VIP G-Quadruplexes Very Important Paper

International Edition: DOI: 10.1002/anie.201503648
German Edition: DOI: 10.1002/ange.201503648

Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase**

*Jia-quan Liu, Shan Xiao, Yu-hua Hao, and Zheng Tan**

Abstract: *G-quadruplex-forming sequences are enriched near transcription start sites (TSSs) in animal genes. They readily form G-quadruplexes in transcription, which in turn regulate transcription. Therefore, the control of G-quadruplex formation is important for their functionality. It is now shown that G-quadruplexes form efficiently on the non-template, but hardly on the template DNA strand in the downstream vicinity of TSSs in DNA duplexes when they are transcribed by the T7 RNA polymerase (RNAP). Structural analysis reveals that the T7 RNAP causes distortion in a DNA duplex both inside and in front of the enzyme. This structural distortion leads to strand-biased G-quadruplex formation when a G-quadruplex-forming sequence is partially fed into the T7 RNAP to a position about seven nucleotides away from the front of RNA synthesis. Based on these facts, we propose a model for the strand-biased formation of G-quadruplexes in transcribed DNA duplexes.*

Nucleic acid sequences with four tandem guanine tracts (G-tracts) can fold into a four-stranded G-quadruplex structure. Such sequences are present throughout the genomes of various organisms, ranging from bacteria to plants and animals.[1] G-quadruplex structures have recently been detected in human cells.[2] The enrichment of potential G-quadruplex-forming sequences near transcription start sites (TSSs) in animal genes[3] suggests that G-quadruplexes play functional roles in transcription and related processes. Furthermore, G-quadruplexes are also implicated in other crucial physiological and pathological processes, such as gene expression, DNA replication, telomere maintenance, and genome instability.[4] For these reasons, G-quadruplexes are promising therapeutic targets in pharmaceutical applications.[5]

In 2013, we found that two different types of G-quadruplexes could form in association with transcription. In the upstream region of a TSS, the formation of an intramolecular DNA G-quadruplex can be induced by remote downstream transcription activity.[6] In the downstream region of the TSS, a distinct type of G-quadruplex, namely a DNA:RNA hybrid G-quadruplex, can be produced by the joint participation of G-tracts from both the non-template DNA strand and the RNA transcript.[7] The progress made in the past decade has undoubtedly consolidated the biological significance of G-quadruplex structures. Whereas the physiological functions of G-quadruplexes may be diverse, one important aspect related to their functionality is how G-quadruplex formation is controlled in different physiological processes. Currently, little information is available concerning this issue.

Herein, we present the first example of how the formation of a G-quadruplex in double-stranded DNA (dsDNA) is governed by the interaction between the RNA polymerase (RNAP) and the DNA that is being transcribed. We found that transcription with T7 RNAP triggered G-quadruplex formation preferentially on the non-template DNA strand, but rarely on the template strand downstream of a promoter. Structural analysis indicated that a DNA duplex was destabilized both inside and in front of the T7 RNAP during transcription, which led to the formation of a G-quadruplex when a G-quadruplex-forming motif was fed into the RNAP to a critical point approximately seven nucleotides (nts) away from the transcription bubble. Our data suggests that the strand-biased formation of G-quadruplexes results from the different constraints that are imposed on the two DNA strands by the RNAP.

To study G-quadruplex formation in dsDNA, we transcribed two DNAs in which a $G_3(TG_3)_3$ core was placed 15 nts downstream of a T7 promoter (Figure 1, top) either on the non-template (G15/5) or template (C15/5) strand. Full-length transcription was conducted in the presence of four NTPs. G-quadruplex formation in the DNAs was detected by native gel electrophoresis in which DNA that bear a G-quadruplex migrates more slowly than the corresponding DNA without a G-quadruplex.[6–8] A heat denaturation/renaturation procedure was used to produce G-quadruplex-bearing DNA as a reference.[7,8] These references demonstrated that the G-core had physically identical abilities to form G-quadruplexes on the two strands (Figure 1, lanes 5 and 8, top gel). However, it behaved differently in transcribed DNA. G-quadruplex formation occurred in a large fraction of the G15/5 DNA (Figure 1, lane 6, top gel). In contrast, only a small amount of G-quadruplex formation was observed for C15/5 (Figure 1, lane 9, top gel). The G-quadruplex formation in the two DNAs was confirmed by DMS footprinting in which guanine residues involved in G-quadruplex assembly were protected

[*] Dr. J.-q. Liu,[+] Dr. S. Xiao,[+] Y.-h. Hao, Prof. Dr. Z. Tan
State Key Laboratory of Biomembrane and Membrane Biotechnology, Institute of Zoology, Chinese Academy of Sciences
Beijing 100101 (P. R. China)
E-mail: z.tan@ioz.ac.cn
Dr. S. Xiao[+]
Key Laboratory of Zebrafish Modeling and Drug Screening for Human Diseases of Guangdong Higher Education Institutes
Department of Cell Biology, School of Basic Medical Sciences
Southern Medical University, Guangzhou 510515 (P. R. China)
[+] These authors contributed equally to this work.

[**] This research was supported by the MSTC (2013CB530802, 2012CB720601) and the NSFC (31470783).
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201503648.

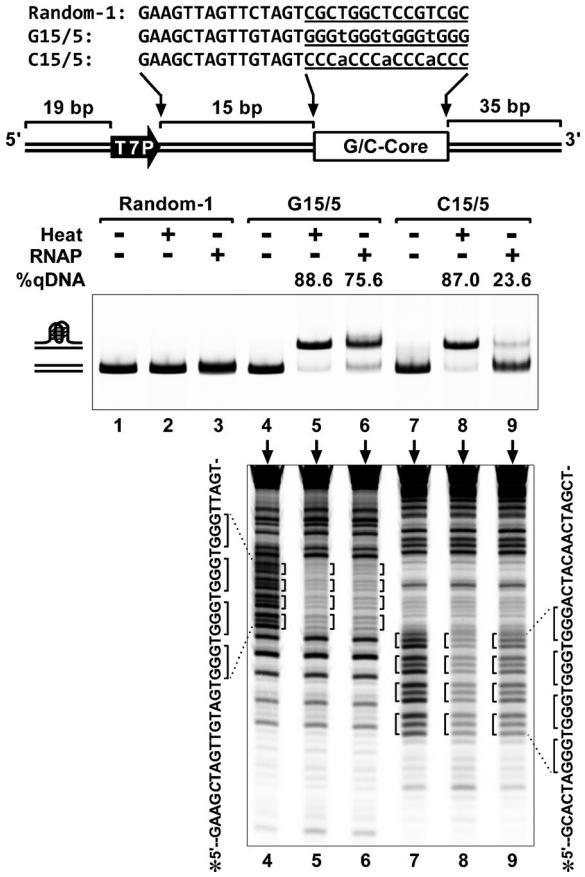

Figure 1. G-quadruplex formation during full-length transcription in dsDNA without (Random-1) or with a $G_3(TG_3)_3$ core on the non-template (G15/5) or template (C15/5) DNA strand downstream of a T7 promoter (T7P). The DNAs (scheme at the top) were transcribed by T7 RNA polymerase (RNAP) in the presence of four NTPs or subjected to heat denaturation/renaturation (Heat). They were either resolved on a native gel after digestion with RNase and protease (top gel) or subjected to DMS footprinting without the digestion (bottom gel). DNA was visualized by the fluorescence of the FAM dye (*) attached at the 5′ ends of the non-template (G15/5) and template (C15/5) DNA strands. The number above the native gel indicates the amount of G-quadruplex-bearing DNA (qDNA) as a percentage of the total DNA in the lane. The scheme on the left side of the top gel indicates the structure of the DNAs.

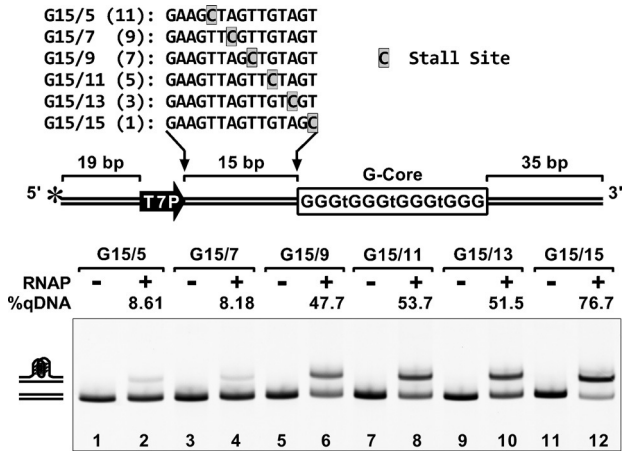

Figure 2. Approaching T7 RNAP triggers G-quadruplex formation on the non-template DNA strand in partially transcribed dsDNA. The DNAs containing a T7 promoter (T7P) and a downstream $G_3(TG_3)_3$ core on the non-template strand were transcribed by T7 RNAP in the presence of ATP, GTP, and UTP. They were subsequently resolved on a native gel and visualized by the FAM dye (*) attached at the 5′ end of the non-template strand. The lack of CTP prevented the RNAP to proceed beyond the C residue on the non-template strand. The numbers in parentheses indicate the distance from the stall site to the 5′ end of the G-core. The numbers above the gels indicate the amount of G-quadruplex-bearing DNA (qDNA) as a percentage of the total DNA in the lane.

from chemical cleavage.[9] Similar and obvious protection to the guanines in the G-core was observed for both the heated and transcribed G15/5 DNAs (Figure 1, lanes 5 and 6, bottom gel). In the C15/5 DNA, however, the protection to the G-core was weaker in the transcribed than in the heated DNA sample (lane 9 vs. lane 8, bottom gel).

As the extent of G-quadruplex formation was identical for the two strands in the heated DNA samples, the strand-biased G-quadruplex formation in the transcribed DNAs implied that G-quadruplex formation on the template strand of C15/5 was suppressed. We recently showed that transcription could lead to the formation of a DNA:RNA hybrid G-quadruplex when a G-core was on the non-template DNA strand.[7,8b,10] As the G-core on the template strand in C15/5 could only form an intramolecular G-quadruplex, we wondered if G-quadruplex formation on the template strand was limited by the lack of RNA participation. To test this hypothesis, we transcribed 6 DNAs in which a stall site (cytosine) was arranged at different positions upstream of the G-core on the non-template strand (Figure 2, top). Transcription was conducted in the absence of CTP, such that an RNAP could approach the G-core at defined distances, but would not pass the G-core to produce G-tract-bearing RNA (Figure S1); therefore, a hybrid G-quadruplex cannot form. In Figure 2, it is shown that G-quadruplex formation hardly occurred when the stall site was nine or more nts away from the G-core (lanes 2 and 4). However, G-quadruplex formation was significantly enhanced when RNA synthesis was only seven nts away from the G-core, resulting in an approximately six-fold increase in G-quadruplex formation (lane 6). Maximal G-quadruplex formation occurred when RNA synthesis stalled right at the 5' boundary of the G-core (Figure 2, lane 12), which was comparable to that in the full-length transcription (Figure 1, lane 6). These results indicate that a lack of RNA participation does not limit G-quadruplex formation.

We then carried out the same type of experiment with DNAs carrying the G-core on the template strand. As shown in Figure 3, G-quadruplex formation was not detected in any of the DNAs transcribed to different extents towards the G-core. To ensure that G-quadruplexes did and did not form in the two sets of DNAs, respectively, a representative DNA from each set was analyzed by DMS footprinting. The G15/15 that had been shown to form G-quadruplexes (Figure 2, lane 12) had its four G-tracts well protected (Figure 4, lane 2 vs. 1). For the C15/15 that did not show visible G-quadruplex formation (Figure 3, lane 7), no protection was detected at the G-tracts (Figure 4, lane 4 vs. 3). These results therefore confirmed that the RNAP triggered G-quadruplex formation in a strand-biased manner when approaching the G-core.

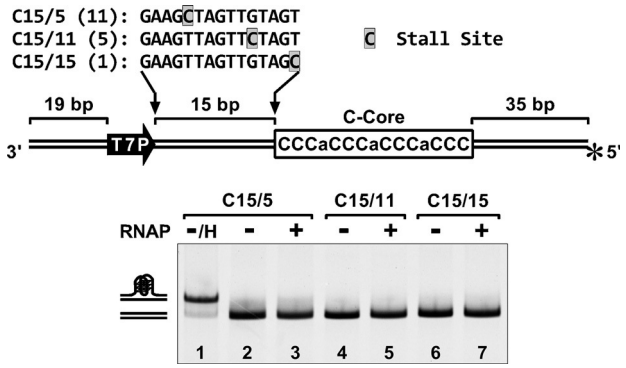

Figure 3. Approaching T7 RNAP does not trigger G-quadruplex formation on the template DNA strand in partially transcribed dsDNA. The DNAs containing a T7 promoter (T7P) and a downstream $G_3(TG_3)_3$ core on the template strand were transcribed and processed as in Figure 2. DNAs were visualized by the FAM dye attached at the 5′ end of the template DNA strand. The DNA in lane 1 was subjected to a heat denaturation/renaturation procedure without transcription to generate G-quadruplex-bearing DNA as a reference. The numbers in parentheses indicate the distance from the stall site to the 5′ end of the G-core.

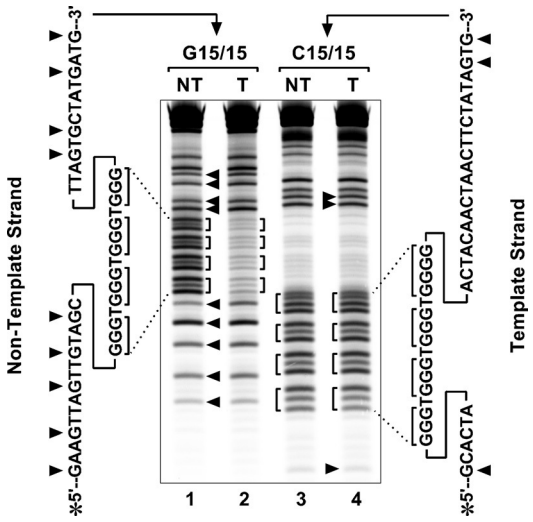

Figure 4. DMS footprinting of G-quadruplex formation triggered by approaching RNAP in partially transcribed dsDNA. The DNAs was not (NT) or was (T) transcribed as in Figure 2 and then immediately subjected to DMS footprinting. DNA fragments were resolved on a denaturing gel and visualized by the FAM dye (*) attached at the 5′ end of the non-template (G15/15) and template (C15/15) DNA strand, respectively.

Although G-quadruplex formation was not seen in C15/5 for partial transcription, it was detected in a small fraction of the DNA when transcribed full-length (Figure 1, lane 9, top gel). The formation of G-quadruplexes in the latter case might be caused by the RNAP that had passed the G-core. Our recent study had shown that G-quadruplex formation could be triggered by downstream transcription activity.[6]

The strand-biased G-quadruplex formation was further demonstrated in transcription with variations in NTP concentration, which determines the translocation rate of polymerase on DNA.[11] As shown in the Supporting Information, Figure S2, G-quadruplex formation on the non-template DNA strand was reduced with a decrease in NTP concentration. In agreement with the results shown in Figure 3, G-quadruplex formation was not detected at any NTP concentrations when the G-core was placed on the template DNA strand (Figure S3). We also tested DNAs with a G-core from the *NRAS* gene (Figure S4). Like the $G_3(TG_3)_3$ core, the *NRAS* G-core can form an intramolecular G-quadruplex of three G-quartet layers. Compared to the $G_3(TG_3)_3$ core, the increase in loop size in the *NRAS* G-core was expected to result in a less stable G-quadruplex.[12] Again, G-quadruplex formation was only detected in the DNAs with the G-core on the non-template, but not in those with the G-core on the template strand when the DNAs were partially transcribed.

Single-stranded guanine-rich nucleic acids spontaneously form G-quadruplex structures in salt solution. In a DNA duplex, however, a G-quadruplex-forming motif is constrained by the flanking sequences at its 5′ and 3′ sides as well as by the complementary C-rich region to maintain its base-paired integrity. These constraints suppress the formation of G-quadruplexes or, in other words, the integrity of the double helix needs to be disrupted for a G-quadruplex to form. To obtain structural insights, we analyzed the structural status of the partially transcribed DNAs using potassium permanganate ($KMnO_4$) footprinting in which thymine residues were preferentially cleaved in melted or distorted DNA duplexes.[13] During transcription, a RNAP denatures a DNA duplex and maintains a small transcription bubble that moves along with the enzyme to accommodate nascent RNA synthesis. In complete agreement with previous reports,[14] we found that a stalled RNAP created a transcription bubble of approximately 10 nts in a random DNA between positions -8 and $+2$ relative to the stall site (Figure 5A, bracket) as judged from the heavy cleavages (▼) on the non-template DNA strand at the transcribed region (green bar).

Next, we examined the C15/15, G15/7, and G15/9 DNA under the same conditions. The DNAs were labeled at the downstream side instead of at the upstream side as for the random DNA. We used this strategy because the intense cleavages in the promoter (Figure 5A, red bars) and bubble (Figure 5A, bracket) regions would shield off the visualization of cleavages at their downstream region if the labels were introduced at the 5′ end of the non-template DNA strand. In C15/15, which had its stall site right at the 3′ side of the G-core, the thymines in the transcription bubble encountered more cleavage in the transcribed than in the untranscribed DNA as expected (Figure 5B, ▼). Furthermore, enhanced cleavage (< 2-fold) was observed at the thymines in a region that began at the stall site and extended downstream (▽). Although transcription did not trigger G-quadruplex formation in the DNA (Figure 3, lane 7 and Figure 4, lane 4), this enhanced cleavage implies that a RNAP also causes distortion in the DNA in front of the enzyme. In the G15/7 DNA, a similar cleavage enhancement was also found downstream of the stall site (Figure 5C, ▽), which could be attributed to the distortion in front of the RNAP. In G15/9 (Figure 5D), the thymines in the transcription bubble (▼) were similarly cleaved to those in the other DNAs. However, a strikingly enhanced cleavage occurred at the thymines

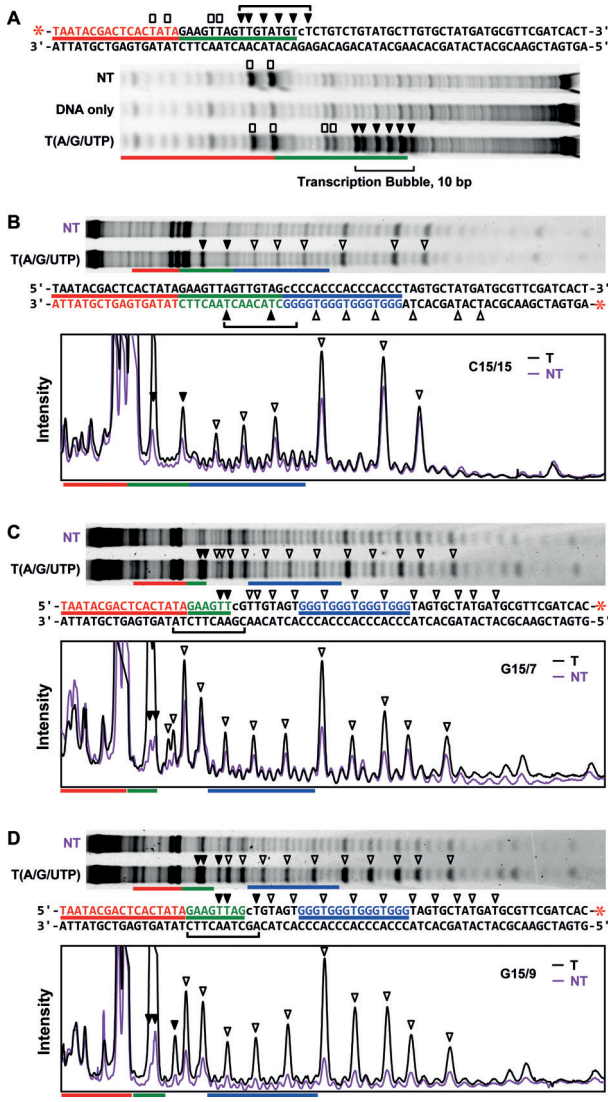

Figure 5. Structural distortion in partially transcribed dsDNA detected by potassium permanganate footprinting. DNA labeled with a FAM dye (*) at the indicated position was incubated with T7 RNAP in the absence (NT) and presence (T) of ATP, GTP, and UTP and then immediately subjected to footprinting. DNA fragments were resolved on a denaturing gel and visualized by the FAM dye. The graph under each gel was produced by digitizing the sample lanes in the gel using the ImageQuant 5.2 software. DNA used: A) Random-2; B) C15/15; C) G15/7; D) G15/9. Red, green, and blue bars indicates promoter, transcribed, and G-core regions, respectively. The brackets under the gels and beside sequences indicate the region of the transcription bubble. Cleavage at thymine residues upstream of (□), within (▼), or downstream (▽) of the transcription bubble is indicated.

downstream of the stall site (▽), which was three to eight times greater than that for un-transcribed G15/9. This enhancement indicates that a greater structural distortion occurred accompanying G-quadruplex formation in the DNA (Figure 2, lane 6).

T7 RNAP performs many abortive transcription cycles before proceeding to an elongation phase to produce a full-length transcript. A stable elongation complex forms when RNA synthesis reaches the position of +13 relative to the promoter. In our aforementioned experiments, partial transcription was limited to 15 nts downstream of TSS; therefore, the strand-biased formation of G-quadruplexes in such proximal G-cores was associated with the initiation stage of transcription. To determine how G-quadruplexes form when a transcription goes beyond the initiation phase, we transcribed DNAs with the $G_3(TG_3)_3$ placed 40 nts downstream of the promoter. In the presence of four NTPs, a strand bias was also observed for these distal G-cores (Figure S5, lane 3 vs. 6). In this case, the amount of G-quadruplex on the non-template was about twice as that on the template DNA strand, in a ratio that was smaller than that (ca. 3-fold) in the initiation stage (Figure 1, lane 6 vs. 9). The reduced ratio could be due to the fact that the transcription cycles are less frequent in the elongation phase than in the initial abortive phase. Under the partial-transcription conditions, the strand bias was still observed, as G-quadruplex formation was only found on the non-template, but not on the template DNA strand (Figure S5, lane 8 vs. 10).

Our structural analysis revealed that the RNAP generated a major structural disruption in the DNA within the RNAP/DNA complex around the transcription bubble. Furthermore, a minor distortion was also generated downstream of the transcription bubble (Figure 5). In the G15/7 DNA, the minor structural distortion extended well beyond the G-core (Figure 5C, ▽), but did not lead to efficient G-quadruplex formation (Figure 2, lane 4). This means that the minor distortion alone was not enough to trigger G-quadruplex formation. The formation of G-quadruplexes on the non-template DNA strand was sensitive to the distance between the transcription bubble and the G-core. It underwent a sharp transition when the T7 RNAP was approaching the G-core, suggesting that the distortion around the transcription bubble is crucial. G-quadruplex formation jumped from a low basal level to a high percentage when the front boundary of RNA synthesis was approximately seven nts away from the upstream border of the G-core (Figures 2, S2, and S4). In contrast, G-quadruplex formation was suppressed when the same G-core was placed on the opposite template strand (Figures 3, S3, and S4).

The strand-biased induction of G-quadruplexes by the approaching RNAP likely resulted from the different interactions between the RNAP and the two DNA strands because our experiments involved only RNAP and DNA without the participation of any third party. In support of this, a previous study has demonstrated different T7 RNAP/DNA contacts for the two DNA strands.[15] Interestingly, T7 RNAP can perform transcription using a single-stranded template without the non-template DNA strand downstream of a promoter.[16] This implies that the interaction between the RNAP and the template DNA strand is necessary and sufficient to guide RNA synthesis and that the non-template strand is not an essential element in transcription. An RNAP needs to pull a DNA into its active site to catalyze the templated RNA synthesis. This process may require a tight grip on the template DNA strand, which, as a result, may suppress the formation of G-quadruplexes on the DNA strand.

Based on our observations and the above considerations, we propose a biased strand-constraining model for the control of G-quadruplex formation in T7 transcription (Figure 6). It is

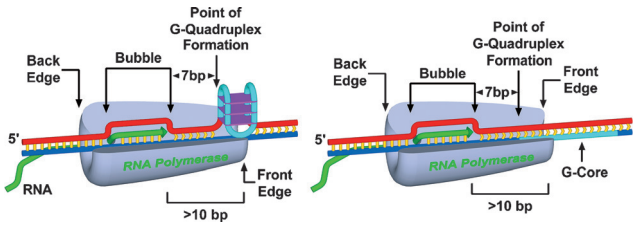

Figure 6. A biased strand-constraining model for the control of G-quadruplex formation in DNA duplexes transcribed with T7 RNAP. The T7 RNAP maintains a transcription bubble to host RNA synthesis when moving along a DNA. According to a previous study, the front edge of a T7 RNAP lies ahead of the 3′ end of RNA synthesis by more than ten nts. Our data show that for a G-core on the non-template DNA strand, the G-quadruplex begins to form when the 3′ end of RNA synthesis is approximately 7 nts away from the G-core. Under these conditions, the G-core has partially entered the RNAP. G-quadruplex formation on the non-template DNA strand indicates that the G-core on this strand has sufficient flexibility to form a G-quadruplex. The inability of the G-core on the template strand to form a G-quadruplex implies that the template strand is tightly constrained by the RNAP, preventing it from forming a G-quadruplex.

known that the front edge of a T7 RNAP lies ahead of the 3′ end of RNA synthesis by more than ten nts.[17] Therefore, when a T7 RNAP has reached the critical point towards a G-quadruplex-forming motif, the motif is partially fed into the RNAP. At this point, structural disruption on a DNA duplex provides an opportunity for the G-quadruplex to form. However, the two DNA strands are differently constrained in the RNAP/DNA complex. The template strand is tightly constrained by the protein but the non-template strand has sufficient freedom and flexibility. As a result, G-quadruplex formation is permitted on the non-template (Figure 6, left), but suppressed on the template DNA strand (Figure 6, right). In summary, the strand-biased formation of G-quadruplexes is granted by the combined effects of structural disruption of the DNA helix and the constraints on the template DNA strand that are imposed by the RNAP.

G-quadruplex-forming motifs are enriched on both the non-template and template DNA strands near the TSS in the genes of warm-blooded animals.[3a–c] The strand-biased formation of G-quadruplexes in DNA duplexes suggests that such motifs may have greater importance on the non-template strand. Suppressing G-quadruplex formation on the template DNA strand is necessary because the template strand has to be in a linear form to serve as template. Our work demonstrates that the phage T7 RNAP is able to control the formation of G-quadruplexes while performing its transcriptional activity. As a single-subunit RNA polymerase, the T7 RNAP shares the basic transcription steps and characteristics with the multi-subunit RNA polymerases.[18] It is unclear if the strand-biased formation of G-quadruplexes also occurs in transcription with other RNA polymerases. Surprisingly, intramolecular G-quadruplex-forming motifs were not found in the T7 phage genome when it was analyzed using the consensus sequence of $G_{\geq 3}$-(N_{1-7}-$G_{\geq 3}$)$_{\geq 3}$.[3a] This finding suggests that the suppression of G-quadruplex formation on the template DNA strand is probably an intrinsic feature of templated RNA synthesis. Therefore, we anticipate that the suppression of G-quadruplex formation on the template DNA strand may also occur with other RNA and even DNA polymerases, which all require a linear template sequence to function.

Keywords: DNA · G-quadruplexes · nucleic acids · secondary structure · transcription

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8992–8996
Angew. Chem. **2015**, *127*, 9120–9124

[1] a) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916; b) H. Takahashi, A. Nakagawa, S. Kojima, A. Takahashi, B. Y. Cha, J. T. Woo, K. Nagai, Y. Machida, C. Machida, *J. Biosci. Bioeng.* **2012**, *114*, 570–575; c) A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2011**, *39*, 4917–4927; d) J. A. Capra, K. Paeschke, M. Singh, V. A. Zakian, *PLoS Comput. Biol.* **2010**, *6*, e1000861; e) Z. Du, Y. Zhao, N. Li, *Nucleic Acids Res.* **2009**, *37*, 6784–6798; f) S. G. Hershman, Q. Chen, J. Y. Lee, M. L. Kozak, P. Yue, L. S. Wang, F. B. Johnson, *Nucleic Acids Res.* **2008**, *36*, 144–156; g) M. A. Mullen, K. J. Olson, P. Dallaire, F. Major, S. M. Assmann, P. C. Bevilacqua, *Nucleic Acids Res.* **2010**, *38*, 8149–8163.
[2] a) L. Yuan, T. Tian, Y. Chen, S. Yan, X. Xing, Z. Zhang, Q. Zhai, L. Xu, S. Wang, X. Weng, B. Yuan, Y. Feng, X. Zhou, *Sci. Rep.* **2013**, *3*, 1811; b) E. Y. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, *Nat. Commun.* **2013**, *4*, 1796; c) G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
[3] a) S. Xiao, J. Y. Zhang, K. W. Zheng, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 10379–10390; b) Z. Du, Y. Zhao, N. Li, *Genome Res.* **2008**, *18*, 233–241; c) J. Eddy, N. Maizels, *Nucleic Acids Res.* **2008**, *36*, 1321–1333; d) A. Verma, K. Halder, R. Halder, V. K. Yadav, P. Rawal, R. K. Thakur, F. Mohd, A. Sharma, S. Chowdhury, *J. Med. Chem.* **2008**, *51*, 5641–5649; e) Y. Zhao, Z. Du, N. Li, *FEBS Lett.* **2007**, *581*, 1951–1956.
[4] a) N. Maizels, L. T. Gray, *PLoS Genet.* **2013**, *9*, e1003468; b) M. Tarsounas, M. Tijsterman, *J. Mol. Biol.* **2013**, *425*, 4782–4789.
[5] a) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275; b) J. Bidzinska, G. Cimino-Reale, N. Zaffaroni, M. Folini, *Molecules* **2013**, *18*, 12368–12395.
[6] C. Zhang, H. H. Liu, K. W. Zheng, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 7144–7152.
[7] K. W. Zheng, S. Xiao, J. Q. Liu, J. Y. Zhang, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 5533–5541.
[8] a) K. W. Zheng, Z. Chen, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2010**, *38*, 327–338; b) J. Y. Zhang, K. W. Zheng, S. Xiao, Y. H. Hao, Z. Tan, *J. Am. Chem. Soc.* **2014**, *136*, 1381–1390.
[9] D. Sun, L. H. Hurley, *Methods Mol. Biol.* **2010**, *608*, 65–79.
[10] a) K. W. Zheng, R. Y. Wu, Y. D. He, S. Xiao, J. Y. Zhang, J. Q. Liu, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2014**, *42*, 10832–10844; b) S. Xiao, J. Y. Zhang, J. Wu, R. Y. Wu, Y. Xia, K. W. Zheng, Y. H. Hao, X. Zhou, Z. Tan, *Angew. Chem. Int. Ed.* **2014**, *53*, 13110–13114; *Angew. Chem.* **2014**, *126*, 13326–13330.
[11] J. H. Kim, R. G. Larson, *Nucleic Acids Res.* **2007**, *35*, 3848–3858.
[12] N. Kumar, B. Sahoo, K. A. Varun, S. Maiti, S. Maiti, *Nucleic Acids Res.* **2008**, *36*, 4433–4442.
[13] B. F. Kahl, M. R. Paule, *Methods Mol. Biol.* **2009**, *543*, 73–85.
[14] C. Liu, C. T. Martin, *J. Mol. Biol.* **2001**, *308*, 465–475.
[15] S. Mukherjee, L. G. Brieba, R. Sousa, *Cell* **2002**, *110*, 81–91.
[16] C. Broxson, J. Beckett, S. Tornaletti, *Biochemistry* **2011**, *50*, 4162–4172.
[17] L. G. Brieba, R. Sousa, *EMBO J.* **2001**, *20*, 6826–6835.
[18] W. T. McAllister, *Cell. Mol. Biol. Res.* **1993**, *39*, 385–391.

Received: April 21, 2015
Published online: June 12, 2015

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase**

*Jia-quan Liu, Shan Xiao, Yu-hua Hao, and Zheng Tan**

anie_201503648_sm_miscellaneous_information.pdf

Supporting Figures

Figure S1

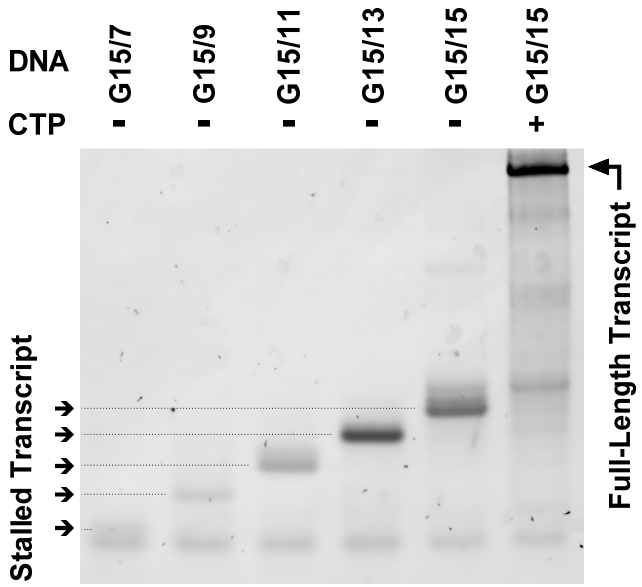

Figure S1. RNA transcripts obtained in partial and full-length transcription of dsDNA. Partial transcription was performed using the indicated dsDNA with T7 RNAP and ATP, GTP, and TTP. Full-length transcription was conducted in the same way except that CTP was supplied. RNA transcripts were resolved on a denaturing gel and visualized by staining with SYBR Gold. Stalling of transcription was reflected by the decrease in migration rate and increase in staining intensity as stall transcripts increased in length.

Figure S2

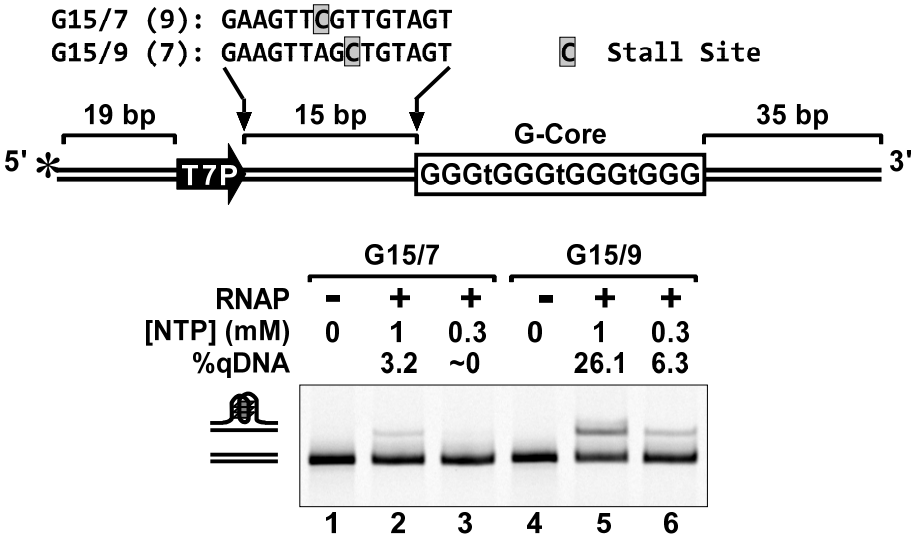

Figure S2. Effect of NTP concentration on the formation of G-quadruplex in partially transcribed dsDNA bearing a $G_3(TG_3)_3$ core on the non-template strand. DNAs were transcribed at the indicated concentration of NTP and the samples were subsequently processed as in Figure 2 in the main text. The number in parentheses in the scheme indicates the distance from the stall site to the 5' end of the G-core. %qDNA shows the amount of G-quadruplex-bearing DNA as percentage of the total DNA in the lane.

Figure S3

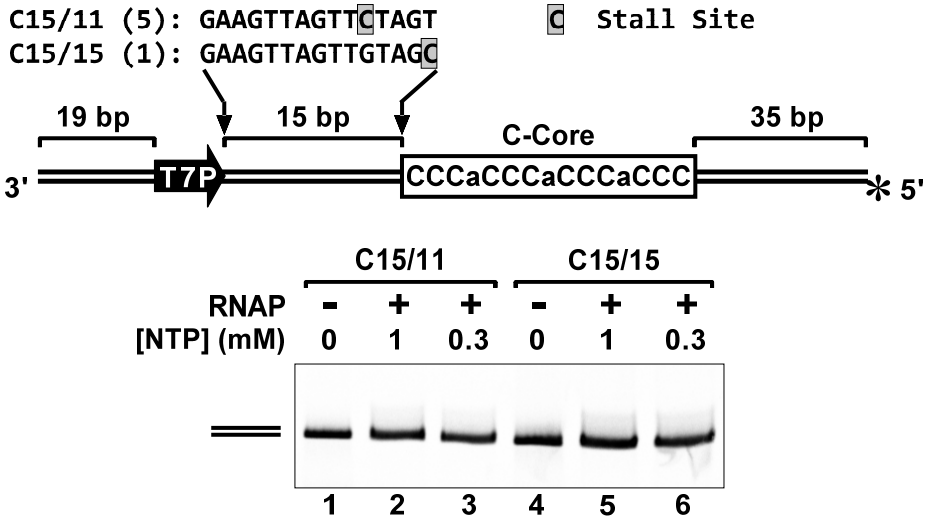

Figure S3. Effect of NTP concentration on the formation of G-quadruplex in partially transcribed dsDNA bearing a $G_3(TG_3)_3$ core on the template strand. DNAs were transcribed at the indicated concentration of NTP and the samples were subsequently processed as in Figure 3 in the main text. The number in parentheses in the scheme indicates the distance from the stall site to the 5' end of the G-core.

Figure S4

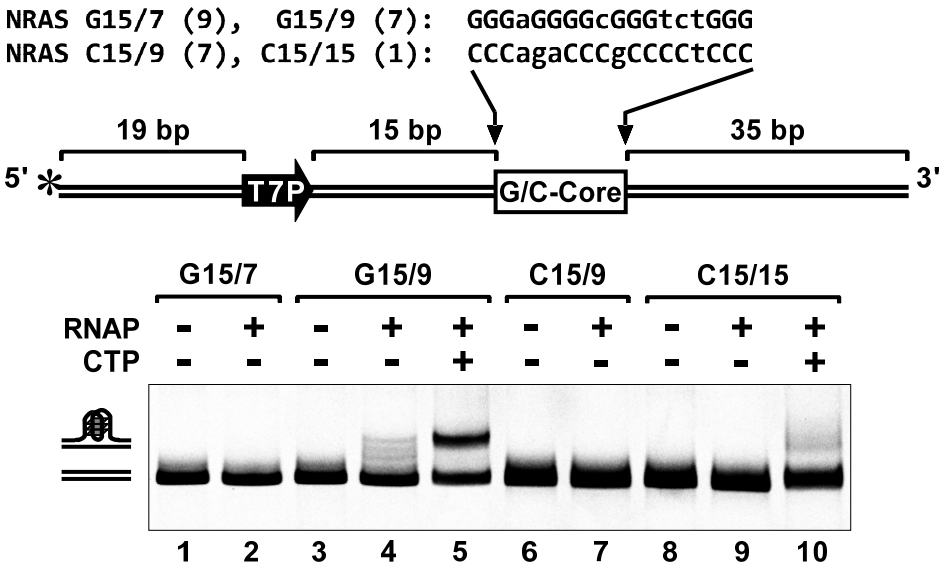

Figure S4. Approaching T7 RNAP induced formation of G-quadruplex on the non-template, but not on the template DNA strand in transcribed dsDNA carrying a G-core ($G_3aG_4cG_3tctG_3$) from the *NRAS* gene. DNAs were transcribed in the presence of ATP, GTP, and UTP with or without a supply of CTP and the samples were subsequently processed as in Figure 2 in the main text. The multiple slow-migrating bands in lane 4 may indicate structural isoforms of G-quadruplexes. The sequences of the NRAS DNAs are the same as the equivalent DNAs carrying the $G_3(TG_3)_3$ core, but with the G/C-core being replaced.

Figure S5

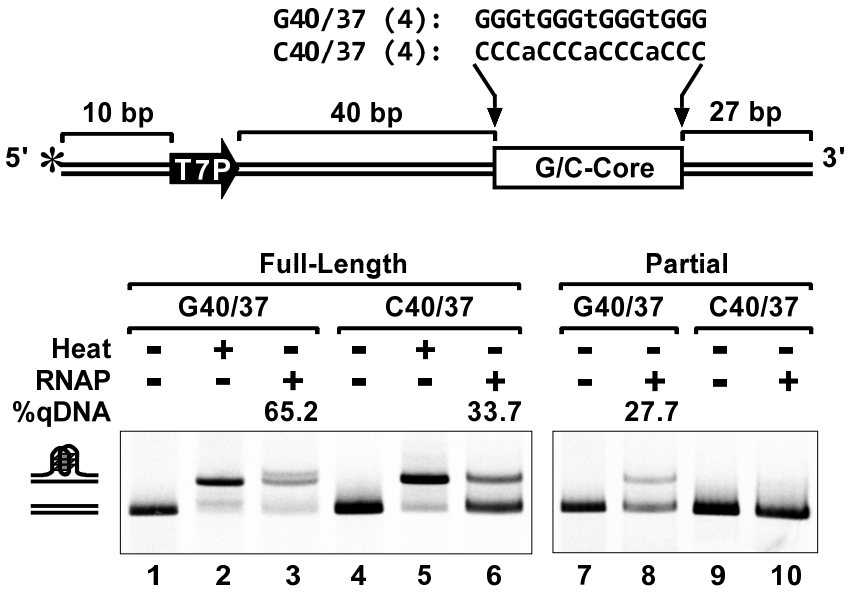

Figure S5. G-quadruplex formation in full-length and partial transcription of dsDNA with a $G_3(TG_3)_3$ core placed 40 nts away downstream of a T7 promoter either on the non-template (G40/37) or template (C40/37) DNA strand. DNAs were transcribed and processed as in Figure 1 (full-length) or 2 (partial) in the main text in the presence of four or three NTPs. The number in parentheses in the scheme indicates the distance from the stall site to the 5' end of the G-core. The number above the native gel indicates the amount of G-quadruplex-bearing DNA (qDNA) as percentage of the total DNA in the lane. The scheme on the left side of the gel indicates the structure of the DNAs.

DNA Sequences

Random-1

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTCTAGTGCTG
GCTCCGTCGCTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

Random-2

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTGTATGTcTCT
GTCTGTATGCTTGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/5

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGcTAGTTGTAGTGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/7

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTcGTTGTAGTGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/9

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGcTGTAGTGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/11

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTcTAGTGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/13

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTGTcGTGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G15/15

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTGTAGcGGGT
GGGTGGGTGGGTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

C15/5

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGcTAGTTGTAGTCCCA
CCCACCCACCCTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

C15/11

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTcTAGTCCCA
CCCACCCACCCTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

C15/15

TGGAGTCCCCTGCAGCGATTAATACGACTCACTATAGAAGTTAGTTGTAGcCCCA
CCCACCCACCCTAGTGCTATGATGCGTTCGATCACTCCATGTGATC

G40/37

TGCTGACAGCTAATACGACTCACTATAGGAGACACGAGACAGCAAGCGAGCAGC
GAAGCACAGtttAGGGTGGGTGGGTGGGTAGATCATGATGCAAGAATCACGAATG

C40/37

TGCTGACAGCTAATACGACTCACTATAGGAGACACGAGACAGCAAGCGAGCAGC
GAAGCACAGtttACCCACCCACCCACCCTAGATCATGATGCAAGAATCACGAATG

Red: T7 promoter; Green: transcribed region; Blue: G/C-core; c, ttt: stall site.

Experimental Section

Transcription

Oligonucleotides were obtained from Sangon Biotech (Shanghai, China). One μM of dsDNA was annealed in 10 mM $LiAsO_2(CH_3)_2$ buffer containing 50 mM LiCl by heating at 95 °C for 10 min and then cooling down to room temperature at a rate of 0.03 °C per second. Fifty nM annealed dsDNA was transcribed with 1.6 U/μl T7 RNA polymerase (Fermentas, MBI) in transcription buffer (40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 50 mM KCl and 40% PEG 200) in the presence of 1 mM GTP, 1 mM ATP and 0.2 mM UTP (partial transcription) or 1 mM NTP (full-length transcription) at 37 °C for 30 min.

Native gel electrophoresis of DNA

Transcription was stopped by incubating with 2 μM T7 promoter-competitive DNA (1,2) and 5 U RNase H at 37 °C for 30 min followed by incubating with 0.8 mg/ml RNase A and 25 mM EDTA at 37 °C for 30 min. Then the samples were treated with 0.5 mg/ml Proteinase K (Fermentas, MBI) at 37 °C for 30 min and loaded on a 8% polyacrylamide gel containing 75 mM KCl, 40% (w/v) PEG 200 and electrophoresed at 4 °C in 1× TBE buffer containing 75 mM KCl. Gels were scanned on a Typhoon 9400 phosphor imager (GE Healthcare, USA).

DMS footprinting

Fifty nM dsDNA was transcribed in 200 μl volume as aforementioned and subjected to DMS footprinting as described (3).

KMnO4 footprinting

Fifty nM dsDNA was transcribed in 200 μl volume as described (4,5). The samples were mixed with 3.75 mM potassium permanganate (KMnO4) and incubated for 2 min at room temperature. The reaction was stopped by addition of stop buffer (0.1 M β-mercaptoethanol,

10 μg sperm DNA, 40 mM EDTA and 0.2% SDS). After phenol/chloroform extraction and ethanol precipitation, the DNA was dissolved in 100 μl 10% piperidine and heated at 90 °C for 30 min. After another phenol/chloroform extraction and ethanol precipitation, the DNA was dissolved in 80% (v/v) deionized formamide in water, denatured at 95 °C for 5 min and resolved on a denaturing 12% polyacrylamide gel.

Analysis of putative G-quadruplex forming motifs in T7 phage genome

T7 genome sequence was downloaded from the NCBI website (www.ncbi.nlm.nih.gov) in fasta format with a GenBank id of V01146.1. It was searched for putative G-quadruplex forming motifs using the program from reference (6).

Denaturing gel electrophoresis of RNA

After transcription, the samples were extracted with equal volume of phenol/chloroform (1/1, v/v), followed by addition of formamide to 75%. The sample was denatured at 95 °C for 5 min, resolved on a 12% denaturing polyacrylamide gel, and stained with SYBR Gold. Gels were scanned on a Typhoon 9400 phosphor imager (GE Healthcare, USA).

References

1. Ujvari, A. and Martin, C.T. (1997) Identification of a minimal binding element within the T7 RNA polymerase promoter. *J. Mol. Biol.*, **273**, 775-781.

2. Zhang, J.Y., Zheng, K.W., Xiao, S., Hao, Y.H. and Tan, Z. (2014) Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.*, **136**, 1381-1390.

3. Zheng, K.W., Chen, Z., Hao, Y.H. and Tan, Z. (2010) Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res.*, **38**, 327-338.

4. Kahl, B.F., Li, H. and Paule, M.R. (2000) DNA melting and promoter clearance by eukaryotic RNA polymerase I. *J. Mol. Biol.*, **299**, 75-89.

5. Huang, J. and Sousa, R. (2000) T7 RNA polymerase elongation complex structure and movement. *J. Mol. Biol.*, **303**, 347-358.

6. Xiao, S., Zhang, J.Y., Zheng, K.W., Hao, Y.H. and Tan, Z. (2013) Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals. *Nucleic Acids Res.*, **41**, 10379-10390.