DNA used, in which a CSB II G-core was placed 36 nt downstream of a T7 promoter. (B) The CSB II G-core was unable to form a G-quadruplex (left panel) unless an additional G-tract was supplied (right panel). G-quadruplex formation was detected on native gel before (N) and after (H) the DNAs were subjected to heat denaturation/renaturation. The asterisk indicates an intramolecular G-quadruplex formed by the G-core shown above the gel. (C) HQ formation in transcription detected by native gel electrophoresis. DNA was transcribed with GTP or dzGTP and the three other NTPs, followed by hydrolysis with RNase T1 + A, RNase A, and RNase A + H, respectively. NT indicates DNA without transcription. (D) Formation of HQ requires G-tract(s) from RNA. DNA without (left) or with (right) mutation only on the template strand within the G-core was transcribed using GTP or dzGTP. After digestion with the indicated RNase, HQ formation was analyzed on a native gel (top gel). RNA product was also analyzed on a denaturing gel without RNase digestion (bottom gel).

implies degradation only at the random region but not at the U/C-free CSB II G-core.

To test the activity of the enzymes toward HQ, we annealed an RNA oligonucleotide with a DNA oligonucleotide (DNA1) to form a partial duplex (Figure 1B). Both carried a CSB II G-core, providing an opportunity for them to jointly form an HQ. The same RNA was also annealed with a similar DNA oligonucleotide (DNA2) that did not bear the G-core, and therefore, HQ was unable to form in the duplex. Both duplexes showed a slower migration than the ssRNA on native gel (lanes 2 and 6 versus lane 1). Digestion with RNase T1, which cut at G in ssRNA, produced a fast band, suggesting cleavage at the G's flanking the A at the 5′ RNA end that was not assembled into HQ (lanes 3 and 7). Digestion with RNase A or H of RNA + DNA1 removed the RNA in the hybridization region (R-loop) and yielded a band (lanes 4 and 5) running faster than the one in the undigested sample (lane 2), but slower than those in the digested samples in which HQ could not form (lanes 8 and 9). This feature indicates that the CSB II motif in the RNA was not released, implying that the G-core in the RNA formed HQ with the one in the DNA1 and was, as a result, resistant to the two RNases. The samples treated with RNase A (lanes 4 and 8) always migrated faster than those treated with RNase H (lanes 5 and 9) because RNase A can also cleave ssRNA besides the R-loop.

It is not clear whether the HQ in the RNA + DNA1 duplex was resistant to T1 as judged from the migration of the dye (Figure 1B, lane 3), because the 5′ RNA end might be tethered off the HQ. Thus, the release of the dye might not reflect the status of the HQ. To further clarify this issue, we prepared a partial duplex of RNA and DNA (Figure 1C). The RNA was labeled with a Cy5 dye at its 5′ end and the DNA with a Cy3 dye at its 3′ end. Both oligomers carried four G_3-tracts that could form either an intramolecular G-quadruplex or HQ. This duplex showed a slower migration (Figure 1C, lane 2) than both the RNA (lane 1) and DNA (lane 6) alone. Digestion with RNase A and H of the R-loop region resulted in two major product bands on the gel (Figure 1C, lanes 4 and 5). The first one from the top represented an HQ because it carried both fluorescent dyes; the second band from the top was a DNA monomer that could be produced by the substrate in which the RNA and DNA formed an intramolecular G-quadruplex separately. These results again demonstrated that the HQ was resistant to both RNase A and H. The duplex was largely resistant to RNase T1 (lane 3). The digestion only showed a marginal effect (lane 3), suggesting that the G-tract flanking the 5′ RNA dye was assembled into HQ. The two minor digestion bands (lane 3) that migrated similarly to those of the products of RNase A and H (lanes 4 and 5) probably indicated cleavage at the interface between the duplex and G-quadruplex where a few nucleotides might be in single-stranded form.

The HQ structure in the digested RNA + DNA1 sample was further confirmed by CD spectroscopy (Figure 1D). The spectrum featured a negative peak near 245 and a positive peak near 265 nm (curve 1), which is characteristic of a parallel G-quadruplex and similar to that of the HQ of telomeric DNA and RNA.[33] On the contrary, this feature was not seen in the RNA + DNA2 sample in which the DNA did not bear a G-tract (curve 2). The HQ in RNA + DNA1 was extraordinarily stable that an obvious melting was not observed when the sample was heated to 95 °C, similar to the human telomeric G-quadruplex in PEG.[34]

The results in the above experiments showed that the R-loop survived the treatment with RNase T1, the R-loop at the CSB II G-core region survived RNase A, and HQ survived all three RNases. According to these results, it is deduced that, for a transcribed DNA that carries RNA in HQ, R-loop, and single-stranded form, and if the G-core is U/C-free, a post-transcription treatment with RNase T1 will retain HQ and the R-loop at the G-core and its flanking region, RNase A will retain HQ and the R-loop at the G-core region, and treatment with both RNase A and H will detect HQ (Figure 2). This post-transcription digestion provided a technique for the identification of the different structures in this work.

HQ and R-Loop Formation in Transcription. We used a dsDNA carrying a CSB II G-core 36 nt downstream of a T7 promoter (Figure 3A) in our transcription experiments. The ability of the G-core to form a G-quadruplex was first tested by heat denaturation/renaturation followed by native gel electrophoresis. DNA possessing a G-quadruplex migrates slower than

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

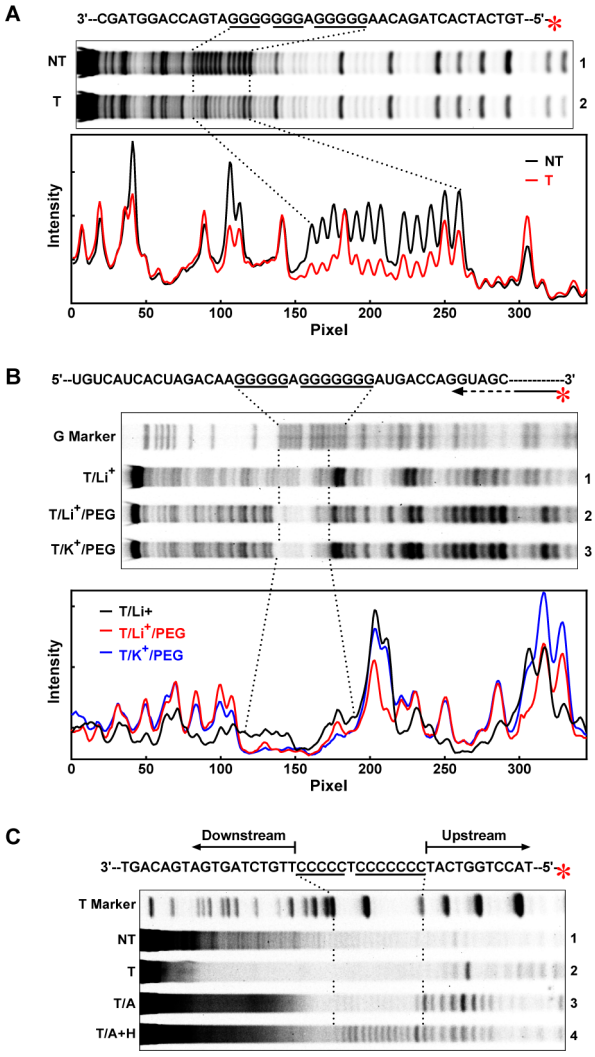


Figure 4. HQ and R-loop formation in multiple-round transcription. (A) Participation of the nontemplate DNA strand in HQ formation detected by DMS footprinting. DNA was transcribed (T) or was not transcribed (NT). Cleavage fragments were resolved on a denaturing gel. (B) Participation of RNA in HQ formation detected by DMS footprinting. DNA was transcribed, and the primer was annealed with RNA and extended by reverse transcriptase. Extension products were resolved on a denaturing gel. T/Li^+, T/Li^+/PEG, and T/K^+/PEG indicate transcription in Li^+, Li^+ + PEG, and K^+ + PEG solution, respectively. The graphs beneath the gels in (A) and (B) show the digital scans of the gels. (C) R-loop formation detected by S1 nuclease footprinting. DNA was transcribed and then treated with RNase A (T/A) or RNase A and H (T/A + H), followed by nuclease S1 digestion. DNA fragments were resolved by denaturing gel electrophoresis. NT indicates DNA without transcription. The asterisk in (A)−(C) indicates an FAM dye covalently linked to the 5′ end of the DNA.

the fully annealed DNA.[20,21,29,30] The results showed that the CSB II G-core was unable to form an intramolecular G-quadruplex by itself, unless an additional G-tract was supplied (Figure 3B, left versus right panel). This result is expected because a stable G-quadruplex normally requires four G-tracts of three or more consecutive guanines and the CSB II motif can only provide three G_3-tracts.

Transcription was performed with T7 polymerase and four NTPs in a solution containing 50 mM K^+ and 40% PEG 200. PEG was added to stabilize G-quadruplex formation in dsDNA to facilitate its detection.[20,21,29,30] After termination of the reaction by adding competitive dsDNA to capture the polymerase,[28] the samples were treated by the indicated RNase to digest RNA in different structural forms. Then an additional digestion with protease K was followed to remove the T7 polymerase and RNases. The DNAs were run on a native gel to detect HQ formation (Figure 3C). In lanes 3 and 4, an extra slower band appeared. This band represented a DNA with an R-loop or a G-quadruplex or both because of their resistance to RNase A. The presence of HQ in this band was indicated by the fact that the band was still present when the sample was digested with RNase H to remove the R-loop (lane 5). Although the CSB II G-core is unable to form a G-quadruplex by itself, the RNA transcript can provide extra G-tracts.[20]

The participation of an RNA transcript in the formation of HQ was supported by transcription in which normal GTP was replaced by dzGTP. Due to the substitution of N7 with C7, dzGTP does not form the Hoogsteen hydrogen bond required for G-quadruplex assembly.[35] In this case, no extra band was observed (Figure 3B, lanes 7−9), as was in the random DNA carrying no CSB II G-core (Figure 3B, lanes 12−14). The participation of RNA was further verified by mutation introduced to the template DNA strand without changing the nontemplate DNA strand (Figure 3D). This modification did not impair transcription (bottom gel), but abolished the supply of G-tracts from RNA. As a result, no G-quadruplex was detected as expected (top gel, right half versus left half).

The involvement of the nontemplate DNA in the formation of HQ was verified by DMS footprinting in which the guanine is protected from methylation and later chemical cleavage because of the Hoogsteen hydrogen bonding at the N7. The results revealed a G_5−G_3−G_3 protection pattern for the G_5AG_7-core (Figure 4A). This suggests that HQ of three G-quartet layers formed, with the guanine in the middle of the G_7 serving as a loop. DMS footprinting was also conducted for RNA (Figure 4B) in which base modification was detected by extending a primer by reverse transcriptase.[32] RNA obtained in transcription in Li^+ solution without PEG was used as a reference (lane 1). Compared with the reference, the RNA obtained in the transcription in Li^+ or K^+ solution in the presence of PEG showed protection at the G-core region (lanes 2 and 3). For the G_7-tract, the protection seemed more obvious at the 5′ side. This may suggest a different involvement of guanines in the HQ assembly. The protection of the RNA in Li^+/PEG solution could be attributed to the ability of PEG to promote G-quadruplex formation under salt-deficient conditions.[36]

To detect if an R-loop was present in the transcribed DNA, we analyzed the template strand by S1 nuclease footprinting (Figure 4C). S1 nuclease cleaves single-stranded DNA and RNA. After transcription, the sample was subjected to treatment with RNase A or A + H before the S1 digestion. Without RNase treatment, the C_7TC_5 stretch and a large portion of its immediate downstream region in the transcribed DNA were obviously protected (lane 2), indicating formation of an R-loop at the G_5AG_7-core and its downstream region, which is in agreement with previous observations for CSB II DNA.[37,38] After digestion with RNase A, the protection of the C_7TC_5 region remained, but that of its downstream region disappeared (lane 3). This suggests persistence of the R-loop at the C_7TC_5-core because of the lack of C and U in the G_5AG_7 region. Because RNase H is expected to cleave the RNA in the whole R-loop, additional cleavage indeed occurred at the

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

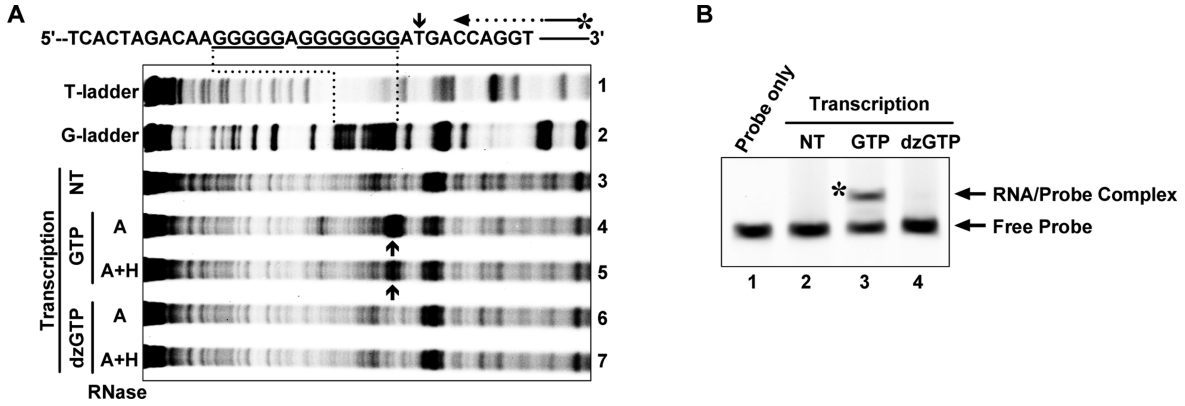


Figure 5. Presence of RNase-resistant RNA in HQ. (A) DNA was transcribed using GTP or dzGTP in the presence of 4S-UTP followed by digestion with the indicated RNase. The 4S-UTP was then cross-linked with DNA by UV irradiation. Cross-linking to the nontemplate DNA strand was analyzed by stalling of primer extension. (B) DNA was transcribed using GTP or dzGTP, followed by digestion with RNase A and H. The DNA was then immobilized onto magnetic beads and washed. RNA was released and hybridized with a C-rich DNA complementary to the CSB II G-core and run on a native gel without K^+ and PEG. NT in both (A) and (B) indicates no transcription.

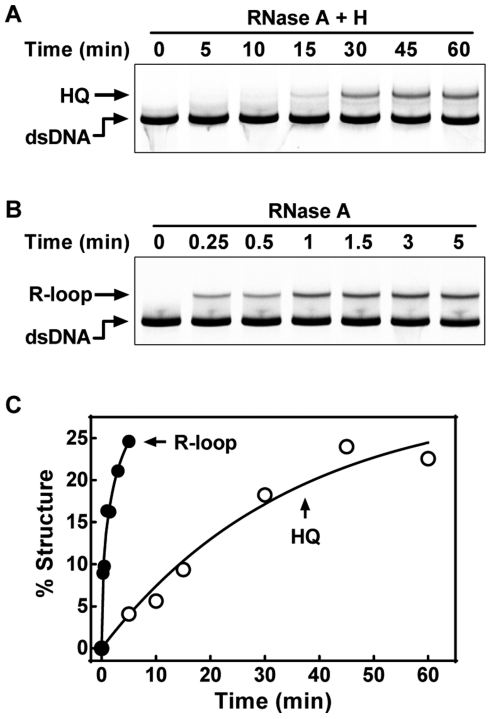


Figure 6. Formation of HQ lags behind that of an R-loop. DNA was transcribed for the indicated period. After the reaction was stopped by addition of excess competitive DNA, the DNA was treated with (A) RNase A (top gel) or (B) RNase A and H (bottom gel) before being resolved by native gel electrophoresis. (C) Quantification of HQ and R-loop formation as a function of time. The retarded band in (A) and (B) was expressed as a percentage of the total DNA in each lane.

C_7TC_5-core (lane 4). These results indicate that an R-loop was formed at the CSB II G-core and its downstream region in the transcription.

Further Evidence of RNA Participation in HQ Formation. Besides G, DMS also methylates N1 of A and N3 of C in RNA. This prevents the natural hydrogen bonding at the affected bases and causes termination of primer extension by reverse transcriptase in DMS footprinting.[39,40] Therefore, the comparison between samples for the RNA was more complicated in reaching a conclusion than for the DNA. To further verify the participation of RNA in HQ, we carried out two additional independent experiments to confirm the presence of RNA in the HQ structure (Figure 5).

First, we incorporated 4-thiouridine (4S-UTP) into the RNA during transcription.[20] After a post-transcription digestion with RNases, the RNA remaining in HQ was cross-linked with the nontemplate DNA strand upon UV irradiation. By primer extension, specific cross-linking was detected downstream of the G-core in the DNA transcribed using GTP (Figure 5A, arrow, lanes 4 and 5, arrowhead), but not in that using dzGTP to prevent the RNA from forming a G-quadruplex (lanes 6 and 7).

Second, we immobilized transcribed DNA onto streptavidin-coated magnetic beads after it was digested by RNases. After washing, the RNA fragment in HQ was released by chelating both K^+ and Mg^{2+}. The RNA was then hybridized with a fluorescent DNA probe complementary to the G-core of the RNA and run on a native gel (Figure 5B). The result showed that the RNA G-core sequence was detected in DNA transcribed using GTP (lane 3, asterisk), but not in that using dzGTP (lane 4) in which HQ could not form. The detection of RNase-resistant G-rich RNA in the above two experiments further confirmed the participation of RNA in HQ formation in the DNA transcribed using GTP.

Formation of HQ Lags Behind That of an R-Loop. Our results in Figure 4 showed that both HQ and an R-loop were present in the transcribed DNA. To investigate how the two structures were formed and the relationship between them, we reduced the transcription efficiency by excluding pyrophosphatase in transcription and examined the order of appearance of the two individual structures. Two sets of DNA samples were arranged with respect to their post-transcription treatment. One was treated with both RNase A and H to detect HQ; the other was treated with RNase A only to reserve both the R-loop and HQ (Figure 2). As shown in Figure 6A, the appearance of a retarded band resistant to both RNase A and H showed that HQ started to form 15 min after transcription was initiated. When transcribed DNAs were only treated with RNase A, a retarded band began to show up in less than 1 min and reached a high magnitude in 5 min (Figure 6B). According to the results in Figure 6A, HQ had not apparently formed yet within this period of transcription. Therefore, this retarded band had to be mainly from the R-loop. The large difference in their time of

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

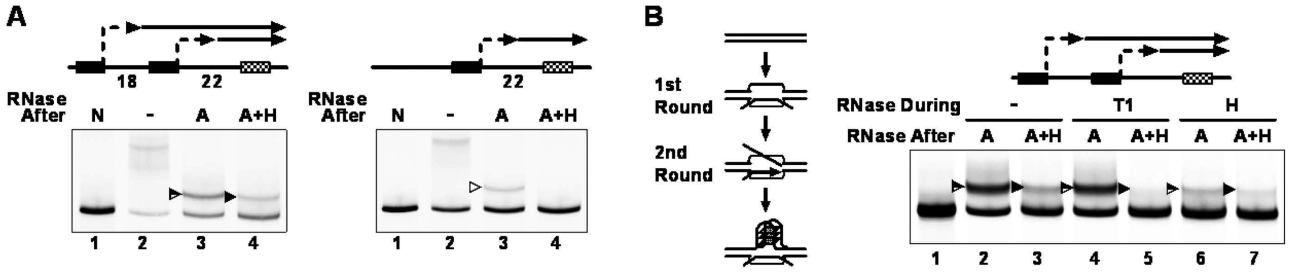


Figure 7. Structural transformation of the RNA transcript in HQ formation. (A) HQ formation in two rounds (left) or one round (right) of transcription. A polymerase was first allowed to proceed to +13 (dashed arrow) by supplying only three NTPs. Then the other NTP was added along with competitive DNA for the polymerase to complete a full-length transcription (solid arrow). After digestion with the indicated RNase, DNA was resolved by native gel electrophoresis. The number in the schemes indicates the number of nucleotides between two adjacent elements. (B) Effect on HQ formation of RNase T1 and H added during transcription. Transcription and processing were conducted as in (A, left), but in the presence of RNase T1 or H. Scheme at the left illustrates HQ formation in the absence of RNase during transcription.

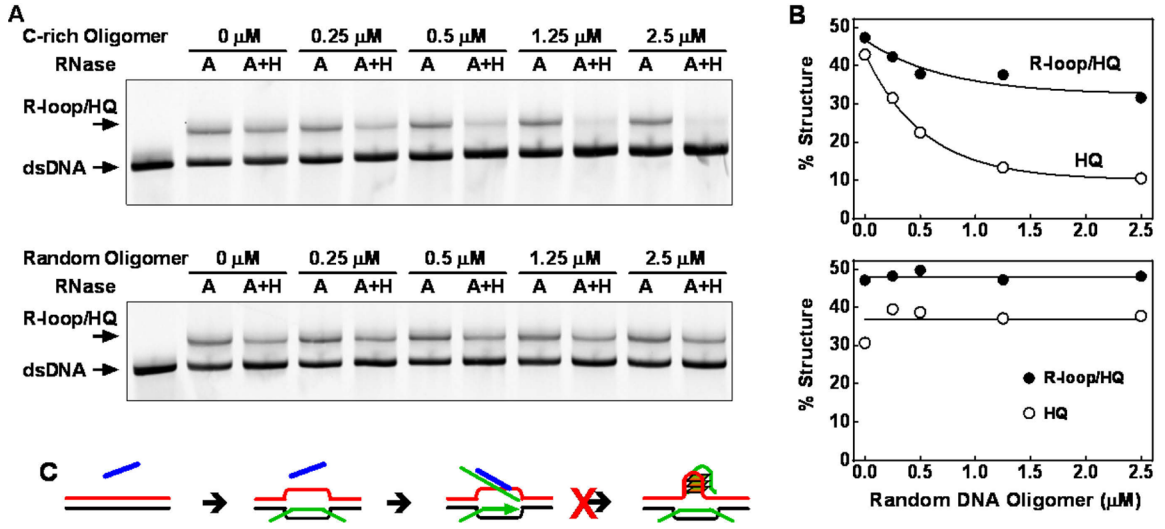


Figure 8. Manipulation of HQ formation by targeting RNA. Transcription was conducted as in Figure 6 in the presence of a C-rich DNA oligomer complementary to the G-rich region of the RNA transcript (A, top gel; B, top graph) or a random sequence (A, bottom gel; B, bottom graph) of various concentrations. Transcribed DNA was processed and structure detected as in Figure 6. (C) Scheme illustrating that an RNA transcript (green) in an R-loop displaced by a subsequent transcription is captured by a complementary DNA oligomer (blue) and prevented from forming HQ.

formation (Figure 6C) implied that the R-loop formed prior to the formation of the HQ.

HQ Formation Follows R-Loop Displacement. The formation of an R-loop leaves the nascent RNA hybridized with the template DNA strand. If multiple rounds of transcription take place in a DNA, it can be imagined that the RNA in an R-loop formed in a previous round of transcription will be displaced during the next round. We anticipated that this RNA displacement would supply RNA for HQ formation. This explains why the formation of the HQ lagged behind that of the R-loop (Figure 6).

To test this assumption, we constructed two DNAs that could undergo one and two rounds of transcription, respectively. The DNAs carried either one or two T7 promoters in a row in the upstream region of a CSB II G-core (schemes in Figure 7A). T7 polymerase undergoes premature transcription abortion cycles in the synthesis of the first 2−8 nucleotides before it reaches the +13 position to engage in a processive state.[41] On this basis, three NTPs were first added to allow the transcription to proceed to the +13 position. Then the other NTP was supplied along with an excess of competitive DNA for the transcription to proceed further to the end of the DNA. When the transcription was finished, the released polymerase would be captured by the competitive DNA, preventing them from further transcription cycles.[28] In this experimental setting, the DNA with one promoter would undergo a single round of transcription, but the DNA with two promoters would be transcribed twice. In the latter case, the RNA synthesized by the promoter closer to the G-core would be detached by the polymerase fired from the other promoter. HQ was detected in the DNA containing two promoters as judged from the RNase H-resistant band (Figure 7A, left panel, lane 4). The HQ must be formed by the detached RNA because only the R-loop (Figure 7A, right panel, lane 3), and not HQ (Figure 7A, right panel, lane 4), was detected in the DNA containing only one promoter.

The above experiments suggested an R-loop → ssRNA → HQ structural cascade gone through by the RNA synthesized from the promoter adjacent to the CSB II G-core (Figure 7B, scheme at left). To verify this, we repeated the two-round transcription experiment, but with RNase T1 or H added during transcription (Figure 7B). Because of its specificity for ssRNA, RNase T1 was expected to intercept the ssRNA displaced from the template before it formed HQ. In support of

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

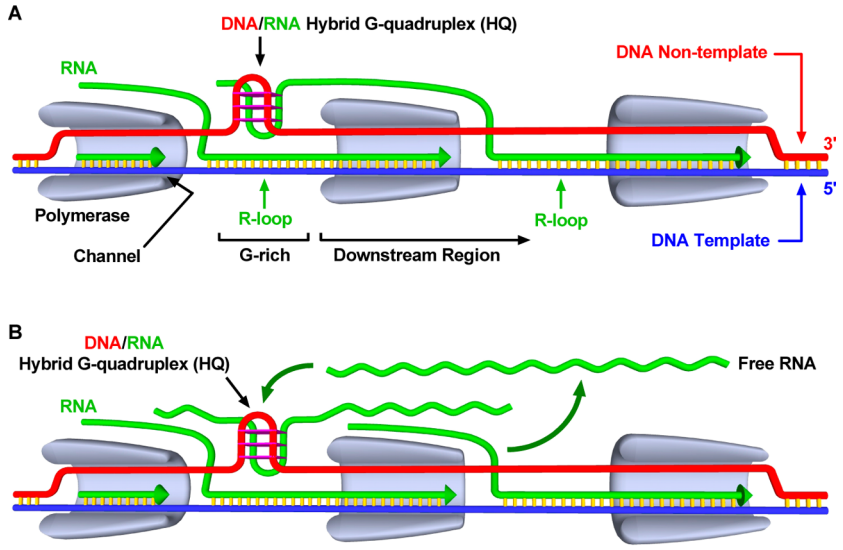


Figure 9. Proposed model of HQ formation in transcription. A nascent RNA remains hybridized with the template DNA strand, forming an R-loop that stretches downstream. The RNA in the R-loop is displaced in a consecutive transcription. The displaced RNA either (A) can form HQ with the nontemplate DNA strand directly or (B) remains free and has a possibility to form HQ later. R-loop formation was drawn here according to the "extended hybrid" mechanism, because the R-loop formation was not affected by RNase T1 added during transcription (Figure 7B). Alternatively, the R-loop can also form by a "thread-back" mechanism. The mechanism by which the R-loop forms should not affect HQ formation.

this expectation, it abolished HQ formation (Figure 7B, lane 5). On the other hand, RNase T1 added during transcription did not affect R-loop formation because the R-loop is first formed and not a substrate of RNase T1 (Figure 7B, lane 4). RNase H added during transcription would degrade the R-loop before it was displaced from the DNA template. As a result, this enzyme dramatically reduced the amount of both the R-loop (Figure 5B, lane 6) and HQ (Figure 7B, lane 7).

Manipulation of HQ Formation by Targeting the Formation Pathway. The above RNase treatment during transcription demonstrated that the different structures appearing in the structural cascade can be used to manipulate HQ formation. To illustrate this possibility, we performed transcription in the presence of a DNA oligonucleotide complementary to the RNA transcript. We anticipated that the oligonucleotide would intercept the RNA transcript by hybridization with the RNA and, as a result, impair HQ formation. Transcription was conducted in the presence of a 23 nt random or C-rich DNA oligonucleotide that can bind the RNA transcript at the region spanning across the CSB II G-core. In Figure 8, it can be seen that the oligonucleotide strongly inhibited HQ formation in a concentration-dependent manner. When the R-loop and HQ were detected as a whole by digestion with only RNase A, their formation was also inhibited, but to a much lower degree. The inhibition of HQ formation by the oligonucleotide in this case would increase the competition from strand annealing of the DNA duplex. As a result, this could indirectly disfavor R-loop formation. As expected, no effect was observed for the random oligonucleotide. These results not only illustrate the possibility of manipulating HQ formation, but also verify the HQ formation pathway (Figure 8C).

DISCUSSION

In this study, we analyzed the mechanism of cotranscriptional HQ formation in DNA possessing G-tracts in the nontemplate strand using the T7 transcription model and revealed a pathway through which HQ forms. During transcription, a DNA duplex is melted in the channel in the polymerase to template RNA synthesis. The separated DNA strands normally reanneal back to restore the duplex form as the polymerase moves forward.[42] The formation of an R-loop keeps the nontemplate DNA single-stranded. The R-loop can form by either an "extended hybrid" or a "thread-back" mechanism.[27] In either case, the RNA in the R-loop will be displaced from the template by the polymerase in the follow-up transcription. Our results (Figures 6 and 7) suggest that the G-tracts in the displaced RNA get a chance to form HQ with the G-tracts in the nontemplate DNA strand (Figure 9).

Guanine enrichment is featured in many physiological elements in genomes, such as immunoglobulin class switch sequences,[22] prokaryotic[43] and mitochondrial[44] replication origins, the MAZ transcription termination element,[45,46] and other transcribed genes.[46,47] Transcription of such DNA is known to result in formation of an R-loop at the G-rich region, leaving the nontemplate strand unpaired.[24,27] This G-rich sequence-dependent R-loop formation has been implicated in immunogenesis,[22] recombination,[23] and genomic instability.[24−26]

Our finding of HQ formation revealed a secondary structure accompanying R-loop formation.[20] The connection between HQ and the R-loop revealed in this work strongly suggests that HQ may play an important role in all these physiological processes in which the R-loop was thought to be involved. In those processes, a unique feature is the requirement of transcription in physiological orientation in which the nontemplate DNA strand is G-rich. For example, transcription of G-rich DNA, a well-recognized source of genome instability, is often associated with R-loop formation and strand bias toward guanine enrichment on the nontemplate strand.[24,48−51] Another example is that the structure formation and transcription blockage of T7 RNA polymerase and mammalian RNA polymerase II take place when guanine-richness is in the nontemplate but not in the template DNA strand.[51−54] This strand discrimination implies a special mechanism in the physiological function of G-rich DNA, which is not yet properly

dx.doi.org/10.1021/ja4085572 | J. Am. Chem. Soc. 2014, 136, 1381−1390

interpreted from our current knowledge. Apparently, the formation of HQ offers a plausible explanation of the orientation dependence because only a G-rich nontemplate can produce G-rich RNA transcripts to allow HQ formation. In this regard, we hypothesize that HQ may be an important player in these processes.

In our recent work, we showed the abundance and prevalence of PHQS motifs in the genes in warm-blooded animals.[21] HQ was shown to influence transcription under both in vitro and in vivo conditions, and their presence inversely correlates with the maximal transcription level of genes in human tissues.[20] On the basis of these facts, targeting HQ formation may offer opportunities to manipulate gene expression in putative therapeutic applications. As illustrated in this work, intercepting the displaced ssRNA (Figure 8) or destroying the R-loop (Figure 7B) may prevent HQ formation. HQ-interacting compounds may directly influence HQ formation and, as a result, affect the related physiological processes.

■ CONCLUSION

We found that, in the T7 transcription model, a newly synthesized RNA transcript remains annealed with its template DNA strand at the G-rich and downstream regions. It is displaced in the next transcription cycle, releasing a single-stranded RNA that can jointly form a DNA:RNA hybrid G-quadruplex with the nontemplate DNA strand. We also demonstrate that the structural cascade R-loop → ssRNA → HQ offers opportunities to interfere with HQ formation.

■ AUTHOR INFORMATION

Corresponding Author

z.tan@ioz.ac.cn

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Grants 2013CB530802, 2012CB720601, and 2010CB945300 from the Ministry of Science and Technology of China (MSTC) and Grants 21072189 and 30970617 from the National Natural Science Foundation of China (NSFC).

■ REFERENCES

(1) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.
(2) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167.
(3) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593.
(4) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584.
(5) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. *J. Am. Chem. Soc.* **2006**, *128*, 1096.
(6) Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nat. Chem. Biol.* **2007**, *3*, 218.
(7) Cogoi, S.; Xodo, L. E. *Nucleic Acids Res.* **2006**, *34*, 2536.
(8) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901.
(9) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(10) Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. *J. Med. Chem.* **2008**, *51*, 5641.
(11) Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2008**, *36*, 1321.
(12) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406.
(13) Du, Z.; Zhao, Y.; Li, N. *Genome Res.* **2008**, *18*, 233.
(14) Du, Z.; Kong, P.; Gao, Y.; Li, N. *Biochem. Biophys. Res. Commun.* **2007**, *354*, 1067.
(15) Zhao, Y.; Du, Z.; Li, N. *FEBS Lett.* **2007**, *581*, 1951.
(16) Du, Z.; Zhao, Y.; Li, N. *Nucleic Acids Res.* **2009**, *37*, 6784.
(17) Rawal, P.; Kummarasetti, V. B.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S. K.; Chowdhury, S. *Genome Res.* **2006**, *16*, 644.
(18) Yadav, V. K.; Abraham, J. K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. *Nucleic Acids Res.* **2008**, *36*, D381.
(19) Bejugam, M.; Sewitz, S.; Shirude, P. S.; Rodriguez, R.; Shahid, R.; Balasubramanian, S. *J. Am. Chem. Soc.* **2007**, *129*, 12926.
(20) Zheng, K. W.; Xiao, S.; Liu, J. Q.; Zhang, J. Y.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 5533.
(21) Xiao, S.; Zhang, J. Y.; Zheng, K. W.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 10379.
(22) Stavnezer, J.; Guikema, J. E.; Schrader, C. E. *Annu. Rev. Immunol.* **2008**, *26*, 261.
(23) Poveda, A. M.; Le Clech, M.; Pasero, P. *Transcription* **2010**, *1*, 99.
(24) Li, X.; Manley, J. L. *Genes Dev.* **2006**, *20*, 1838.
(25) Aguilera, A.; Gomez-Gonzalez, B. *Nat. Rev. Genet.* **2008**, *9*, 204.
(26) Duquette, M. L.; Huber, M. D.; Maizels, N. *Cancer Res.* **2007**, *67*, 2586.
(27) Aguilera, A.; Garcia-Muse, T. *Mol. Cell* **2012**, *46*, 115.
(28) Ujvari, A.; Martin, C. T. *J. Mol. Biol.* **1997**, *273*, 775.
(29) Zhang, C.; Liu, H. H.; Zheng, K. W.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2013**, *41*, 7144.
(30) Zheng, K. W.; Chen, Z.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2010**, *38*, 327.
(31) Morris, M. J.; Negishi, Y.; Pazsint, C.; Schonhoft, J. D.; Basu, S. *J. Am. Chem. Soc.* **2010**, *132*, 17831.
(32) Wells, S. E.; Hughes, J. M.; Igel, A. H.; Ares, M., Jr. *Methods Enzymol.* **2000**, *318*, 479.
(33) Xu, Y.; Ishizuka, T.; Yang, J.; Ito, K.; Katada, H.; Komiyama, M.; Hayashi, T. *J. Biol. Chem.* **2012**, *287*, 41787.
(34) Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185.
(35) Fletcher, T. M.; Sun, D.; Salazar, M.; Hurley, L. H. *Biochemistry* **1998**, *37*, 5536.
(36) Kan, Z. Y.; Yao, Y. A.; Wang, P.; Li, X. H.; Hao, Y. H.; Tan, Z. *Angew. Chem., Int. Ed.* **2006**, *45*, 1629.
(37) Xu, B.; Clayton, D. A. *Mol. Cell. Biol.* **1995**, *15*, 580.
(38) Xu, B.; Clayton, D. A. *EMBO J.* **1996**, *15*, 3135.
(39) Fourmy, D.; Yoshizawa, S. *Wiley Interdiscip. Rev.: RNA* **2012**, *3*, 557.
(40) Tijerina, P.; Mohr, S.; Russell, R. *Nat. Protoc.* **2007**, *2*, 2608.
(41) Gong, P.; Martin, C. T. *J. Biol. Chem.* **2006**, *281*, 23533.
(42) Jiang, M.; Ma, N.; Vassylyev, D. G.; McAllister, W. T. *Mol. Cell* **2004**, *15*, 777.
(43) Masukata, H.; Tomizawa, J. *Cell* **1990**, *62*, 331.
(44) Lee, D. Y.; Clayton, D. A. *J. Biol. Chem.* **1998**, *273*, 30614.
(45) Ashfield, R.; Patel, A. J.; Bossone, S. A.; Brown, H.; Campbell, R. D.; Marcu, K. B.; Proudfoot, N. J. *EMBO J.* **1994**, *13*, 5656.
(46) Gromak, N.; West, S.; Proudfoot, N. J. *Mol. Cell. Biol.* **2006**, *26*, 3986.
(47) Mechti, N.; Piechaczyk, M.; Blanchard, J. M.; Jeanteur, P.; Lebleu, B. *Mol. Cell. Biol.* **1991**, *11*, 2832.
(48) Kim, N.; Jinks-Robertson, S. *Nat. Rev. Genet.* **2012**, *13*, 204.
(49) Gan, W.; Guan, Z.; Liu, J.; Gui, T.; Shen, K.; Manley, J. L.; Li, X. *Genes Dev.* **2011**, *25*, 2041.
(50) Shinkura, R.; Tian, M.; Smith, M.; Chua, K.; Fujiwara, Y.; Alt, F. W. *Nat. Immunol.* **2003**, *4*, 435.

(51) Kim, N.; Jinks-Robertson, S. *DNA Repair* **2011**, *10*, 953.
(52) Tornaletti, S.; Park-Snyder, S.; Hanawalt, P. C. *J. Biol. Chem.* **2008**, *283*, 12756.
(53) Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N. *Genes Dev.* **2004**, *18*, 1618.
(54) Belotserkovskii, B. P.; Liu, R.; Tornaletti, S.; Krasilnikova, M. M.; Mirkin, S. M.; Hanawalt, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12816.