G-Quadruplexes

DOI: 10.1002/ange.201408719

# Formation of DNA:RNA Hybrid G-Quadruplex in Bacterial Cells and Its Dominance over the Intramolecular DNA G-Quadruplex in Mediating Transcription Termination**

*Ren-yi Wu, Ke-wei Zheng, Jia-yu Zhang, Yu-hua Hao, and Zheng Tan**

***Abstract:*** *DNA with four guanine tracts can fold into G-quadruplexes that are targets of transcription regulation. We recently found that hybrid DNA:RNA G-quadruplexes (HQs) can form during in vitro transcription. However, it is unclear whether they can form in cells. Evidence is presented supporting their formation in plasmids in bacterial cells. The formation of the HQs is indicated by a unique pattern of prematurely terminated transcripts under two conditions where the RNA transcripts do or do not participate in G-quadruplex assembly and further supported by a number of chemical and biochemical analysis. HQs dominate over the intramolecular DNA G-quadruplexes (DQ) in mediating the transcription termination when both structures are able to form. These findings provide the first evidence of HQ formation in cells and suggest that the competition/conversion between HQ and DQ may regulate transcription and serve as drug target in pharmaceutical applications.*

G-quadruplex-forming motifs are widely present in genomes, and G-quadruplexes are implicated in crucial physiological and pathological processes such as gene expression regulation, DNA replication initiation, DNA recombination, genome instability, and telomere maintenance.[1] Therefore, G-quadruplexes are considered important therapeutic targets for diseases such as cancer.[2] Hybrid DNA:RNA G-quadruplexes (HQs) are a new type of G-quadruplexes we recently found to form in DNA during in vitro transcription.[3] Because HQs can form with as few as two G-tracts on the non-template DNA strand instead of four for a intramolecular DNA G-quadruplex (DQ), potential HQ-forming motifs are more prevalent and abundant than the DQ-forming motifs in genes. Their unique distribution in animal genomes and correlation with transcription suggest that HQs regulate transcription.[3b] Herein we address the question whether HQ can form in cells and how it competes with the DQ in their formation and effect on transcription. We observed a formation of HQ in plasmids in *Escherichia coli* (*E. coli*) cells, and our data show that HQ dominates over DQ in mediating premature termination (PT) of transcription.

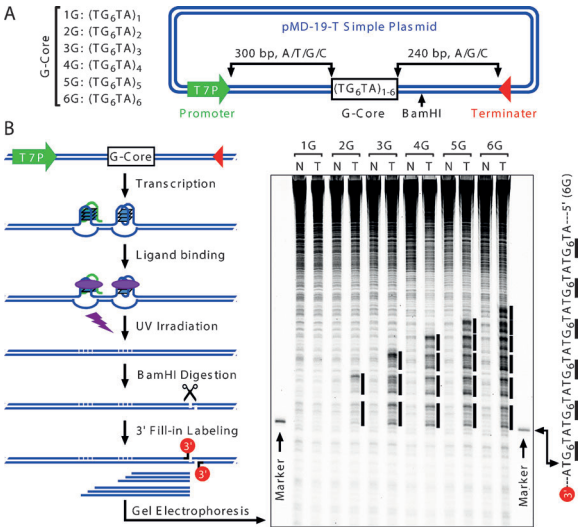


***Figure 1.*** G-quadruplex formation in plasmids detected by ligand-induced photocleavage. Arrays of increasing number of $G_6$ tracts were inserted into a modified pMD19T-simple vector between a T7 promoter and a VSV terminator. Transcribed plasmids were incubated with a ligand Zn-TTAPc, followed by UV irradiation to cleave the G in the G-quadruplexes. The plasmids were then cut at the proximal BamHI site downstream of the $G_6$ tracts and labeled by filling-in with fluorescein-12-dUTP. The cleavage fragments were resolved on a denaturing gel and separated from the opposite labeled full-length DNA strand owing to their large difference in size. The N and T above the lanes denote no transcription and transcription, respectively. Solid bars on gel indicate cleavage at the $G_6$ tracts. The drawings of G-quadruplexes in this and other figures are only symbolic and do not intend to indicate their folding topologies.

We made plasmids bearing 1–6 repetitive $G_6$ tracts between a T7 promoter and a terminator (Figure 1 A). The $G_6$ tract was chosen because it forms more stable G-quadruplexes and causes stronger PT than shorter G-tracts, thus leading to greater detection sensitivity but still having the ability to function as a single G-tract. A G-tract with more than six G tends to break into two G-tracts, using the G in the middle as a loop.[3a] On the other hand, 535 motifs of such tandem G-tracts with a minimum of six G were found in 298 bacterial genomes. G-quadruplex formation in the plasmids transcribed in vitro with T7 RNA polymerase (RNAP) was first detected by photocleavage using a ligand Zn-TTAPc.[4] This ligand specifically binds to G-quadruplex and cleaves the G within it upon UV irradiation. After a linearization by a restriction enzyme and labeling with a fluorescent dye

[*] R.-y. Wu, Dr. K.-w. Zheng, J.-y. Zhang, 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

[**] This work was supported by Grants 2013CB530802, 2012CB720601 from MSTC and 21072189 from NSFC.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201408719.

(Figure 1B, scheme), the DNA were analyzed by denaturing gel electrophoresis. As shown in Figure 1B (right), starting from the 2G plasmid, cleavage signal was found at each $G_6$ tract, indicating that G-quadruplex formed in the 2G-6G plasmids. The formation of G-quadruplexes in the 2G-6G plasmids was verified by restriction enzyme cleavage at the two restriction sites, NcoI and KpnI, arranged at the 5′ and 3′ side of the G-cores, respectively (Supporting Information, Figure S1A). This was indicated by the circular plasmid survived (Figure S1B, lanes 4–8, 12–16) and fully agrees with the results in Figure 1B. The presence of G-quadruplex in the 2G and 3G plasmids suggested a formation of HQ in at least these two plasmids because they are unable to form DQ.

Whereas the results in both Figure 1 and the Supporting Information, Figure S1 suggested a formation of HQ in the 2G and 3G plasmids, it was not clear if HQ also formed in the 4G-6G plasmids because they are able to form DQ. To find this out, we examined if RNA was present in the G-quadruplexes in these plasmids. RNA in a HQ is resistant to RNases T1, A, and H that digest the different forms of RNA except those in the HQ.[3a] Transcribed plasmids were first treated with the three RNases, and then the RNA cleavage fragments were removed by filtration. The plasmid concentrate was mixed with a FAM-labeled C-rich DNA oligomer (5′-ACCCCCCAT-3′) complementary to the G-core, denatured and re-natured, and the samples were resolved on a native gel (Figure 2A). The RNA released from a HQ is

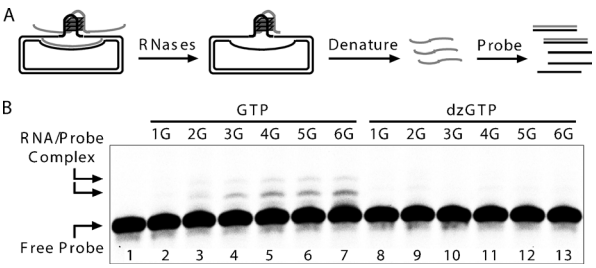


***Figure 2.*** Detection of RNA in HQ by DNA probe shift. Transcribed plasmid was treated with RNases T1, A, and H to digest the single-stranded RNA and R-loop. After the removal of the RNA cleavage fragments by filtration on a 50 kD cut-off filter, the RNAs protected in the HQs were released by denaturation, hybridized with a FAM-labeled DNA probe, and resolved on a native gel.

expected to hybridize with the DNA and retard its migration. In Figure 2B (lanes 3–7), two slower-migrating bands appeared in the 2G-6G plasmids transcribed with GTP, implying that HQ formed in these plasmids. The two slower-migrating bands might present an involvement of different number of RNA G-tracts. To verify that the RNA indeed came from HQ, we transcribed the plasmids with a substitution of 7-deaza-GTP (dzGTP) for GTP. Because the N7 is replaced by a carbon in dzGTP, the $G_6$ tracts in the RNA transcripts were prevented from forming G-quadruplex.[5] As a result, no RNA fragments were detected (Figure 2B, lanes 9–13), thus confirming the formation of HQ in the 2G-6G plasmids.

G-quadruplex on the non-template DNA strand causes transcription termination.[3d,6] Because a HQ can form with a minimum of two G-tracts on the DNA instead of four for a DQ, a G-quadruplex-forming motif will form more HQs than DQs and, as a result, creates more transcription termination sites. Therefore, the type of G-quadruplexes formed in transcription can be determined from the number of termination sites. To check this out, we first transcribed the plasmids using dzGTP to inhibit the participation of RNA. In this case, only DQ could form in those plasmids bearing at least four $G_6$ tracts, but not in those with only 1–3 G-tracts. The transcription clearly showed premature terminations (PT) at the G-core in the plasmids containing 4–6 $G_6$ tracts, but not in those bearing less than four G-tracts as expected (Figure 3, left half gel, filled arrowheads). The number of

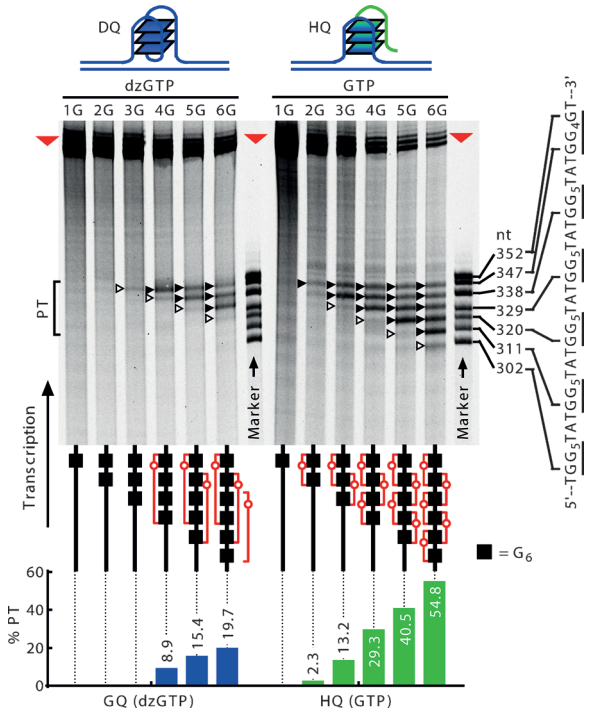


***Figure 3.*** Premature termination (PT) of transcription in plasmids indicates a formation of HQ. Transcription was spiked with 1/20 fluorescein-12-UTP to visualize the transcripts on a denaturing gel. The region between the G-core and terminator was T-free such that PTs can be compared with the full-length transcripts (red arrowhead) without length calibration. The red brackets in the schemes below the gel show combination of $G_6$ tracts for the formation of DQ or HQ to produce the corresponding PTs (filled arrowheads on gel). Circle indicates positions of major PT. Open arrowheads indicate minor PTs possibly caused by minor HQ structures with a 3/1 (left half gel) or 1/3 (right half gel) DNA/RNA G-tract combination. Graph at the bottom shows PT as % of the total transcripts. Sizes of markers are at the right side of the gel.

major PT bands correlated with the number of positions at which a DQ could form (red brackets, scheme under dzGTP lanes). For example, two major PT bands were detected for the 5G plasmid in which a DQ could form using either the four $G_6$ tracts at the 5′ side or those at the 3′ side of the G-core (scheme under dzGTP/5G lane).

We then transcribed the plasmids using normal GTP, which permits the $G_6$ tract in the RNA to participate in G-

quadruplex assembly. In this case, HQ is able to form starting from the 2G plasmid, which can use two $G_6$ tracts from the DNA and two from the RNA (Figure 3, scheme under GTP/2G lane). In agreement with this, a major PT band appeared for the 2G plasmid (GTP/2G lane, filled arrowhead). For the 3G plasmid, the two major PT bands implied a formation of two HQs that used the two $G_6$ tracts either from the 5′ or the 3′ side of the G-core (scheme under GTP/3G lane). As the number of $G_6$ tracts increases, the number of positions (red brackets under GTP lanes) at which such a HQ can form will also increase. Indeed, an increment in the number of PT bands was observed (right half of the gel, filled arrowheads). Comparison of the PT bands with the markers suggests that PT occurred at or before the 3′-most DNA G-tract of a HQ/DQ. This feature explains why all the PTs at the furthest 3′ end aligned at the same position no matter if the transcription was conducted with dzGTP or normal GTP.

Besides the 2/2 DNA/RNA G-tract combination, a HQ, in principle, could also involve three $G_6$ tracts from the DNA and one from the RNA (3/1 DNA/RNA), or alternatively, one from the DNA and three from the RNA (1/3 DNA/RNA). The 3G plasmid, for example, would produce one PT band in the former case, but three PT bands in the latter case. The appearance of two major PT bands for the 3G plasmid suggested that the 2/2 DNA/RNA G-tract combination might be the preferred option. Similar inferences should also hold true for the other plasmids transcribed with GTP (GTP/4G-6G lanes).

Transcription of G-rich DNA results in a formation of R-loop, a nascent RNA transcript that remains annealed with the template DNA strand after its synthesis.[7] We recently showed that HQ formation relies on the RNA transcript displaced from R-loop, such that a HQ could only start to form at the second or later rounds of transcriptions.[3a] In the aforementioned experiments, the substitution of dzGTP for GTP prevented the RNA from forming HQ with DNA. To ensure that the PT pattern with dzGTP was caused by the lack of RNA participation, a single-round transcription with GTP was conducted. We supplied only GMP, ATP, UTP, and CTP to load a T7 RNAP to the +13 position before the first C on the template DNA strand; then GTP was added together with a competitive DNA for the transcription to proceed downstream but for only once. The competitive DNA prevented the RNAP from re-initiating transcription.[3a] Under this condition, the RNA could only form an R-loop, but not HQ. As a result, the PT pattern (Figure 4, GTP, 1G-6G) followed that in the multiple-round transcription with dzGTP (Figure 3, left half gel). This result thus verified that the PT in the multiple-round transcription with GTP (Figure 3, right half gel) was the result of HQ formation that depended on the supply of RNA. Furthermore, the major PT bands in the single-round transcription were not followed by a faint faster-moving band as those in the multiple-round transcriptions in Figure 3 (open arrowheads). We speculate that, in the multiple-round transcription, the faint bands might represent minor terminations caused by a 3/1 DNA/RNA G-tract combination when dzGTP was used or a 1/3 DNA/RNA G-tract combination when GTP was used. In the case of dzGTP, the 3/1 combination might form a less-stable HQ in which the

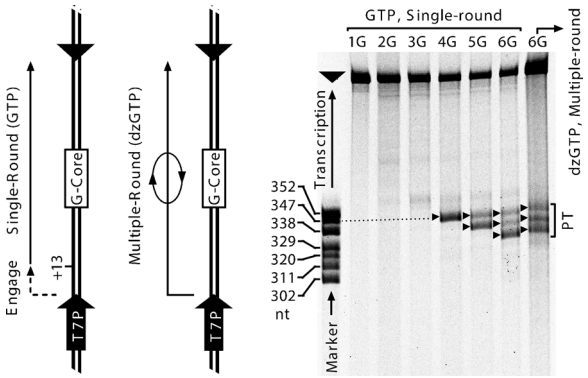


***Figure 4.*** Premature termination (PT) in single-round transcription with GTP and multiple-round with dzGTP confirms HQ formation in multiple-rounds of GTP transcription. For the single-round transcription, a RNAP was first loaded to +13 (dashed arrow in scheme) in the absence of GTP. It then proceeded downstream after the missing GTP and competitive DNA was supplied. Multiple-round transcription was conducted in the presence of all the four NTPs as in Figure 3. RNA transcripts were resolved on a denaturing gel.

G-quartet was stabilized by seven Hoogsteen hydrogen bonds instead of eight. In the latter case, minor PT bands by the 3/1 combination were masked by the major ones.

The course-effect connection between the formation of HQ and the PT (Figure 3 and Figure 4) was confirmed by a G-quadruplex-stabilizer-enhanced termination. We conducted transcription in solution containing $K^+$, $Na^+$, and $Na^+$ in the presence of a G-quadruplex stabilizer Zn-TTAPc.[4] $Na^+$ is a weaker stabilizer than $K^+$.[8] As a result, the overall PT was reduced in the $Na^+$ compared with that in the $K^+$ solution (Supporting Information, Figure S2, dotted bars in graph). However, it was dramatically enhanced by the Zn-TTAPc. This effect was more drastic for the PTs at the first G-quadruplex (filled bars). This cation species-dependent and stabilizer-responsive PT supported the formation of HQ as a cause of the transcription termination.

We attempted to examine if the HQ formation and PT could also occur in *E. coli* cells. We transformed *E. coli* with the 1G-6G plasmids which were transcribed inside the bacteria by an endogenously expressed T7 RNAP. The corresponding RNA transcripts were purified, converted into cDNAs, and amplified by PCR (Figure 5, scheme). The RNA transcripts from an in vitro transcription were processed in parallel to serve as references. The PCR product of the in vitro transcripts displayed a PT pattern (Figure 5, left half gel, filled arrowheads) similar to that of the original PT products (Figure 3, right half gel). An identical PT pattern was also observed for the PCR products of the in vivo transcripts (Figure 5, right half gel, filled arrowheads). These two results implied that a same PT of transcription occurred both in vitro and in vivo, therefore, supporting the formation of HQs in the plasmids in the *E. coli* cells.

Compared to the direct visualization of RNA transcripts (Figure 3, Figure 4; Supporting Information, Figure S2), the PCR products of RNA transcripts were accompanied with many non-specific bands (Figure 5). This could be attributed to the reduced specificity in the PCR in which only the

  

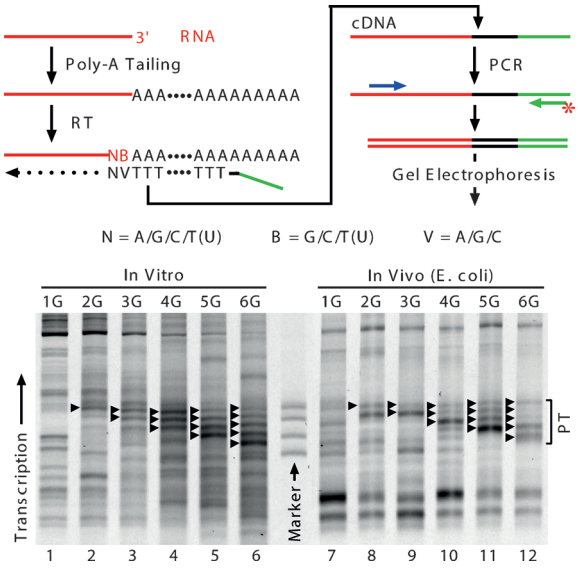


***Figure 5.*** PCR analysis of premature termination (PT) pattern in *E. coli* cells. Plasmids were introduced into the *E. coli* strain BL21 (DE3) and transcribed by an endogenously expressed T7 RNAP. The plasmids were also transcribed in vitro by T7 RNAP. The RNA transcripts were polyadenylated at the 3′ end and converted into cDNA using a downstream junction primer followed by PCR amplification using 7-deaza-dGTP to suppress G-quadruplex formation in DNA. PCR products were resolved on a native gel.

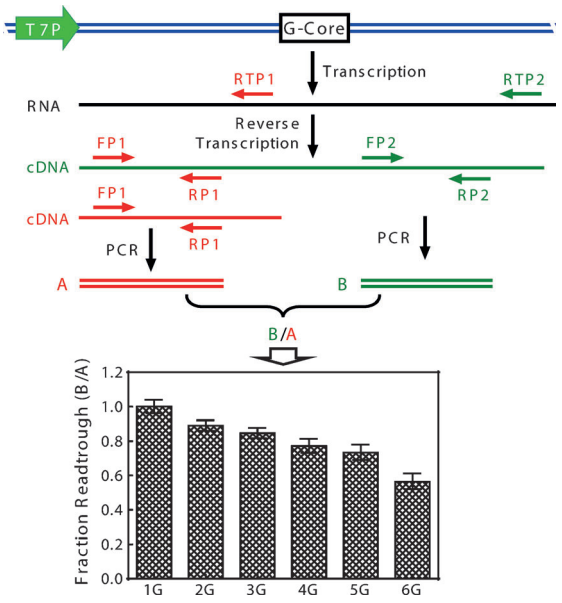


***Figure 6.*** Reading-through of transcription at the G-cores in *E. coli* cells. Plasmids were introduced into the *E. coli* and transcribed as in Figure 5. Transcripts were converted into cDNAs and quantitated by real-time PCR at the two regions indicated using the indicated primers (horizontal arrows).

upstream primer was specific to the PT transcripts. In the *E. coli* cells, the RNA transcripts could be modified by a variety of nucleases and processing enzymes to yield more complicated fragments that could be amplified, making the situation worse. It was noticed that the PCR products of the full-length in vivo transcripts stopped at the terminator did not show a decrease in intensity when the G-core changed from 1G to 6G as the corresponding in vitro transcripts did (Figure 5, lanes 7–12 versus 1–6). In *E. coli*, RNA is subject to exonucleolytic cleavages in the 3′ to 5′ direction.[9] This might spoil the uniformity of the full-length transcripts and ruin their detection by PCR. To find out how the transcription went through the G-cores in the *E. coli*, we amplified two segments of the transcripts upstream and downstream of the G-cores, respectively. In Figure 6, it can be seen that the amount of transcripts passed the G-cores decreased with an increase in the number of $G_6$ tracts. This conclusion was further verified by northern blotting (Supporting Information, Figure S3).

Our RNA markers were made by transcription using mutated G-core to avoid non-specific termination at G-quadruplexes. Therefore, the position of the PTs could not be precisely determined. We sequenced the prematurely terminated RNA transcripts (Supporting Information, Figure S4). For the 3G plasmid, major PTs occurred at the two G-tracts from the 3′ side under both the in vitro and in vivo conditions. This result agreed with that in the 3G lanes in Figure 3 (right half gel), Figure 5 and Supporting Information, Figure S2 and can be attributed to a HQ of 2/2 DNA/RNA combination at two alternative positions. For the 6G plasmid, PTs beyond the third G-tract from the 5′ side were rarely detected and the whole PT pattern resembled that of the 3G plasmid. Such difficult-to-detect transcripts could be due to their ability to form G-quadruplexes, which is known to suppress PCR amplification.[10] In both the 3G and 6G plasmids, minor PTs were seen at the 5′-most G-tract. All the PTs seemed to preferentially occur at the first two G of the G-tracts. Based on the results in Figures 3–5 and Supporting Information, Figures S2 and S4, the structures responsible to the observed PTs are summarized in Figure 7.

Although methods are becoming available for the detection of G-quadruplex in cells,[11] a method for directly distinguishing the type of G-quadruplexes, that is, HQ or DQ, is still unavailable. Our attempt to apply the in vitro methods to detect HQ in vivo ended up with little success. For example, the Zn-TTAPc and other ligands killed *E. coli* even at low concentrations. To this end, the PT pattern analysis provides a way to deduce a judgment and the results obtained offered the first evidence for the formation of HQ in cells. Furthermore, our data (Figure 3, Figure 6; Supporting Infor-

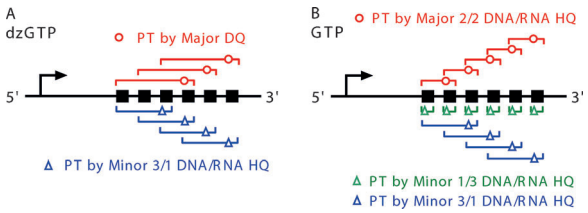


***Figure 7.*** Assumed structures responsible to the premature terminations (PTs) in the 6G plasmid transcribed with (A) dzGTP or (B) GTP. Black square indicates G6 tract. Bracket indicates the required DNA G-tracts to form a DQ or HQ. ○ major PT, △ minor PT positions. PT on left side corresponds to bottom on denaturing gel.

mation, Figures S2 and S3) show that transcription level can be controlled by the number of G-quadruplexes. Whereas the formation of HQ and PT in the 2G plasmid is marginal, they become efficient when the G-tracts increase to three or more (Figure 1, Figure 3; Supporting Information, Figure S2). As a result, the PTs increase as the number of G-tracts increases. Moreover, a much greater PT in the transcription with GTP than with dzGTP (Figure 3) indicates that the HQ dominated over the DQ in mediating the PT when both structures can, in principle, be formed. This could be the result of a dominant formation and/or a more robust effect of the HQs. This observation suggests that the structural competition and conversion between the HQ and DQ may play a role in the regulation of transcription, and as a result serve as a potential target in therapeutic applications.

Received: September 1, 2014
Revised: November 14, 2014
Published online: January 22, 2015

**Keywords:** bacteria · DNA:RNA hybrids · G-quadruplexes · termination · transcription

[1] a) N. Maizels, L. T. Gray, *PLoS Genet.* **2013**, *9*, e1003468; b) M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770 – 780.

[2] a) G. W. Collie, G. N. Parkinson, *Chem. Soc. Rev.* **2011**, *40*, 5867 – 5892; b) S. Balasubramanian, S. Neidle, *Curr. Opin. Chem. Biol.* **2009**, *13*, 345 – 353; c) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261 – 275.

[3] a) J. Y. Zhang, K. W. Zheng, S. Xiao, Y. H. Hao, Z. Tan, *J. Am. Chem. Soc.* **2014**, *136*, 1381 – 1390; b) K. W. Zheng, S. Xiao, J. Q. Liu, J. Y. Zhang, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 5533 – 5541; c) S. Xiao, J. Y. Zhang, K. W. Zheng, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2013**, *41*, 10379 – 10390; d) 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; e) 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.

[4] K. W. Zheng, D. Zhang, L. X. Zhang, Y. H. Hao, X. Zhou, Z. Tan, *J. Am. Chem. Soc.* **2011**, *133*, 1475 – 1483.

[5] T. M. Fletcher, D. Sun, M. Salazar, L. H. Hurley, *Biochemistry* **1998**, *37*, 5536 – 5541.

[6] S. Tornaletti, S. Park-Snyder, P. C. Hanawalt, *J. Biol. Chem.* **2008**, *283*, 12756 – 12762.

[7] A. Aguilera, T. Garcia-Muse, *Mol. Cell* **2012**, *46*, 115 – 124.

[8] C. C. Hardin, A. G. Perry, K. White, *Biopolymers* **2000**, *56*, 147 – 194.

[9] S. R. Kushner, *J. Bacteriol.* **2002**, *184*, 4658 – 4665; discussion page 4657.

[10] H. Han, L. H. Hurley, M. Salazar, *Nucleic Acids Res.* **1999**, *27*, 537 – 542.

[11] 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.

  