

# Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks

Boris P. Belotserkovskii<sup>1</sup>, Alexander J. Neil<sup>1</sup>, Syed Shayon Saleh<sup>1</sup>, Jane Hae Soo Shin<sup>1</sup>, Sergei M. Mirkin<sup>2</sup> and Philip C. Hanawalt<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Stanford University, Stanford, CA 94305, USA and <sup>2</sup>Department of Biology, Tufts University, Medford, MA 02155, USA

Received March 8, 2012; Revised November 27, 2012; Accepted November 28, 2012

## ABSTRACT

The ability of DNA to adopt non-canonical structures can affect transcription and has broad implications for genome functioning. We have recently reported that guanine-rich (G-rich) homopurine-homopyrimidine sequences cause significant blockage of transcription *in vitro* in a strictly orientation-dependent manner: when the G-rich strand serves as the non-template strand [Belotserkovskii *et al.* (2010) Mechanisms and implications of transcription blockage by guanine-rich DNA sequences., *Proc. Natl Acad. Sci. USA*, 107, 12816–12821]. We have now systematically studied the effect of the sequence composition and single-stranded breaks on this blockage. Although substitution of guanine by any other base reduced the blockage, cytosine and thymine reduced the blockage more significantly than adenine substitutions, affirming the importance of both G-richness and the homopurine-homopyrimidine character of the sequence for this effect. A single-strand break in the non-template strand adjacent to the G-rich stretch dramatically increased the blockage. Breaks in the non-template strand result in much weaker blockage signals extending downstream from the break even in the absence of the G-rich stretch. Our combined data support the notion that transcription blockage at homopurine-homopyrimidine sequences is caused by R-loop formation.

## INTRODUCTION

Structural properties of DNA sequences, in particular their ability to adopt non-B form DNA conformations (e.g. quadruplex DNA, triplex DNA, Z-DNA and

others) can affect DNA transcription, and that has been implicated in many biologically important phenomena, including mutagenesis, recombination and transcription-coupled DNA repair [reviewed in (1–4)].

DNA sequences in which the non-template strand is enriched with guanines, especially with contiguous guanine stretches, have recently received increased attention owing to their propensity to form stable extended RNA/DNA hybrids (R-loops) or other stable DNA/RNA complexes during transcription (5–10). Yu *et al.* (11) directly demonstrated that such complexes are present in mammalian cells and that their lengths can exceed 1 kb.

The stability of DNA/RNA complexes generated at these sequences is often ascribed to the formation of guanine quadruplexes (G-DNA) in the non-template DNA strand that prevent RNA displacement in the course of transcription (7–9,12). Various types of triplex structures, including H-DNA-like structures stabilized by the transcript (13–16) and parallel triplexes with RNA as the central strand (13,17), have also been suggested to form during transcription of guanine-rich (G-rich) homopurine sequences. Furthermore, a triplex containing G-rich RNA was recently implicated in gene repression by a non-coding transcript (18). Note, however, that a hybrid duplex formed by the G-rich homopurine RNA and its complementary DNA, Pu-RNA/Py-DNA, is much more stable than the corresponding Pu-DNA/Py-DNA or Py-RNA/Pu-DNA duplexes (19–21). Thus, this difference in stability alone could be sufficient for the R-loop formation (22,23).

A propensity of certain DNA regions to form R-loops has been correlated with the obstruction of transcription in these regions. For example, Daniels and Lieber (5) observed an overall reduction of transcription *in vitro* through the immunoglobulin switch region, which is prone to R-loop formation, and Rajagopal *et al.* (24) observed RNA polymerase pausing in this same region in mammalian cells.

\*To whom correspondence should be addressed. Tel: +1 650 723 2424; Fax: +1 650 725 1848; Email: hanawalt@stanford.edu

An important consequence of stalled transcription complexes and extra-stable RNA/DNA hybrids is replication fork blockage [reviewed in (25,26)]. For example, (G)<sub>n</sub>•(C)<sub>n</sub> repeats stall replication in *Escherichia coli* when transcribed in the direction in which G-stretches are placed on the non-template strand. It was suggested that these repeats first arrest transcription, which is then followed by the collision of the replication machinery with the stalled transcription complex (27). Supporting this idea, we have shown that G-rich homopurine-homopyrimidine sequences block T7 RNAP transcription *in vitro* in the orientation in which the non-template strand is homopurine, but not in the opposite orientation (28). Our initial observations were most consistent with the model that formation of R-loops, which increased the probability of transcription pausing or termination, could be central for the effect (28,29). In the present work, we elaborate and provide further support for the R-loop model by studying the effects of various DNA base substitutions and strand breaks [which were shown to facilitate R-loop formation (30)], on transcription blockage caused by homopurine-homopyrimidine sequences.

## MATERIALS AND METHODS

### DNA substrates

A general scheme for constructing our transcription substrates is shown in Figure 1. DNA sequences (shown below) were cloned between the BamHI and XhoI sites of the pUC GTG-TS plasmid (31). These sequences usually contain the sequence of interest (e.g. homopurine-homopyrimidine repeats), which we refer to as an 'insert', and flanking sequences, which might contain nicking sites. In all sequences, the non-template (sense) DNA strand is shown in the 5'-to-3' direction, and the first nucleotide is localized 236-nt downstream from the start of transcription.

Plasmid pG32 (28) contains an insert with 32 guanines in a row (bold).

GATCGGTACCTCTAGA(G)<sub>32</sub>TCTGCACCGTGG

All sequence modifications of this insert are shown in Figure 2.

Plasmids pN-aga-(insert), pN-tc-(insert)-cc-N and pN-(9)-(insert)-(9)-N contain the sequences:

GATCGGCTTCTTCT↓AGA(insert)TCTGCACCGTGG,  
GATCGCTTTC↓TC(insert)CC↓TCAGCG

and

GATCGCTTTC↓CGGGCACGT(insert)TGCGGGC  
CC↓TCAGCG,

respectively. The recognition sequences for nicking endonucleases Nt.BspQI and Nt.BbvCI, located 5' (upstream) or 3' (downstream) from the insert, respectively, are shown in bold italic, whereas nicking sites are shown by arrows. The sequences and positions of inserts are also shown in the respective figures or figure legends. (In the names of the plasmids, regular, instead of subscript, font

is used to indicate the number of nucleotides (e.g. G16 stands for G<sub>16</sub>).

Plasmids pN-NTS and pN-TS contain the sequences

GATCGGTACCTCTAGACC↓TCAGCTCTGCACCG  
TGG

and

GATCGGTACCTCTAGAGCTGA•GGTCTGCACCG  
TGG,

respectively.

An Nt.BbvCI site (shown in bold italic) was cloned in opposite orientations to ensure that either the non-template (pN-NTS) or the template strand (pN-TS) is cleaved (The nicking site on the non-template strand is shown by an arrow, and the site on the non-template strand opposite to the nicking site on the template strand is shown by a dot).

Transcription substrate preparation, *in vitro* transcription and data analysis were performed, as previously described (28,32,33) (See also Supplementary Materials and Methods for details).

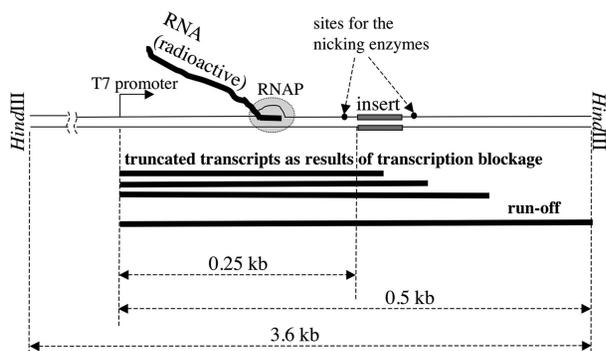
## RESULTS

### Experimental strategy

The basic scheme of the *in vitro* transcription experiments performed in this work is shown in Figure 1. A DNA sequence of interest, further referred to as 'insert' was positioned 0.25-kbps downstream from the T7 promoter in a plasmid. In some cases, an insert was flanked by the sites for nicking enzymes (The actual sequences are shown in Materials and Methods, and in the corresponding figures). To obtain linear transcription templates, the plasmids were digested by the *Hind*III restriction enzyme 0.5-kbps downstream from the promoter. Transcription was performed in the presence of all four NTPs plus a smaller amount of α-<sup>32</sup>P-CTP, resulting in transcripts that were uniformly labeled over their lengths. Transcripts were analysed by denaturing polyacrylamide gel-electrophoresis.

In the absence of transcription blockage, the only transcript, referred to as 'run-off', appears when RNA polymerase reaches the end of the linear substrate. When RNA polymerase (RNAP) is blocked before reaching the end of the template, shorter, truncated transcripts appear, and their lengths reflect the distance from the promoter to the blockage site. The ratio of intensities of truncated transcript(s) to the run-off band characterizes the probability of blockage (As our RNA is 'body-labeled' with radioactive cytosine, our calculations were normalized on the number of cytosines within the transcript sequence).

For circular plasmids, that do not have strong termination signals, transcription is expected to continue until spontaneously terminating at some random sites, to produce a heterogeneous mixture of fairly long RNA products. In reality, however, this mixture of long RNA fragments forms a distinct band on a gel, the intensity of which was used as a normalization factor in the calculation of relative blockage.



**Figure 1.** Experimental strategy. DNA strands are shown as thin black lines, except within a specific insert (e.g. homopurine-homopyrimidine sequence) where they are shown by thick gray lines; RNA strand is shown as a thick black line; RNA polymerase (RNAP) is shown as a gray oval with dotted borders. The distance from the transcription start site to the insert is  $\sim 0.25$  kb; the sites for nicking enzymes are localized 2–9 nt from the insert. In most experiments, HindIII-digested templates were used, which produce run-off  $\sim 0.5$  kb.

Absolute values of blockage for circular substrates could be estimated using a ribozyme-based approach developed by Grabczyk and Usdin (16), as previously described (28).

### Roles of G-richness and homopurine-homopyrimidine character of the sequence in blockage

Figure 2 (left panel) shows the results of transcription from linear DNA templates with various insert sequences. Here and later in the text, the sequence given is that of the non-template strand. For the C32 insert (lane 1), the only well-pronounced signal corresponds to the full-length transcription product (run-off). However, for a series of homopurine inserts (lanes from 2 to 10), additional, truncated transcription products appear, and their intensities increase with increased G-richness (see graphs in Figure 2). The most pronounced truncated product (indicated by the white block arrow) reflects transcription blockage closely downstream from the insert; thus, we call it the ‘repeat-exiting signal’. In addition, weaker multiple blockage sites are localized within the insert that altogether form the ‘diffuse signal’ (white oval). When A is replaced by either T or C, the blockage is strongly reduced (compare lane 8 with lanes 11 and 13, and lane 9 with lanes 12 and 14). Thus, both G-richness and the homopurine-homopyrimidine character of the sequence are important for blockage.

When supercoiled DNA templates were used (Figure 2, right panel), the relative strengths of blockage signals for the various inserts were similar to those in linear templates, although the overall blockage was more pronounced. These results confirm our previous report that negative supercoiling facilitates blockage (28).

### A single-strand break (nick) in the non-template strand upstream of an insert dramatically increases the blockage

Based on these and previous results (28), we suggest that transcription blockage at G-rich repeats could be caused by R-loop formation (see Discussion). Roy *et al.* (30)

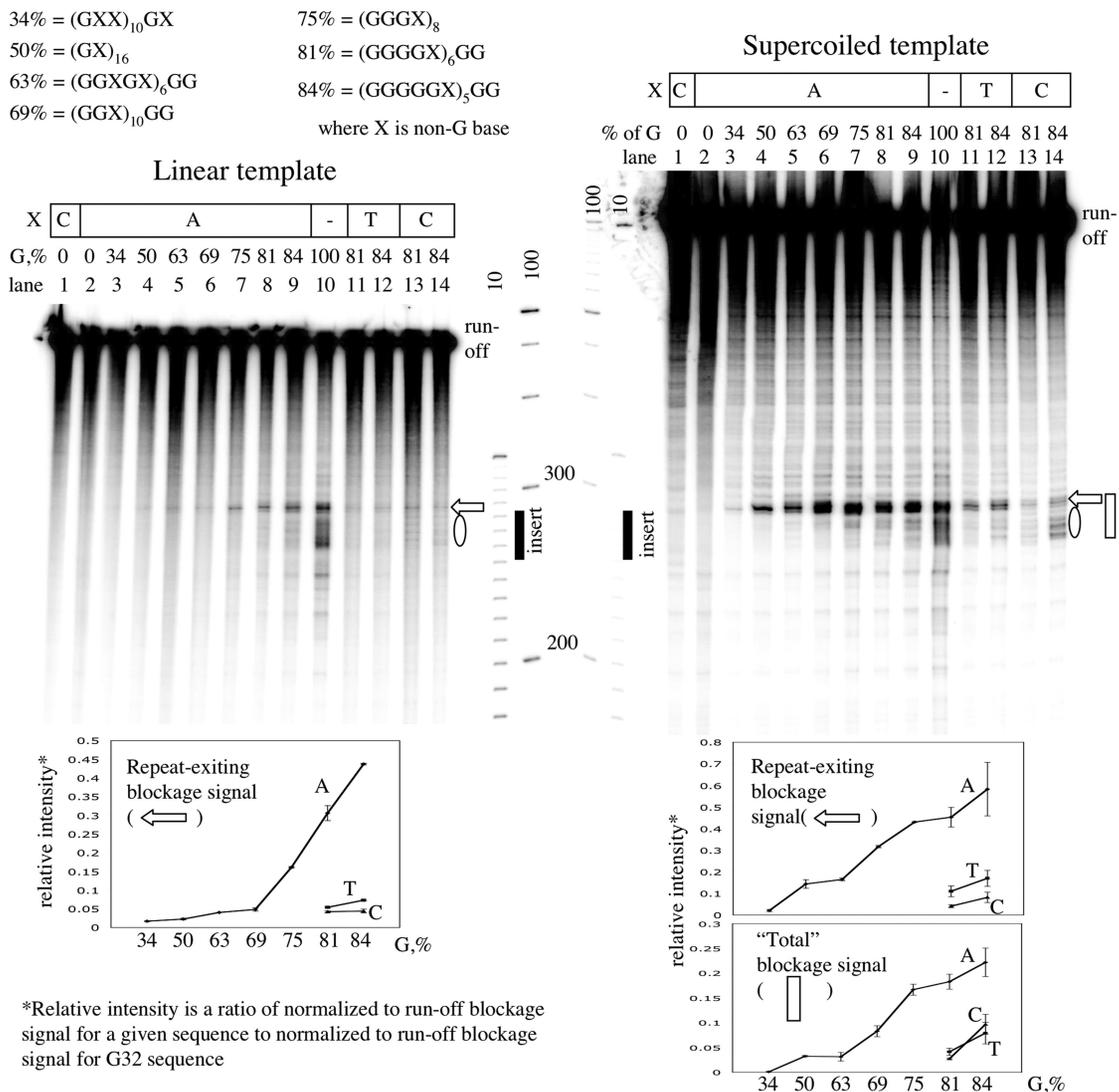
previously showed that R-loop formation is strongly facilitated by a nick in the non-template strand; one would expect, therefore, that a nick in the non-template strand would exacerbate blockage produced by G-rich inserts. To test this prediction, we placed a nicking site in the non-template strand three nucleotides (AGA) upstream from either the G16 or the C16 insert, or in the parental plasmid without any special insert, creating pN-aga-G16, pN-aga-C16 and pN-aga-0 plasmids (see Materials and Methods).

Figure 3, lanes 1–6, shows the effect of a nick in linear DNA templates under our standard transcription conditions [Heading (A, U, G = 1; C = 0.1) refers to NTP concentrations in the transcription reaction, and this will be described in detail in the next subsection]. A nick in the non-template strand by itself (lane 1) does not produce significant blockage under these conditions, albeit some irregular and minor (barely above the background) signals that are not present in the non-nicked control (lane 2) can be seen downstream from the nick. However, the situation is dramatically different in the case of the G16 insert. In the absence of a nick (lane 4), the G16 insert produces only a weak (0.6% of run-off) repeat exiting blockage signal near the position +270. In contrast, a cluster of strong blockage signals appears in the presence of a nick (lane 3), beginning from the strongest signals in the downstream half of the G16 insert, including the repeat-exiting signal, and extending downstream as an irregular ladder of decreasing blockage signals, detectable almost all the way up to the run-off position. The sum of the two strongest signals comprises 40% of the run-off signal. Thus, the nick increases blockage by nearly two orders of magnitude. Interestingly, in case of the C16 insert (lanes 5 and 6), similarly positioned nick-induced blockage signals appear as well. The overall blockage, however, is significantly weaker than that observed for the G16 insert, and the blockage signals are shifted downstream compared with those for the G16 insert. Furthermore, the signals within and in close vicinity of the C16 insert are not the most intense ones, in contrast to the result for the G16 insert.

### Higher NTP concentrations strongly increase blockage signals downstream from the nick in the non-template strand

The experiments described earlier in the text (Figure 2 and lanes 1–6 in Figure 3) were performed under our standard transcription conditions (see Supplementary Materials and Methods), with trace amounts of radioactively labeled CTP, and ‘cold’ adenosine triphosphate, UTP, GTP and CTP; the concentration of cold CTP was one-tenth that for the other cold NTPs, to increase the incorporation of radioactive CTP. These conditions are referred to as (A, U, G = 1; C = 0.1).

Surprisingly, when the concentration of cold CTP was increased up to that of the other NTPs (A, U, G, C = 1) (Figure 3, lanes 7–12), the overall percentage of blockage for the nicked substrates strongly increased; in addition, the distribution of intensities of blockage signals was

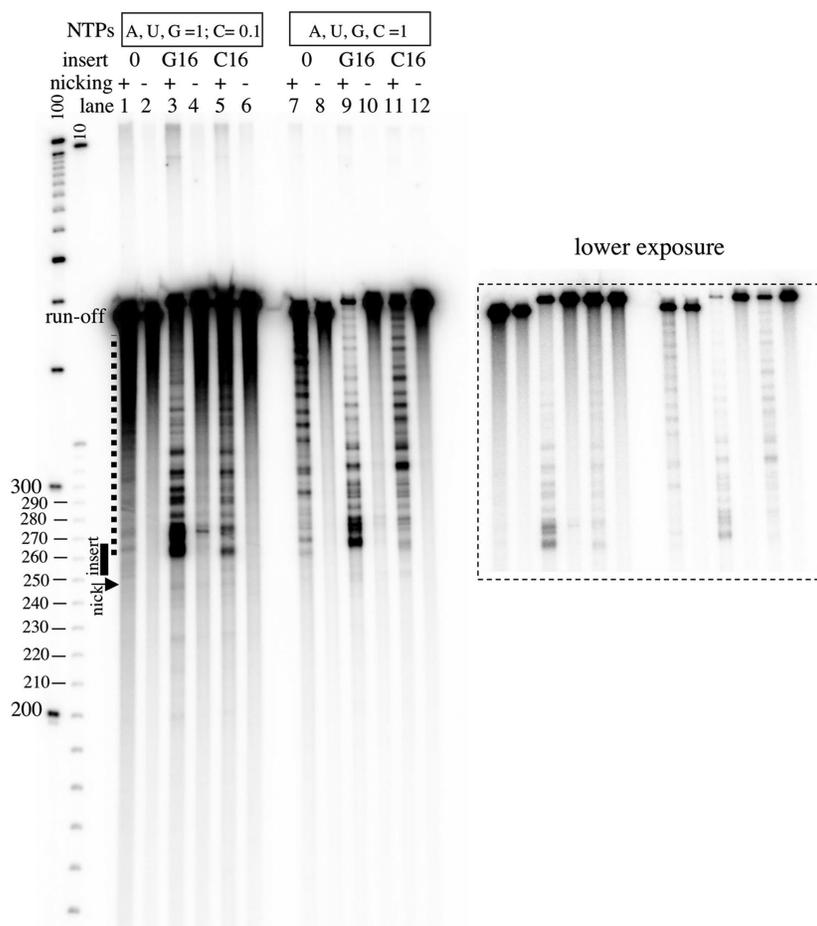


**Figure 2.** Transcription blockage for various DNA insert sequences. The sequences of insert (non-template strand) are shown in the top left corner. All inserts are 32-nt long and start 252-nt downstream from the promoter. Position of the insert on gel (shown thick black lane) is estimated according to denatured DNA ladders with steps of 10 nucleotides and 100 nucleotides, designated by vertical 10 and 100, respectively. The white block arrow shows the repeat-exiting blockage signal, and the white oval shows diffuse blockage signal; for the supercoiled DNA (right panel), the 'total' blockage area is shown by white rectangle. The run-off in the case of linear DNA (left panel) corresponds to the transcript of defined length, whereas in the case of supercoiled DNA (right panel), the apparent run-off is likely to be an unresolved mixture of long products of spontaneous transcription termination at multiple sites within the plasmid. The relative intensities of the blockage signals for various sequences are shown on graphs below the respective panels. In these graphs, the intensities of the respective blockage signals were first normalized to the run-off intensities, and then the ratios of these normalized intensities to the normalized intensity of the G32 insert were plotted versus the G-content of the insert. Each data point is the average of two experimental results, and error bars show deviations from the average. Letters A, T and C designate the non-G base.

shifted in the downstream direction. Under these conditions, for the G16 insert (lane 9), the vast majority of transcripts were truncated, and the sum of the two strongest blockage signals exceeded the run-off product by 2-fold [five times more than for the same substrate under (A, U, G = 1; C = 0.1) conditions].

Moreover, an irregular ladder of nick-induced blockage signals was seen even without any special sequence insert under these conditions (Figure 3, lane 7 versus lane 8). This irregular ladder becomes pronounced some distance downstream from the nick (although the exact starting

point is difficult to define), and it continues up to the run-off position. Interestingly, blockage signals at the same positions, although significantly weaker and without any 'starting area', were also detected in negatively supercoiled DNA (Supplementary Figure S2). This ladder appears only when a nick is positioned in the non-template strand. A nick in the template strand results in the appearance of one sharp blockage signal located approximately at the position of the nick, and a much weaker signal ~10-nt upstream from the nick (Supplementary Figure S3, lane 3). Overall blockage in



**Figure 3.** Effect of a nick in the non-template strand on transcription blockage for various sequences under various conditions. Linearized plasmids pN-aga-0, pN-aga-G16, pN-aga-C16 (see Materials and Methods) were used in these experiments. G16 and C16 inserts (from 252 to 267 nt from the promoter) and nick position (after 248 nt from the promoter) are shown by the thick black line and black arrow, respectively. The vertical dotted line at the left shows the approximate area where blockage signals occur. Headings (A, U, G = 1, C = 0.1) and (A, U, G, C = 1) designate ribonucleotide composition during transcription (see Supplementary Materials and Methods). The incorporation of radioactive CTP per transcript is 10-fold larger at A, U, G = 1, C = 0.1 conditions than at A, U, G, C = 1 conditions, which makes radioactive signals in the first case stronger, despite the fact that lowering CTP concentration decreases intensity of transcriptions. Proper normalization and control experiments with radioactive adenosine triphosphate showed that, in fact, at A, U, G = 1, C = 0.1 conditions, the intensity of transcription is  $\sim 3$ -fold lower than at A, U, G, C = 1 conditions (Supplementary Figure S1).

this case was  $\sim 25\%$ , meaning that RNA polymerase passed the nick with roughly 75% effectiveness and transcribed up to the run-off position [a result similar to that reported previously for small gaps (34) and nicks (35) in the template strand]. An absence of the 'long-range' downstream effect of a nick in the template strand suggests that it does not strongly facilitate RNA sequestration in the form of an R-loop, in contrast to a nick in the non-template strand. This implies that R-loop formation in some cases could render a nick in the non-template strand more harmful for gene expression than a nick in the template strand (36).

Interestingly, the blockage triggered by the C16 insert in the nicked substrates, while less than that by the G16 insert, is still greater than that for the random sequence (Figure 3, lanes 11, 9 and 7, respectively). This result might reflect the fact that the rC/dG duplex, while less stable than the rG/dC duplex, is still more stable than the average mixed RNA/DNA duplex.

Nick-induced transcription blockage was strongly reduced when we reduced the concentrations of NTPs by 3-fold without changing the mixture composition (designated A, U, G, C = 1/3, Supplementary Figure S4, lanes 7–12, higher exposure) such that blockage was only evident for the G16 insert (lane 9).

In the absence of a nick, the repeat-exiting signal was less pronounced when we increased the NTP concentrations (Figure 3, lane 4 versus lane 10). Note, however, that as the blockage signal in this case is weak, it is difficult to judge whether this is owing to a general reduction of transcription blockage or broadening in the distribution of blocking signals.

Nick-enhanced transcription blockage was also observed for a number of other naturally occurring sequences containing G-stretches, including the human telomeric sequence and the sequence from the c-Myc promoter (Supplementary Figures S5 and S6). Remarkably, a stretch of just eight guanines is sufficient

to cause 40% transcription blockage in the presence of a nick (Supplementary Figure S7).

#### **A non-template strand nick downstream from the insert increases blockage, but not as strongly as a nick upstream from it**

To test the effect of nick location, we constructed plasmids pN-tc-G16-cc-N and pN-tc-C16-cc-N, in which nicks in the non-template strand appeared either 2-nt upstream or 2-nt downstream from the insert (see Materials and Methods). Supplementary Figure S8 shows the results of experiments with these plasmids under various conditions. In Supplementary Figure S8A and B, the substrates are circular, and the non-nicked substrates are negatively supercoiled, whereas in Supplementary Figure S8C, the substrates are linear.

For the nick upstream from the insert (lanes 1 and 4 in A and B), the pattern of blockage is similar to that observed for the same inserts in Supplementary Figure S4. For a nick localized downstream from the insert (lanes 2 and 5), the blockages began further downstream and were weaker than these for the upstream nick under all conditions. For the G16 insert, the downstream nick-mediated blockage is stronger than that observed for negatively supercoiled DNA under (A, U, G, C = 1) conditions (Supplementary Figure S8A, lanes 3 and 6), but it becomes comparable with negatively supercoiled DNA under (A, U, G, C = 1/3) conditions (Supplementary Figure S8B and C, respectively, lane 2 versus lane 3). Switching from A, U, G, C = 1 to A, U, G, C = 1/3 conditions practically eliminated nick-induced blockages for the C16 insert such that they were pronounced only for the G16 insert. This sequence-specific, nick-enhanced blockage required the nick to be positioned close to the insert. Increasing the distance between the nick and the insert from 2 nt to 9 nt, dramatically decreased the blockage (Supplementary Figure S9).

#### **Blockages produced by the G-rich insert with nick are at least partially irreversible and increase on an increase in RNAP concentration**

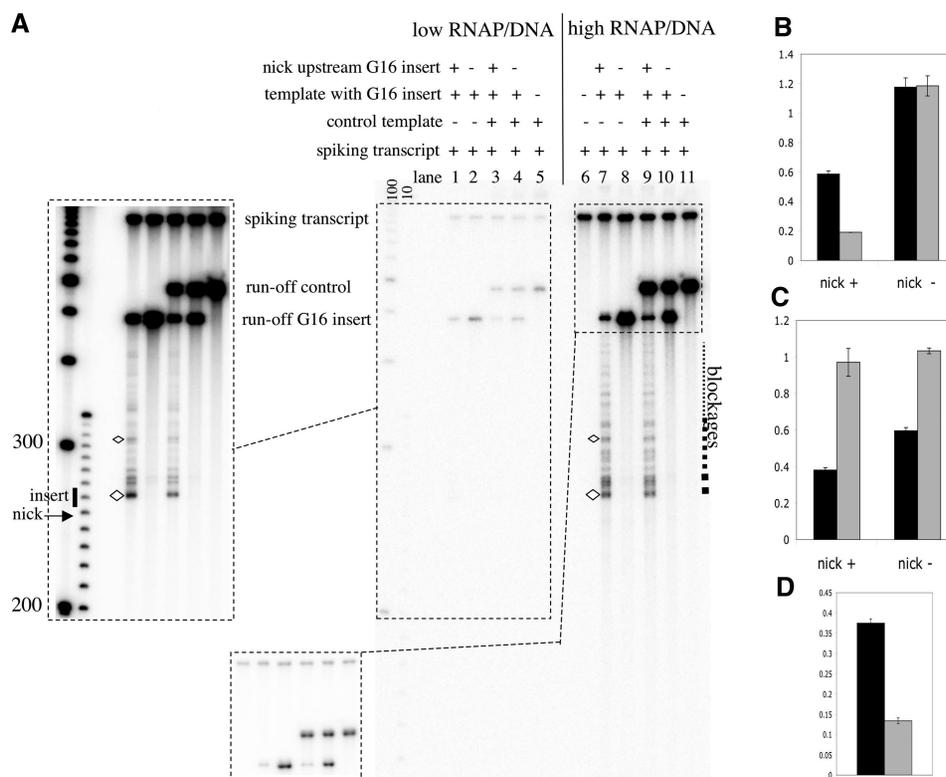
To obtain further insight into the mechanism of transcription blockage in our system, we applied the strategy developed by Grabczyk and Usdin (16), which is based on the competition between different transcription DNA templates present in the same reaction mixture during a multiple-round transcription experiment (see Supplementary Discussion for detailed rationale). This strategy allows one to determine the nature of transcription inhibition by a specific sequence: if its inhibitory effect is due to decrease in elongation rate (such as caused by reversible pausing), the transcription rates for the plasmid containing an inhibitory insert and for the control plasmid should be the same as long as RNAP/DNA ratio is small (i.e. when the templates compete for RNAP). In contrast, if inhibition is due to the inhibition of transcription initiation or to irreversible elongation blockage, the equalizing of transcription rates should not occur. Figure 4A shows the result of the experiment, and Figure 4B–D shows respective quantitations.

Figure 4B shows that for the G16 insert with a nick, the equalizing of transcription rates with the control template does not occur. Because both nick and insert are localized far (~0.25 kb) from the promoter, they should not affect the initiation step. Thus, we concluded that the blockage is at least partially irreversible. Also, the transcription inhibitory effect 'in trans' for the G16 insert with nick (which produces strong blockage) was not much stronger than that for the G16 insert without a nick (which produces barely detectable blockage) (Figure 4C); thus, the blockage is not accompanied by strong RNAP sequestration. We also found that the blockage increases with an increase in RNAP concentration (Figure 4D), suggesting that collisions between RNAP molecules might additionally contribute to the blockage (see Discussion).

## **DISCUSSION**

We have systematically studied the effects of nucleotide substitutions and nicks on T7 RNA polymerase blockage by G-rich inserts. For DNA inserts of the same length, the maximal blockage was observed for pure (G)*n*(C)*n* stretches when the G-rich strand was the non-template (sense) strand for transcription. Substitutions of Gs by any other nucleotides decreased the blockage, but G-to-A substitutions (preserving the homopurine-homopyrimidine nature of the sequence) were much less disturbing than were G-to-C or G-to-T substitutions. Thus, both G-richness and the homopurine-homopyrimidine nature of the sequence are required for effective blockage. It was previously shown that G-rich, homopurine RNA forms much more stable hybrids with its DNA complement than does C-rich homopyrimidine RNA (19–21), suggesting that the formation of extra-stable RNA/DNA hybrids, presumably in the form of R-loops, is responsible for blockage. According to thermodynamic analysis based on the HyTher program (Supplementary Figure S10), our data correlate better with the differences in stabilities between RNA/DNA and DNA/DNA hybrids of the same composition than with the absolute stabilities of RNA/DNA hybrids. This is consistent with R-loop formation, which implies a competition between the nascent RNA and the non-template DNA strand for binding to the template DNA strand.

The requirement for the sequence to be a G-rich homopurine-homopyrimidine also correlates with the stability of a PuPuPy intramolecular triplex (H-r DNA), which would be additionally stabilized by the RNA/DNA hybrid (15,16). The H-DNA model predicts stronger blockage by symmetric versus asymmetric G-to-A substitutions [reviewed in (37,38)]; this was not the case for the G-to-A substitutions in the G32 sequence (28). For other sequences, in which homopurine stretches were interrupted by either pyrimidine stretches (28) or by mixed sequence stretches (32), the effects of sequence variations suggest a contribution from triplex formation to the blockage. An interplay between the RNA-stabilized H-DNA and the R-loop would likely depend on the sequence and ambient conditions, and, in



**Figure 4.** Transcription in a mixture of different DNA templates at low and high RNAP/DNA ratios. As a test plasmid, pN-aga-G16 (further referred as G16) was used, which contains the G16 insert with the site for the nicking enzyme localized in the non-template strand 3-nt upstream from the insert. This plasmid produces strong blockage when it contains nick, but <1% blockage without the nick; as a control plasmid, hTel-C was used, which does not produce detectable blockage. The sequence between the promoter and insert (0.25 kb) was identical for control plasmid and G16 plasmid. Both test and control plasmids were linearized by HindIII. To eliminate errors due to product losses during their purification and gel loading, a 'spiking transcript' was made in separate transcription reaction from the template pN-aga-hTel-C linearized by DraIII (which also does not produce any detectable blockages) and added to all reactions after transcription was stopped but before purification of transcription products (see Supplemental Materials and Methods for details). Transcription reactions were performed 12  $\mu$ l at A, U, G, C = 1 conditions (see Supplemental Materials and Methods). For low RNAP/DNA conditions (lanes 1–5), 0.24 units of T7 RNAP and 200 ng of each DNA template were used per one reaction, whereas in high RNAP/DNA (lanes 7–11) conditions, 20 units of T7 and 10 ng of each DNA template were used; thus, T7 RNAP/DNA ratio varied 1670 times between these conditions. The transcription reaction contains either G16-plasmid alone (lanes 1, 2, 7, 8) or control plasmid (lanes 5 and 11), or their equimolar mixture (lanes 3, 4, 9, 10). Lanes 1, 3, 7 and 9 as well as 2, 4, 8 and 10 correspond to G16-plasmid with or without nick upstream of the G16 insert, respectively. To provide convenient intensity for quantitation, the transcription reaction for spiking transcript was also performed at both low RNAP/DNA and high RNAP/DNA conditions, except that the amounts of template were 120 ng and 6 ng, respectively. The spiking reaction was stopped by addition of 3  $\mu$ l of 100 mM ethylenediaminetetraacetic acid, and then 2  $\mu$ l of respective spiking transcripts were added to all transcription mixtures after the stop buffer. Also, 2  $\mu$ l of spiking transcription reaction alone at high RNAP/DNA condition was processed as the rest of the samples and loaded on gel (lane 6). For convenience of visual analysis, a higher exposure (for lanes 1–5) and a lower exposure (for lanes 6–11) of the areas of the gel within dashed-bordered boxes are shown. For the nicked G16 plasmid, well-pronounced blockage signals could be seen. At high RNAP/DNA ratio, the intensities of blockage signals relative to run-off increase; additionally, the distribution of intensities of blockages noticeably shifts in downstream (i.e. towards run-off) direction. For example, the ratio of intensity of the strongest blockage signal (larger white diamond) to the intensity of run-off was about three times larger, and its ratio to the intensity of one of the weaker downstream blockage signals (smaller white diamond) was about two times smaller at high RNAP/DNA conditions than under low RNAP/DNA conditions. The charts show ratios of radioactive signal intensities for various transcripts, further referred to as 'ratios of transcripts'. All intensities are normalized to the number of cytosines within the transcript; thus, they represent the molar ratios of transcripts. They were also normalized to the intensities of the spiking transcript signals in the same lanes. The data correspond to the average of two experiments, and deviations from the average are shown by the error bars. Results for low RNAP/DNA and high RNAP/DNA conditions are shown in black, and for high RNAP/DNA conditions in gray. (A) Radio-autograph of the gel; (B) Ratio of transcripts obtained from G16 plasmid (either with or without nick) in mixture with the control plasmid to the transcripts obtained from the control plasmid in the same mixture. For G16 plasmid without nick, this ratio is close to 1 at both low and high RNA/DNA conditions, which would be expected for two plasmids with similar initiation and elongation rates. For the G16-plasmid with nick, at high RNAP/DNA conditions this ratio is  $\sim$ 0.2, and at lower RNAP/DNA conditions, it increases up to  $\sim$ 0.6, but does not reach 1, suggesting that the blockage is at least partially irreversible (See Supplementary Discussion); (C) Ratios of transcripts obtained from the control plasmid in mixture with G16-plasmid to the transcripts obtained from the reaction containing the control plasmid alone. The G16-plasmid used in this experiment is either with a nick (nick +) or without a nick (nick -). At high RNAP/DNA conditions, these ratios are close to 1, consistent with an excess of free RNAP in solution and, consequently, no interference between different templates. At low RNAP/DNA, for the mixture with non-nicked G16 plasmid, this ratio approaches 0.5, which suggests that most RNAP in solution are in the DNA-bound state and are evenly distributed between templates. In the case of nicked G16 plasmid, this ratio is  $\sim$ 1.6 times smaller, which might indicate some additional RNAP sequestration by the nicked G16 plasmid; (D) Ratio of transcripts obtained from G16-plasmid with nick to transcripts obtained from G16 plasmid without nick. Reactions were performed with each of the templates, separately. This ratio changes  $\sim$ 3-fold (from 0.13 to 0.38), i.e. the difference between these two templates decreases, on switching from high to low RNAP/DNA conditions.

fact, the former could be a precursor of the latter (39). We therefore cannot exclude the possibility that a short triplex formed transiently in a part of the insert (therefore, insensitive to its overall symmetry) could facilitate R-loop formation by trapping the displaced non-template strand.

In contrast, quadruplex formation does not constrain a sequence to be homopurine homopyrimidine. In fact, quadruplexes with  $X = T$  or  $X = C$  are more stable than are homopurine quadruplexes with  $X = A$  for the (GGGX)-motifs (40,41). In our studies, transcription blockage was much stronger when  $X = A$  in that sequence motif (Figure 5). These data reaffirm our previous conclusion that there is no significant contribution to transcription blockage from quadruplex formation in our case (28). Overall, our results support the conclusion that transcription blockage by G-rich, homopurine-homopyrimidine sequences is primarily owing to the R-loop formation.

Roy *et al.* (30) defined various factors, which facilitate R-loop formation, and offered a mechanistic explanation for their actions. Among these factors are negative supercoiling and a nick in the non-template strand. We reasoned that if transcription blockage was due to R-loop formation, these same factors should strongly increase it. We have demonstrated strong increases in transcription blockage by negative supercoiling (28). In the present study, a dramatic increase in transcription blockage was seen when a nick was positioned in the non-template strand close to the G-rich insert. Moreover, an irregular ladder of minor blockage signals extending downstream from the nick appeared even in the absence of the G-rich insert. Similar, albeit weaker, signals were observed in negatively supercoiled DNA (Supplementary Figure S2). That is again consistent with R-loop formation, which is known to be facilitated by negative supercoiling [(5,30,42–44) and references therein].

Although there is a strong correlation between the requirements for R-loop formation and for transcription blockage, the question remains, why should R-loops facilitate transcription blockage?

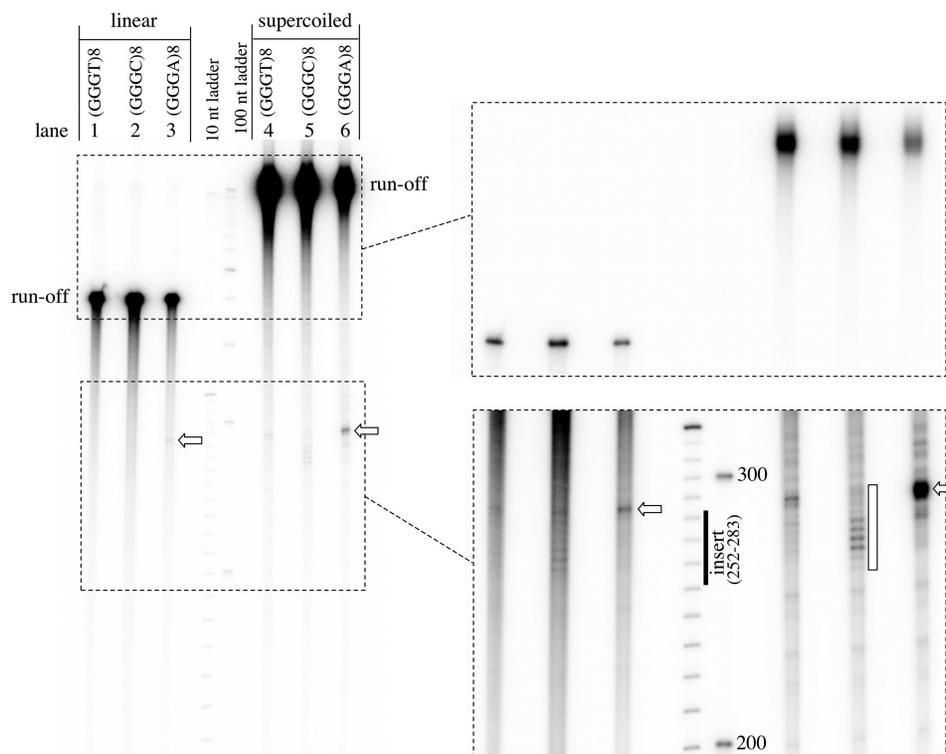
At first glance, the simplest explanation is that during multiple rounds of transcription, RNA polymerases collide with the R-loops formed by preceding RNA polymerases, thereby interfering with the transcription outcome. In accordance with this mechanism, a 2-fold decrease in the total yield of transcription through a long (300 nt), artificially pre-formed R-loop was observed (45). We have shown previously, however, that transcription blockage also occurs during single-round transcription in our system (28). Furthermore, if the blockage signals were produced on collisions with the previously formed R-loops, one would expect them to occur upstream of the causative sequences while we observe blockage signals located predominantly downstream from them. Thus, collision with pre-formed R-loops, although theoretically possible in our system, cannot explain our results.

We have previously argued that the continuation of transcription beyond the formation of a stable R-loop should trigger topological constraints owing to sterical clashes between DNA and nascent RNA, which could

destabilize the elongation complex and result in transcription blockage (29). A nick in the non-template strand should remove topological constraints. Therefore, if these constraints were essential for RNAP blockage, a nick should alleviate the blockage, despite the fact that it facilitates R-loop formation. In contrast, we show that a nick dramatically increases blockage by G-rich sequences and that it triggers transcription blockage even for random DNA sequences. These observations suggest that there must be a fundamental mechanism explaining transcription blockage by R-loops that does not require specific sequences or topological constraints, even though it can be modulated by these factors.

We believe that this fundamental mechanism is grounded in the existence of different configurations of the transcription elongation complex during normal transcription in comparison with the transcription complex in the presence of an R-loop. During normal transcription elongation, the nascent RNA is separated from its DNA template (46,47) and is 'guided' from the transcription complex through a positively charged groove (exit channel) on the inner surface of the RNA polymerase (48). Interaction between the nascent RNA and the surface of the exit channel strongly contributes to the stabilization of the transcription complex [(42,49) and references therein]. R-loop formation could alter this normal elongation mode: according to one model, the R-loop is formed immediately upstream of the normal RNA/DNA hybrid within the transcription complex, with only a brief (if any) separation of the nascent RNA from the DNA template (17,42,44). In this scenario, the R-loop disrupts proper interactions between the nascent RNA and the exit channel, thus destabilizing the transcription complex, similar to the manner in which a hairpin in the nascent RNA destabilizes the transcription complex at termination signals (42). In the alternative ('thread-back') model for R-loop formation (23,30), the nascent RNA is extruded from the transcription complex normally, but this is followed by re-hybridization with the template strand upstream of the elongation complex. In this model, the R-loop formation does not disrupt interactions between the RNAP and the nascent transcript at first, but this disruption could occur later, on growing of RNA/DNA duplex upstream the transcription complex (Figure 6). In any case, the R-loop would likely create some additional strains and distortions in the elongation complex, increasing the probability of transcription blockage. In support of this interpretation, R-loop formation has been shown to exacerbate weak transcription termination in stretches of oligo(dT)/oligo(dA) (50).

The R-loop-mediated exacerbation of weak transcription pausing/termination signals could explain the irregular ladder of blockage signals extending downstream from the causative sequence into seemingly random sequences (Supplementary Figure S11). Even for random sequences, there is usually some probability of spontaneous transcription pausing/termination at each position, which varies depending on the physical properties of the DNA as well as DNA/RNA hybrids, and other factors. Indeed, ubiquitous minor pauses during transcription elongation have been observed in single-molecule experiments



**Figure 5.** Comparison of blockages produced by GGGX repeats. The left panel shows the entire gel, whereas top and bottom right panels show the run-off area with reduced exposure times, and the blockage area with higher exposure, respectively. For linear templates (lanes 1–3), the repeat-exiting blockage signal (white block arrow) is clearly pronounced over the background only for X = A. For supercoiled templates (lanes 4–6), the repeat-exiting blockage signal is pronounced also for X = T, but its value (normalized on ‘run-off’) is 13 times smaller, than for X = A. For X = C, the repeat-exiting blockage signal is not pronounced; however, several blockage signals are evident within the insert. The explanation for this pattern is presently unclear; however, it could be compared with the blockage signals produced by other sequences by estimating the total blockage signal over the area shown by the white rectangle, which includes all blockages. To perform the comparison, a signal was calculated over the same area also for GGG A and GGG T repeats. For GGG T and GGG C repeats, this blockage was seven and three times less than that for GGG A repeats, respectively.

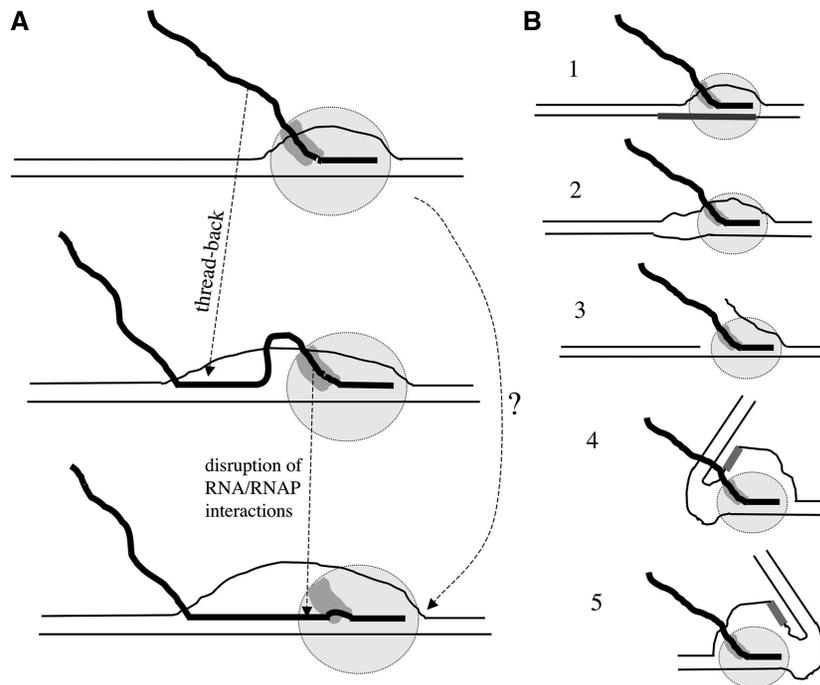
(51,52). Normally, however, these pausing/termination signals are too weak to be detected in our assay; only the presence of additional factors that interfere with transcription elongation (R-loops in our case) reveal those ‘hidden’ signals.

Thus, the strong inhibitory effect of G-rich, homopurine-homopyrimidine sequences on transcription could simply be caused by the high likelihood of R-loop formation owing to the superior stability of RNA/DNA hybrids formed by these sequences. An additional problem for RNAP could be to ‘peel off’ the nascent RNA from the extra-stable puRNA/pyDNA hybrid in the transcription elongation complex during RNA extrusion (28).

When a nick is placed in close vicinity to the G-rich, homopurine-homopyrimidine sequence (further referred as an ‘R-loop prone sequence’), the transcription blockage becomes much stronger, indicating that the nick and an R-loop prone sequence act cooperatively in increasing the probability of R-loop formation. The effect of an upstream nick is significantly stronger than that of a downstream nick (Supplementary Figure S8). The nick in the non-template strand facilitates R-loop formation by decreasing RNA-displacing propensity of the

non-template strand, thus facilitating a thread back of the transcript (30). This means that the effect of the nick presumably would manifest itself when RNAP has passed the nick, and the nick is localized closely behind the transcription complex. For the upstream nick, it means that the sequence within and, probably, a few nucleotides behind the transcription complex is the one which forms the extra-stable RNA/DNA hybrid. That would stabilize a short stretch of RNA/DNA duplex during reversible RNA invasion, thus increasing probability of R-loop formation. In contrast, for the downstream nick, the sequence within and a few nucleotides behind the transcription complex is random, and it would therefore require for RNA to invade a few nucleotides further upstream to reach the sequence, which forms extra-stable RNA/DNA hybrid. That could explain the difference in effects for the two different nicks locations.

The detailed mechanisms of R-loop and/or triplex-mediated transcription blockage could also be different in various systems. For example, Grabczyk and Usdin (16) showed that a reduction in transcription through long GAA repeats is primarily owing to the reversible sequestration of RNAP by these repeats. In our system, sequestration does not seem to be the primary source of



**Figure 6.** Possible mechanism for R-loop interference with transcription. (A) The basic mechanism. Normally, the nascent RNA (shown by the thick black line) interacts with certain area (shown by stripped patch) of RNA polymerase (shown by gray oval with dotted line border). R-loop formation (either via thread-back mechanism or some other mechanism shown by dashed line with question mark) disrupts (possibly, partially) this interaction, thus destabilizing the elongation complex and making it more prone to stalling or/and dissociation. (B) Factors that exacerbate the blockage by facilitating R-loop formation: (1) Sequence that forms extra-stable RNA/DNA hybrid (shown by thick patterned line); (2) negative supercoiling, which increases fluctuative opening of DNA; (3) nick in the non-template strand, which decreases propensity of the non-template strand to displace RNA; involvement of the part of the non-template DNA strand (shown by the thick patterned line) in triplex formation with the DNA duplex upstream (4) or downstream (5) of the transcription complex, which would sequester non-template DNA strand thus decrease its propensity to displace RNA. In addition to facilitating R-loop formation, some of these factors could additionally exacerbate blockage by other mechanisms. For example, extra-stable RNA/DNA hybrid inside the transcription complex could interfere with the nascent RNA separation, and triplexes could create obstacles for RNAP movement.

blockage, and at least some blocks appear to be irreversible (Figure 4).

For the substrates containing nicks, we also observed an increase in RNAP blockage when the RNAP/DNA ratio was increased (Figure 4). That indicates that collisions between RNA polymerases might also contribute to the transcription blockage by forcing paused enzymes out from their templates (53). In our case, such collisions might occur when a transcribing RNAP overtakes and displaces the RNAP slowed or paused on the R-loop formation. An increase in the RNAP/DNA ratio not only facilitates the blockage but it also shifts the distribution of blockage sites in the downstream direction. The latter observation argues against an alternative explanation that transcription blockage occurs when a new RNAP becomes trapped by the previously stalled RNAPs, as one would expect blocking sites to shift upstream in this case.

In light of these observations, the increased blockage on increasing NTP concentration is probably owing to surge in transcription initiation rate, which would have a similar effect as an increase in RNAP concentration. In support of this view, our experimental data suggest that an increase in NTP concentration from A, U, G, C = 1/3 to A, U, G, C = 1 strongly increases transcription initiation rates (Supplementary Figure S12).

Interestingly, the increased transcription blockage on an increase in the T7-RNAP/DNA ratio contrasts with results reported for bacterial RNA polymerase, in which collisions facilitated elongation [reviewed in (54)]. This difference could be explained by the frequent back-tracking characteristic of the bacterial RNAP [reviewed in (54)], which does not occur with the T7 RNAP.

The interference of extra-stable RNA/DNA hybrid-forming sequences with transcription elongation could be a general property for various RNA polymerases, and this is likely responsible for transcription pausing at G-rich sequences *in vivo* (55). Exacerbation of this pausing by nicks (which can result from DNA damage, intermediates in DNA repair, topoisomerase actions, etc.) might cause significant blockage in natural sequences that contain short homopurine stretches, such as the telomeric sequences and c-Myc promoter sequences, both of which have been shown to be transcribed *in vivo* in the direction in which the purine-rich strand is the non-template (sense) [reviewed in (56,57)].

Transcription blockage and R-loop formation have been implicated in various types of genomic instabilities [e.g. see (1,26,58–62)]. It is tempting to speculate that these instabilities could be exacerbated by nicks in the non-template strand.

It is also interesting to consider how the nick-enhanced blockage might impact transcription-coupled repair (TCR), the specialized subpathway of nucleotide excision repair that targets blocked RNA polymerases to efficiently remove the responsible lesions from the transcribed DNA strands (3,63,64). It was hypothesized that if transcription were blocked in undamaged DNA, TCR might occasionally activate futile cycles of DNA repair, eventually leading to mutagenesis and destabilizing the genome (65). Alternatively, this gratuitous form of TCR might also attract ligases and other repair factors, which would seal the nicks and promote genome stability.

In addition, transcription blockage by R-loop formation has been implicated in regulation of RNA transactions, including RNA decapping (66) and sequestration (67).

Thus, transcription blockage can have multiple biological effects, of which some are deleterious, but others may be beneficial.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–12, Supplementary Materials and Methods and Supplementary Discussion.

## ACKNOWLEDGEMENTS

The authors thank Robert Landick, Laurent Lacroix, Steven Block and members of his lab, as well as the members of the authors' research group for helpful discussions.

## FUNDING

National Institutes of Health (NIH) from the National Cancer Institute [CA077712 to P.C.H.]; National Institute of General Medicine [GM60987 to S.M.M.]; Undergraduate Research Grants at Stanford (to A.J.N., S.S. and J.H.S.S.). Funding for open access charge: NIH from the National Cancer Institute to the laboratory of P.C.H. at Stanford University [2RO1 CA077712].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Aguilera, A. (2002) The connection between transcription and genomic instability. *EMBO J.*, **21**, 195–201.
2. Bacolla, A. and Wells, R.D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.*, **48**, 273–285.
3. Hanawalt, P.C. and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell. Biol.*, **9**, 958–970.
4. Zhao, J., Bacolla, A., Wang, G. and Vasquez, K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life. Sci.*, **67**, 43–62.
5. Daniels, G.A. and Lieber, M.R. (1995) RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.*, **23**, 5006–5011.
6. Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
7. Eddy, J. and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.*, **36**, 1321–1333.
8. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
9. Eddy, J. and Maizels, N. (2009) Selection for the G4 DNA motif at the 5' end of human genes. *Mol. Carcinog.*, **48**, 319–325.
10. Wongsurawat, T., Jenjaroenpun, P., Kwok, C.K. and Kuznetsov, V. (2012) Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.*, **40**, e16.
11. Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E. and Lieber, M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.*, **4**, 442–451.
12. Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
13. Reaban, M.E., Lebowitz, J. and Griffin, J.A. (1994) Transcription induces the formation of a stable RNA-DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.*, **269**, 21850–21857.
14. Reaban, M.E. and Griffin, J.A. (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature*, **348**, 342–344.
15. Grabczyk, E. and Fishman, M.C. (1995) A long purine-pyrimidine homopolymer acts as a transcriptional diode. *J. Biol. Chem.*, **270**, 1791–1797.
16. Grabczyk, E. and Usdin, K. (2000) The GAA\*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res.*, **28**, 2815–2822.
17. Karamychev, V.N., Panyutin, I.G., Neumann, R.D. and Zhurkin, V.B. (2000) DNA and RNA Folds in Transcription Complex as Evidenced by Iodine-125 Radioprobe. *J. Biomol. Struct. Dyn.*, **17**, 155–167.
18. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. and Akoulitchev, A. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, **445**, 666–670.
19. Roberts, R.W. and Crothers, D.M. (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science*, **258**, 1463–1466.
20. Lesnik, E.A. and Freier, S.M. (1995) Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry*, **34**, 10807–10815.
21. Gyi, J.I., Conn, G.L., Lane, A.N. and Brown, T. (1996) Comparison of the thermodynamic stabilities and solution conformations of DNA:RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes. *Biochemistry*, **35**, 12538–12548.
22. Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell. Biol.*, **29**, 3124–3133.
23. Roy, D., Yu, K. and Lieber, M.R. (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell. Biol.*, **28**, 50–60.
24. Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R. and Gearhart, P.J. (2009) Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J. Exp. Med.*, **206**, 1237–1244.
25. Mirkin, E.V. and Mirkin, S.M. (2007) Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.*, **71**, 13–35.
26. Potenski, C.J. and Klein, H.L. (2011) R we there yet? R-loop hazards to finishing the journey. *Mol. Cell.*, **44**, 848–850.
27. Krasilnikova, M.M., Samadashwily, G.M., Krasilnikov, A.S. and Mirkin, S.M. (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO J.*, **17**, 5095–5102.
28. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M. and Hanawalt, P.C. (2010) Mechanisms and

- implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl Acad. Sci. USA*, **107**, 12816–12821.
29. Belotserkovskii, B.P. and Hanawalt, P.C. (2011) Anchoring nascent RNA to the DNA template could interfere with transcription. *Biophys. J.*, **100**, 675–684.
  30. Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L. and Lieber, M.R. (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol. Cell. Biol.*, **30**, 146–159.
  31. Tornaletti, S., Patrick, S.M., Turchi, J.J. and Hanawalt, P.C. (2003) Behavior of T7 RNA polymerase and mammalian RNA polymerase II at site-specific cisplatin adducts in the template DNA. *J. Biol. Chem.*, **278**, 35791–35797.
  32. Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.*, **282**, 32433–32441.
  33. Ditlevson, J.V., Tornaletti, S., Belotserkovskii, B.P., Teijeiro, V., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2008) Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res.*, **36**, 3163–3170.
  34. Zhou, W. and Doetsch, P.W. (1994) Transcription bypass or blockage at single-strand breaks on the DNA template strand: effect of different 3' and 5' flanking groups on the T7 RNA polymerase elongation complex. *Biochemistry*, **33**, 14926–14934.
  35. Neil, A.J., Belotserkovskii, B.P. and Hanawalt, P.C. (2012) Transcription blockage by bulky end termini at single-strand breaks in the DNA template: differential effects of 5' and 3' adducts. *Biochemistry*, **51**, 8964–8970.
  36. Khobta, A., Lingg, T., Schulz, I., Warken, D., Kitsera, N. and Epe, B. (2010) Mouse CSB protein is important for gene expression in the presence of a single-strand break in the non-transcribed DNA strand. *DNA Repair (Amst.)*, **9**, 985–993.
  37. Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65–95.
  38. Mirkin, S.M. and Frank-Kamenetskii, M.D. (1994) H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 541–576.
  39. Grabczyk, E., Mancuso, M. and Sammarco, M.C. (2007) A persistent RNA-DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res.*, **35**, 5351–5359.
  40. Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.*, **581**, 1657–1660.
  41. sGuedin, A., De Cian, A., Gros, J., Lacroix, L. and Mergny, J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686–696.
  42. Gopal, V., Briebe, L.G., Guajardo, R., McAllister, W.T. and Sousa, R. (1999) Characterization of structural features important for T7 RNAP elongation complex stability reveals competing complex conformations and a role for the non-template strand in RNA displacement. *J. Mol. Biol.*, **290**, 411–431.
  43. Drolet, M. (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.*, **59**, 723–730.
  44. Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H. and Pearson, C.E. (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.*, **39**, 1749–1762.
  45. Tous, C. and Aguilera, A. (2007) Impairment of transcription elongation by R-loops in vitro. *Biochem. Biophys. Res. Commun.*, **360**, 428–432.
  46. Daube, S.S. and von Hippel, P.H. (1994) RNA displacement pathways during transcription from synthetic RNA-DNA bubble duplexes. *Biochemistry*, **33**, 340–347.
  47. Yin, Y.W. and Steitz, T.A. (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
  48. Jiang, M., Ma, N., Vassilyev, D.G. and McAllister, W.T. (2004) RNA displacement and resolution of the transcription bubble during transcription by T7 RNA polymerase. *Mol. Cell*, **15**, 777–788.
  49. Montesana, P.E., Chin-Bow, S.T., Sousa, R. and McAllister, W.T. (2000) Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *J. Mol. Biol.*, **302**, 1049–1062.
  50. Tomizawa, J. and Masukata, H. (1987) Factor-independent termination of transcription in a stretch of deoxyadenosine residues in the template DNA. *Cell*, **51**, 623–630.
  51. Zhou, J., Ha, K.S., La Porta, A., Landick, R. and Block, S.M. (2011) Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. *Mol. Cell*, **44**, 635–646.
  52. Dalal, R.V., Larson, M.H., Neuman, K.C., Gelles, J., Landick, R. and Block, S.M. (2006) Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. *Mol. Cell*, **23**, 231–239.
  53. Zhou, Y. and Martin, C.T. (2006) Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced “bumping”. *J. Biol. Chem.*, **281**, 24441–24448.
  54. Nudler, E. (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell*, **149**, 1438–1445.
  55. Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y. and Maizels, N. (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.*, **39**, 4975–4983.
  56. Azzalin, C.M. and Lingner, J. (2008) Telomeres: the silence is broken. *Cell Cycle*, **7**, 1161–1165.
  57. Marcu, K.B., Bossone, S.A. and Patel, A.J. (1992) myc function and regulation. *Annu. Rev. Biochem.*, **61**, 809–860.
  58. Lin, Y., Hubert, L. Jr and Wilson, J.H. (2009) Transcription destabilizes triplet repeats. *Mol. Carcinog.*, **48**, 350–361.
  59. Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D. and Napierala, M. (2010) R loops stimulate genetic instability of CTG-CAG repeats. *Proc. Natl Acad. Sci. USA*, **107**, 692–697.
  60. Salinas-Rios, V., Belotserkovskii, B.P. and Hanawalt, P.C. (2011) DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. *Nucleic Acids Res.*, **39**, 7444–7454.
  61. Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M. and Aguilera, A. (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.*, **30**, 3106–3119.
  62. Kim, N. and Jinks-Robertson, S. (2011) Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair (Amst.)*, **10**, 953–960.
  63. Mellon, I., Spivak, G. and Hanawalt, P.C. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*, **51**, 241–249.
  64. Mellon, I. and Hanawalt, P.C. (1989) Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature*, **342**, 95–98.
  65. Hanawalt, P.C. (1994) Transcription-coupled repair and human disease. *Science*, **266**, 1957–1958.
  66. Geisler, S., Lojek, L., Khalil, A.M., Baker, K.E. and Collier, J. (2012) Decapping of Long Noncoding RNAs Regulates Inducible Genes. *Mol. Cell*, **45**, 279–291.
  67. Liu, C.R., Chang, C.R., Chern, Y., Wang, T.H., Hsieh, W.C., Shen, W.C., Chang, C.Y., Chu, I.C., Deng, N., Cohen, S.N. et al. (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell*, **148**, 690–701.