Supporting Information

Belotserkovskii et al. 10.1073/pnas.1007580107

SI Text

SI Methods. ***Estimation of the percentages of blockage.*** The intensities of the bands were measured by phosphoimaging, as described in more details in ref. 1. In our calculations we assume that the transcriptional products are homogeneously labeled, and consequently the intensities of their signals are proportional to their lengths. Therefore, for linear DNA templates the molar percentage of the truncated blockage products are estimated from the following equation:

$$p_s = \frac{\frac{I_s L_{r\text{-off}}}{I_{r\text{-off}} L_s}}{\frac{I_s L_{r\text{-off}}}{I_{r\text{-off}} L_s} + 1}, \qquad \text{[S1]}$$

where p_s is the percentage of the blockage, I_s and $I_{r\text{-off}}$ are intensities, and L_s and $L_{r\text{-off}}$ are lengths of specific blockage product and runoff product, respectively. For linear G32 the percentage of blockage estimated by Eq. **S1** was $12.6 \pm 1.1\%$.

For supercoiled templates, the quantitative estimation for the percentage of blockage is complicated by the fact that, due to the absence of specific termination sites in our plasmids, the apparent "runoff" product is heterogeneous. To estimate the percentage of blockage inserts G20T4G8 and C20T4C8 were cloned into a ribozyme-containing vector (1–3). The ribozyme cleaves RNA with about 80% efficiency 0.56 kb downstream of the insert of interest, and the cleavage product can be used for normalization of the blockage signal:

$$p_s = \frac{\frac{\theta I_s L_{\text{ribo}}}{I_{\text{ribo}} L_s}}{\frac{\theta I_s L_{\text{ribo}}}{I_{\text{ribo}} L_s} + 1}, \qquad \text{[S2]}$$

where p_s is the percentage of the blockage, I_s and I_{ribo} are intensities, and L_s and L_{ribo} are lengths of specific blockage product and ribozyme cleavage product, respectively, and θ is the efficiency of ribozyme cleavage.

Multiple-round transcription. Transcription reactions were performed as described in 12 μL of buffer containing 10 ng of DNA substrate, 33 mM TrisHCl (pH 7.9), 5 mM $MgCl_2$, 8.3 mM NaCl, 1.7 mM spermidine, 4.2 mM DTT, 0.17 mM each ATP, GTP, and UTP, 0.017 mM CTP, about 10 μCi (α-^{32}P) CTP, 20 units of T7 RNA polymerase, and 16 units of RNasin (both from Promega Corp.) for 30 min at 37 °C. When GTP in the transcription reaction was substituted by ITP (Sigma-Aldrich) or 7-deazaGTP (TriLink BioTechnologies), 0.67 mM GMP (Sigma-Aldrich) was added to the reaction (4). The reaction was stopped by adding 94 μL of stop buffer containing 1% SDS, 106 mM TrisHCl (pH 7.6), 160 mM NaCl, 13.2 mM EDTA, 25 μg tRNA, and 10 μg Proteinase K, and after 15 min incubation at the room temperature, transcription products were ethanol-precipitated and analyzed on the sequencing gel as described in ref. 1.

Single-round transcription.
For the single-round transcription experiment, for two of our plasmids, G20T4G8 and C20T4C8, a 37-bp-long T-less sequence GGGAGACCACAACGGGACAGACCAGACCGCACAGAGC was placed immediately downstream from the T7 promoter. Except for the last three nucleotides, this sequence is identical to the 34-bp-long sequence used by Zhou and Martin (5). First, to initiate transcription and obtain stalled elongation complex 37 nt downstream from the promoter, 10 ng of template DNA were incubated with T7 RNAP in 10 μL of U-less mixture containing 0.2 mM ATP, and GTP, 0.02 mM CTP, 10 μCi (α-^{32}P) CTP, 20 units of T7 RNA polymerase, and 16 units of RNasin for 5 min at 37 °C. The buffer conditions were 20 mM TrisHCl (pH 7.9), 15 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA, and 0.05% Tween20 (6, 7). To continue elongation, 2 μl of mixture containing 1 mM UTP (which allows further elongation) and 1.5 mg/mL heparin (which prevents new transcription initiation but allows elongation of already existing stalled transcription complexes) was added (producing the final concentration of heparin 250 μg/mL, and the final concentration of UTP 0.17 mM), and the incubation was continued at 37 °C for the designed time. Stopping transcription reaction and purification and analysis of the transcription products were the same as for the multiple-round transcription (see the main text).

Transcription-dependent oligonucleotide hybridization.
5′-Oligonucleotide labeling was performed with ^{32}P-γ-ATP (PerkinElmer) using T4 polynucleotide kinase (Invitrogen), followed by purification with a nucleotide removal kit (Qiagen). Transcription was performed in 12 μL of buffer containing 200 ng of DNA substrate, 33 mM TrisHCl (pH 7.9), 5 mM $MgCl_2$, 8.3 mM NaCl, 1.7 mM spermidine, 4.2 mM DTT, 0.17 mM of each NTP, and 20 units of T7 RNA polymerase for 20 min at 37 °C. Then 1.2 μL of 100 mM EDTA and 1.2 μL of RNaseA/T mixture (Ambion) were added and the mixture was incubated 1 h at 37 °C. Then 1.2 μL of 1.25 M NaCl and 1.5 μL radioactively labeled oligonucleotide (approximate concentration 300 nM) were added to the mixture, and the incubation continued for 1 h at room temperature. For experiments involving RNaseH, transcription and RNaseA and T treatment were performed as described above, except the reactions were upscaled 5-fold. Next, the products were purified using a PCR purification kit (Qiagen) and recovered in 30 μL elution buffer (Qiagen). Next, to 7.5 μL of this solution, 1.2 μL of 10× RNaseH buffer (NEB), 2.3 μL H_2O, and 1 μL of RNaseH (5 units/μL, NEB) were added and the mixture was incubated at 37 °C for 1 h. Next, 1.2 μL of 100 mM EDTA, 1.2 μL H_2O, 1.2 μL of 1.25 M NaCl, and 1.5 μL radioactively labeled oligonucleotide were added and incubation was performed for 1 h at room temperature. The samples were placed in a precooled (−20 °C) rack. For analysis, the samples were loaded on 1% agarose gels and electrophoresis was performed overnight at about 1 V/cm at 4 °C. The gel was stained with EtBr, photographed, placed on a Hybond membrane, dried, and exposed (without separation from the membrane) to a phosphoimager screen.

SI Discussion. ***Possible sterical constraints for further transcription elongation produced by R-loop formation.*** During transcription the nascent RNA is separated from the DNA template upon exiting the transcription complex. However, in some cases, for example when an exceptionally stable rPu/dPy duplex is formed, the RNA-DNA hybrid persists behind the transcription complex forming an R loop. According to suggested models, the R- loop is formed by RNA rehybridization with the template DNA strand in the unwound region of DNA duplex immediately upstream of the transcription bubble (8, 9).

Here we will analyze what happens when RNAP has just formed a stable R loop during transcription of a template sequence that is prone to R-loop formation, and now begins to transcribe an "ordinary" DNA sequence.

One option is that RNAP will exit the "R-loop synthesis mode" and continue transcription in a "normal" way, i.e., with separation

of the RNA-DNA hybrid, whereas the region of the transcript that forms an exceptionally stable RNA-DNA duplex remains base-paired with DNA within the R-loop (Fig. S7*A Top*).

The problem is that in this case, because the position of the R loop remains fixed relative to the DNA, and the RNA polymerase rotates relative to DNA while following a helical path during transcription, the nascent transcript produced after R-loop formation must wrap around the DNA duplex region localized between the R loop and the elongating RNAP (Fig. S7*A Bottom*). Such wrapping, if it really happens, would be extremely entropically unfavorable because it would dramatically decrease the conformational space available for nascent RNA, and it would create a force opposing transcription elongation. The second option is that after R-loop formation within the R-loop-prone sequence, the nascent transcript continues to form an RNA-DNA hybrid behind the RNA polymerase when it continues transcription into a regular downstream sequence, which normally would not support R loop formation (Fig. S7*B*). As we mention in the main text, this mode of transcription elongation could be more prone to spontaneous stalls/termination than normal transcription, because rewinding of the DNA/DNA duplex behind the RNAP facilitates transcription elongation, and R-loop formation prevents this rewinding. Besides, because RNAP remains bound to the nontemplate strand during transcription, sterical clashes between the nontemplate strand and the downstream RNA-DNA duplex might occur during elongation, and this also would be expected to increase the probability of spontaneous transcription termination.

Thus, in both cases stable R-loop formation interferes with further transcription elongation, which could explain "secondary" blockage signals downstream of the G-rich inserts. (The minor blockage bands upstream from the inserts are probably due to secondary collisions with already stalled RNAPs.) Thus, we would expect that whenever a stable R loop is formed, it interferes with further transcription. It could occur, for example, during transcription of immunoglobulin S regions and c-MYC gene, where biological effect of R loops was implicated (10). Stalled RNAP could contribute to the effect of R loop per se, for example, by attracting some protein factors.

1. Belotserkovskii BP, et al. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem* 282:32433–32441.
2. Grabczyk E, 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.
3. Ditlevson JV, et al. (2008) Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res* 36:3163–3170.
4. Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* 180:51–62.
5. Zhou Y, Martin CT (2006) Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision-induced "bumping." *J Biol Chem* 281:24441–24448.
6. Maslak M, Martin CT (1994) Effects of solution conditions on the steady-state kinetics of initiation of transcription by T7 RNA polymerase. *Biochemistry* 33:6918–6924.
7. Mentesana PE, Chin-Bow ST, Sousa R, McAllister WT (2000) Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *J Mol Biol* 302:1049–1062.
8. Roy D, Lieber MR (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.
9. Roy D, Yu K, Lieber MR (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol* 28:50–60.
10. Duquette ML, Pham P, Goodman MF, Maizels N (2005) AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24:5791–5798.

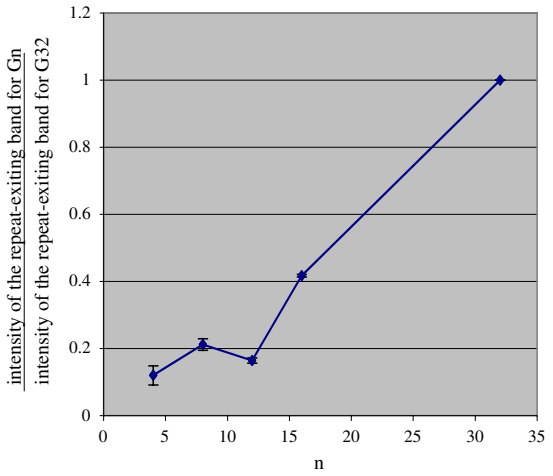

Fig. S1. Relative intensities of exit blockage bands for Gn inserts for linear DNA templates. Intensities of the bands were measured by phosphoimaging, and the average intensities of areas with the same size as the band immediately above and below the bands were subtracted as a background. The intensities of the blockage bands were normalized to corresponding runoff bands, and these normalized intensities are presented as a ratio to the normalized intensity for the G32 insert. The intensities of repeat-exiting blockage bands increase with the length, except for a minor deviation from this trend between $n = 8$ and $n = 12$.

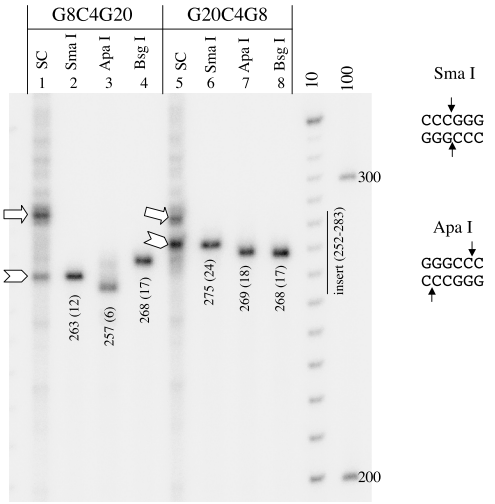

Fig. S2. More accurate estimation for the positions of the interruption blockage bands using transcripts from plasmids digested with restriction enzymes as mobility controls. Numbers below the runoff bands for transcripts from digested plasmids show the distances from the promoter and (in parentheses) the distance from the start of the insert. Sma I cleaves the transcription template DNA strand precisely at the downstream junction of the C4-interruption, and Apa I cleaves the transcription template DNA strand 2 nt before the C4-interruption (see the scheme at *Right*). Bsg I cleaves 17 nt from the start of the insert regardless of the sequence. Interruption bands are shown by white chevrons, and repeat-exiting bands are shown by white arrows. It can be seen that for both G8C4G20 and G20C4G8 the interruption bands are localized very close to the run-off bands from the Sma I digest of the corresponding plasmids.

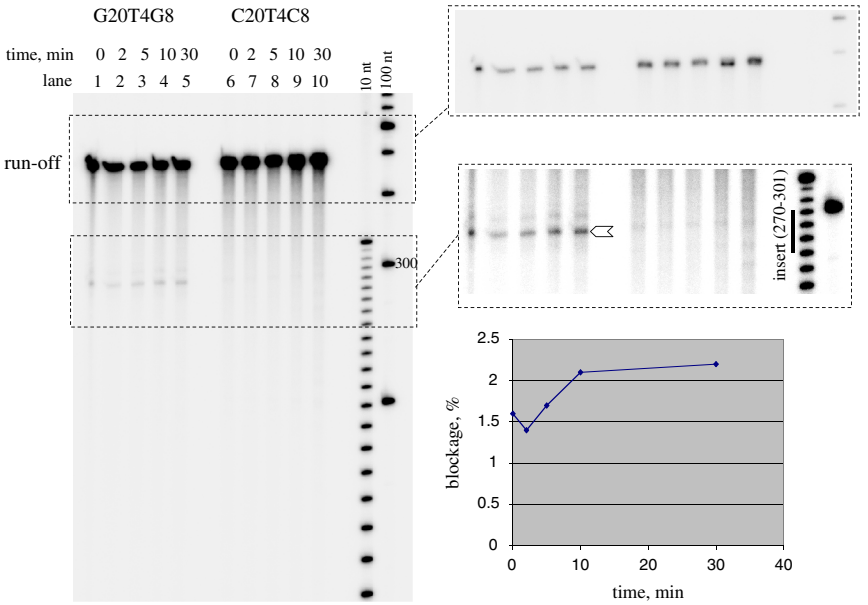

Fig. S3. Single-round transcription experiments for linear DNA templates with inserts G20T4G8 and C20T4C8. In this case the length of the runoff product is 480 nt, and the position of inserts was 270–301 nt from promoter instead of usual 252–283 nt because of extra T-less sequence cloned downstream the promoter (see *SI Materials and Methods*). Areas in the dashed rectangles at the right show lower exposure of runoff products and more contrast image of the blockage area. In the lower dashed rectangle, the white chevron sign shows main blockage product, which position is consistent with T4 interruption in G20T4G8 sequence (290 nt). In these experiments, first the elongation complexes stalled at the end of T-less sequence about 37 nt downstream from T7 promoter were obtained by transcription in the presence of U-less mixture, and then UTP and heparin were added to allow further elongation, and to suppress further initiation, respectively (See *SI Methods* for details). After addition of the heparin/UTP mixture the transcription reaction mixture was placed at 37 °C, and the aliquots were taking and mixed with the stop solution in time intervals indicated above the lanes in the figure. Time "0" (lanes 1 and 6) actually corresponds to about 1 min at RT during mixing and brief spinning of the reaction mixture. It is seen that the reaction is essentially completed during this time, and the percentage of blockage does not change significantly during further incubation (see the graph at *Lower Right*) suggesting that the blockage is irreversible, at least at the time scale of the experiment.

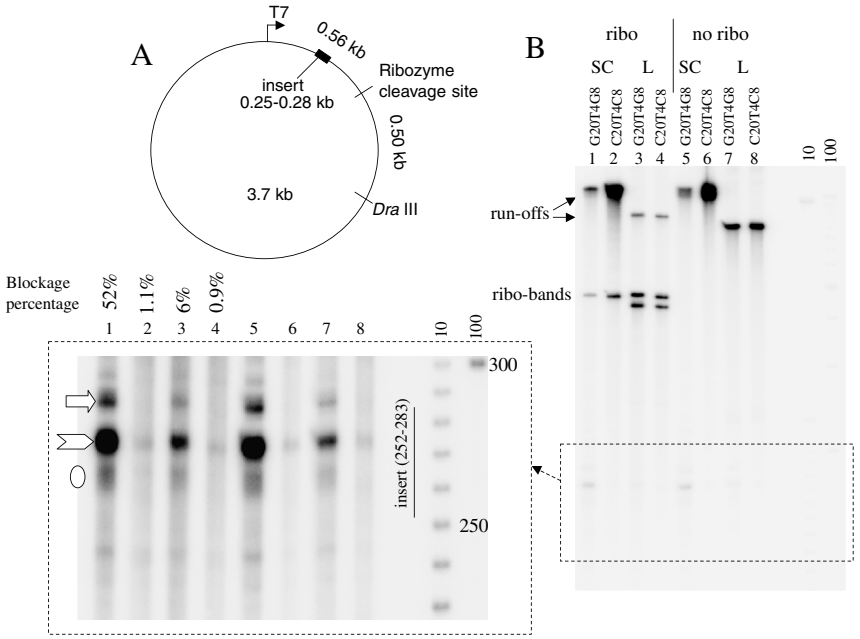

Fig. S4. Effect of negative supercoiling (sc) on transcription blockage. (*A*) Scheme of the ribozyme-containing plasmid used for measuring the effect of supercoiling on blockage. The ribozyme cleavage site is localized 0.56 kb from the T7 promoter and 0.5 kb from the DraIII cleavage site, which was used to obtain the linear substrate. Consequently, with linear substrates ribozyme cleavage yields two products, the longer one (0.56 kb) corresponds to the distance from the promoter to ribozyme cleavage site, and the shorter one (0.5 kb) corresponds to the distance between the ribozyme cleavage site and the end of template at the Dra III cleavage site, plus residual uncleaved DraIII-run-off. From the ratio of cleaved and uncleaved runoff we estimated the probability of cleavage, which was 80%. In the case of circular (including sc) ribozyme-containing plasmid, transcription yields the 0.56-kb product plus long plasmid-size products. This long plasmid-size product is not resolved from the noncleaved runoff product under our conditions; thus in the case of supercoiled DNA, only one cleavage product is expected to be distinguishable. (*B*) Transcription experiment with substrates with (lanes 1–4) and without (lanes 5–8) ribozyme. As predicted from the scheme in *A*, the linear DNA template produces two closely migrating ribozyme cleavage products (ribo-bands), whereas in supercoiled DNA, only one riboband is seen. The blockage band area in the dashed-lined rectangle is shown after a higher exposure. The block arrow, chevron, and oval show repeat-exiting, interruption, and diffuse blockage bands, respectively. The percentages of blockage, shown above lanes with ribozyme-containing substrates, were calculated from Eq. **S2** (only repeat-exiting and interruption bands were taken into account). It is seen that for G20T4G8 insert blockage signals strongly increase with negative supercoiling, suggesting involving some unusual structure in the blockage effect. In contrast, a weak blockage signal within C20T4C8 insert does not depend on supercoiling, suggesting that it is some purely sequence effect, which does not involve a structure formation.

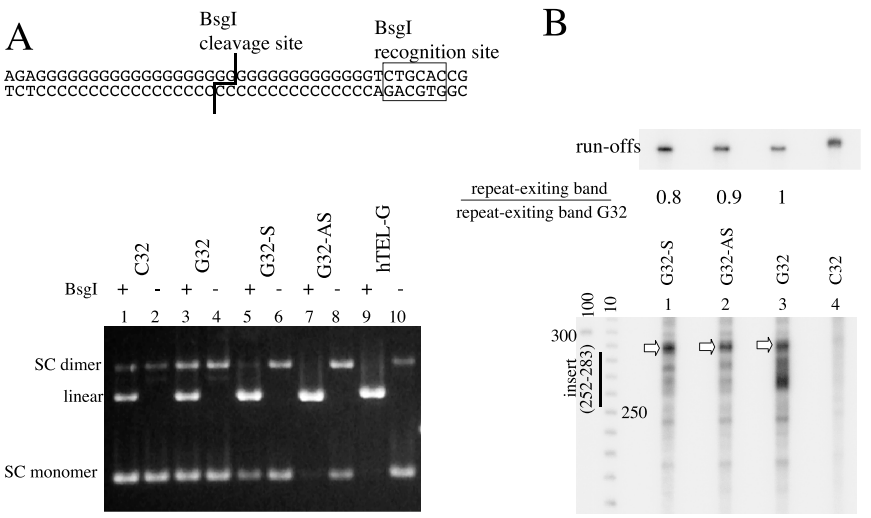

Fig. S5. Transcription blockage does not correlate with triplex-forming potential. (*A*) Protection from BsgI cleavage is consistent with expected triplex-forming potential. BsgI is a restriction enzyme that cleaves arbitrary sequences in double-stranded DNA that are localized at a fixed distance from its specific recognition site (see the scheme at the top with G32 sequence as an example). In this case, it cleaves within the G32 insert or its derivatives, converting supercoiled plasmid into the linear form. If the cleavage site occurs within an unusual DNA structure (for example, a triplex), this would inhibit the cleavage, and this inhibition will increase with increased formation of the structure. It can be seen that the protection (i.e., the ratio of remaining supercoiled to linear DNA) is much stronger for G32-S than for G32-AS plasmid (in the latter, the cleavage is practically complete). Protection in G32 and C32 is stronger than in G32-S, most likely because PyPuPu triplexes formed in mixed G,A sequences are less stable than triplexes formed by pure G sequences. The insert hTEL-G contains the sequence $(TTAGGG)_{17}$, which under these conditions does not form any unusual structure, and, as expected, does not provide any protection. (*B*) Symmetric and asymmetric G-to-A substitutions in the G32 sequence have the same effect on blockage. White block arrows show exit bands, and the numbers above the lanes show the ratio of intensities for repeat-exiting bands of G32-S and G32-AS to the repeat-exiting band of G32. All repeat-exiting band intensities are normalized to the intensities of corresponding runoff bands. These results were for linear DNA templates.

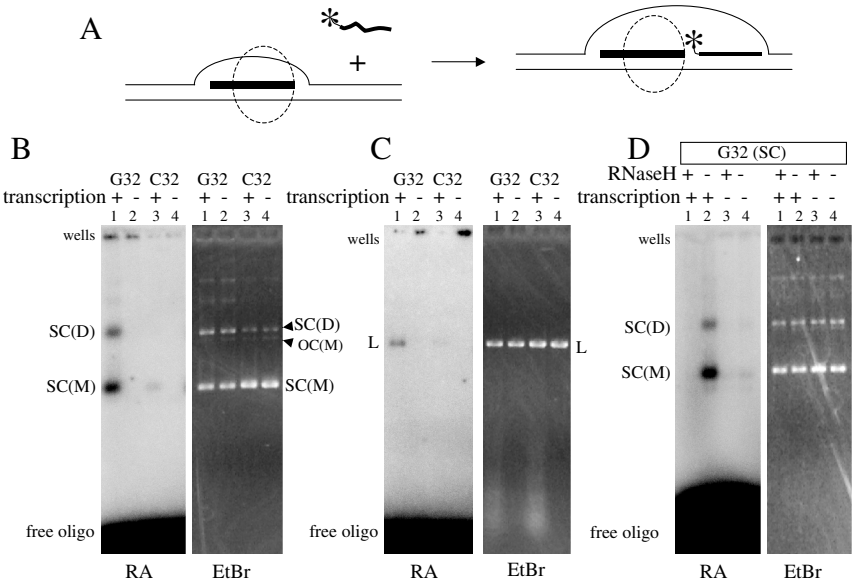

Fig. S6. Transcription-dependent oligonucleotide hybridization is much stronger for the sequence for which the blockage is observed. (*A*) Scheme of the hybridization experiment. Plasmid DNA is shown in thin lines, RNA is shown as a thick line, the radioactively labeled DNA oligonucleotide TCTGCACCGTGGTCGAGAGAT, which is complementary to the template strand immediately downstream of all the inserts used in this study, is shown as a line with asterisk, which symbolize a radioactive labeling. Before the oligonucleotide hybridization the transcription reaction was stopped by EDTA and treated by RNaseA/T mixture, which digests RNA not base-paired with DNA (see *SI Methods*). T7 RNAP (shown as a dashed-line circle) could either remain bound to RNA-DNA hybrid, or dissociate. (*B*) Hybridization with supercoiled DNA. (*C*) Hybridization with linear DNA. (*D*) RNaseH treatment prior to addition of the oligonucleotide abolishes the oligonucleotide hybridization. SC, OC, and L designate negatively supercoiled, open circular (nicked), and linear plasmids, and M and D designate monomer and dimer, respectively. "RA" stands for radioautograph; EtBr stands for ethidium bromide staining.

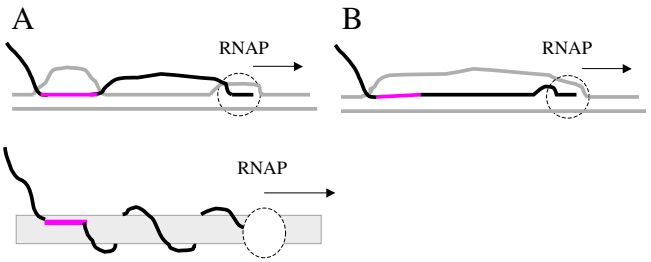

Fig. S7. Possible pathway of transcription elongation after stable R-loop formation. RNAP is shown as a dashed-lined circle, DNA is shown in gray, the RNA region that forms stable R loop is shown in magenta; the rest of the RNA is shown in black. (*A*) RNAP exits the "R-loop mode" and continues transcription in a normal way (i.e., with separation of the RNA/DNA hybrid). In this case, because the position of the R loop is fixed relative to the DNA duplex, and the RNAP rotates relative to DNA during transcription, this pathway will lead to wrapping of the nascent transcript around the DNA. (*B*) RNAP continues R-loop synthesis into ordinary downstream DNA sequences.

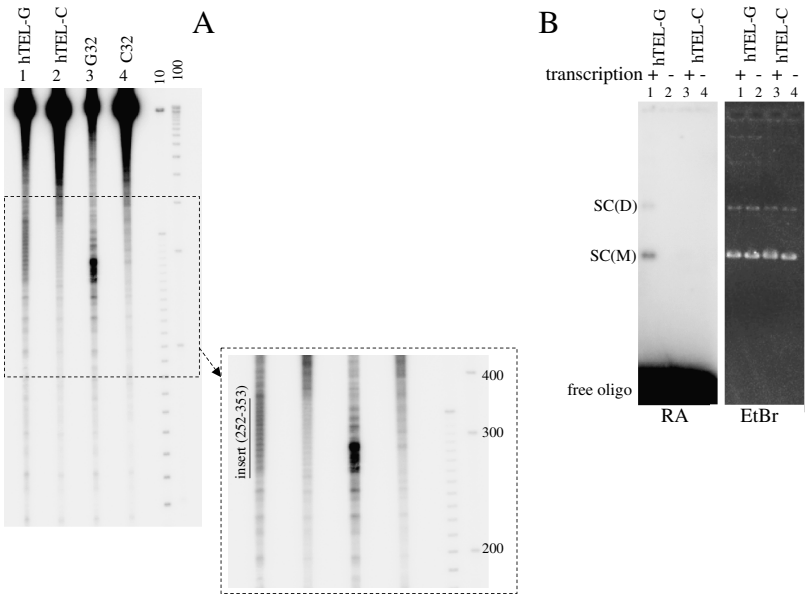

Fig. S8. Orientation-dependent partial transcription blockage and enhanced oligonucleotide hybridization for the human telomeric sequence. Inserts hTEL-G and hTEL-C contain the sequences $(TTAGGG)_{17}$ and $(CCCTAA)_{17}$, respectively. (*A*) Transcription of supercoiled DNA templates. There is an increased density of truncated transcription products within the hTEL-G, but not the hTEL-C insert (compare lanes 1 and 2), which could be interpreted as an increased probability of blockage within the insert. This effect was much less pronounced than for the G32 versus C32 insert (lanes 3 and 4), and we were able to detect it only in supercoiled DNA. (*B*) Transcription-dependent oligonucleotide hybridization. Experiments and designations are similar to those in Fig. S6, except the complementary oligonucleotide (i.e., the one that binds to the nontemplate instead of the template strand immediately downstream of the insert) was used.