Mechanisms and implications of transcription blockage by guanine-rich DNA sequences

Boris P. Belotserkovskii[a], Richard Liu[a], Silvia Tornaletti[b], Maria M. Krasilnikova[c], Sergei M. Mirkin[d,1], and Philip C. Hanawalt[a,1]

[a]Department of Biology, Stanford University, Stanford, CA 94305; [b]Department of Anatomy and Cell Biology, University of Florida, Gainesville, FL, 32610; [c]Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA 16802; and [d]Department of Biology, Tufts University, Medford, MA 02155

Contributed by Philip C. Hanawalt, June 4, 2010 (sent for review April 12, 2010)

Various DNA sequences that interfere with transcription due to their unusual structural properties have been implicated in the regulation of gene expression and with genomic instability. An important example is sequences containing G-rich homopurine-homopyrimidine stretches, for which unusual transcriptional behavior is implicated in regulation of immunogenesis and in other processes such as genomic translocations and telomere function. To elucidate the mechanism of the effect of these sequences on transcription we have studied T7 RNA polymerase transcription of G-rich sequences in vitro. We have shown that these sequences produce significant transcription blockage in an orientation-, length- and supercoiling-dependent manner. Based upon the effects of various sequence modifications, solution conditions, and ribonucleotide substitutions, we conclude that transcription blockage is due to formation of unusually stable RNA/DNA hybrids, which could be further exacerbated by triplex formation. These structures are likely responsible for transcription-dependent replication blockage by G-rich sequences in vivo.

R-loops | DNA supercoiling | Hoogsteen base pairing | inosine | 7-deazaquanosine

Sequence-specific modulation of transcription, including transcription blockage or impediment, plays an important role in DNA transactions, for example, transcription-related mutagenesis and recombination (reviewed in refs. 1 and 2) and could also be responsible for several severe genetic diseases (reviewed in refs. 3–5).

Among the DNA sequences that could affect transcription are GC-rich homopurine-homopyrimidine (hPu/hPy) stretches. These sequences could form unusual DNA structures, including triplexes and G quadruplexes (reviewed in refs. 3–5), which have been implicated in several transcription-dependent phenomena (for example, see refs. 6–9).

Another important property of these sequences is a dramatic asymmetry in the stabilities of RNA/DNA duplexes: The rPu/dPy duplex is significantly more stable, whereas the rPy/dPu duplex is less stable than a DNA/DNA duplex of the same sequence (10). The increased stability of rPu/dPy duplexes is likely responsible for stable R-loop formation by these sequences (11), although alternative DNA structures might also be involved (8, 12, 13).

The simplest example of GC-rich hPu/hPy sequences, the G_n/C_n repeats, is abundant in various genomes, including transcribed domains (14, 15).

The G_{32}/C_{32} stretch was previously shown to stall DNA replication in *Escherichia coli* plasmids in vivo (16). Remarkably, this effect was observed only when the sequence was transcribed, which led to a model stipulating that this sequence stalled an elongating RNA polymerase, and the stalled transcription complex, in turn, blocked the replication machinery (16).

To elucidate the mechanism of transcription blockage by this sequence, we have studied its effect on T7 RNA polymerase (T7 RNAP) transcription in vitro, using various sequence modifications and solution conditions that allowed us to discriminate between possible DNA and RNA structures. We conclude that transcription stalling is triggered by the unusual stability of rG/dC duplexes, leading to R-loop formation, which could be additionally stabilized by DNA triplex formation between the displaced nontemplate DNA strand and duplex DNA.

Results

G_n/C_n Stretches Cause Transcription Blockage in a Length-, Orientation- and Supercoiling-Dependent Manner. Typical results of an in vitro transcription experiment are shown in Fig. 1. Transcription was performed in the presence of radiolabeled NTP; thus, the labeling is roughly proportional to the transcript lengths. The most intense signal, at the top of the gel, corresponds to runoff products, and the additional rapidly migrating products correspond to truncated transcripts, which we interpret to be a consequence of transcription blockage.

There is no evidence of blockage within the C32 sequence (Fig. 1*A*, lane 1), whereas for the G32 sequence (Fig. 1*A*, lane 2) there are well-pronounced blockage signals that start as a diffuse band (marked as a white oval) about 10 bp into the insert and ending with a sharper band (marked as a white block arrow) at, or closely following, the downstream flank of the insert. This sharp band is a common feature of the blockage patterns observed in this study, and we refer to it as a "repeat-exiting band." The transcription blockage products comprise about 12% of total transcription products for linear DNA with G32 insert (see *SI Methods* for calculation). Some minor blockage signals can be additionally observed upstream and downstream from the insert, especially for the negatively supercoiled DNA templates (see below).

For shorter inserts, the intensity of the repeat-exiting band decreases and the diffuse band becomes undetectable (Fig. 1*A*, lanes 3–6; see also the graph in Fig. S1). Surprisingly, while a long C sequence (C32) causes no blockage (Fig. 1*A*, lane 1), small interruptions of C4 within the G32 sequence (inserts G20C4G8 and G8C4G20) produce additional sharp blockage signals (shown by white chevrons), whose positions roughly coincide with the positions of the interruptions (Fig. 1*A*, lanes 7 and 8; see also Fig. S2 for mapping). A similar (or even stronger) effect is produced by the T4 interruption (sequence G20T4G8; Fig. 1*A*, lane 10). The latter observation suggests that the additional blockage signal produced by the C4 interruption is not due to interactions between C and G in the DNA insert or its RNA transcript. Note that the oligoT sequence might, in principle, provide some

Author contributions: B.P.B., S.T., M.M.K., S.M.M., and P.C.H. designed research; B.P.B., R.L., and M.M.K. performed research; B.P.B., S.M.M., and P.C.H. analyzed data; and B.P.B., S.M.M., and P.C.H. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: hanawalt@stanford.edu or sergei.mirkin@tufts.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1007580107/-/DCSupplemental.

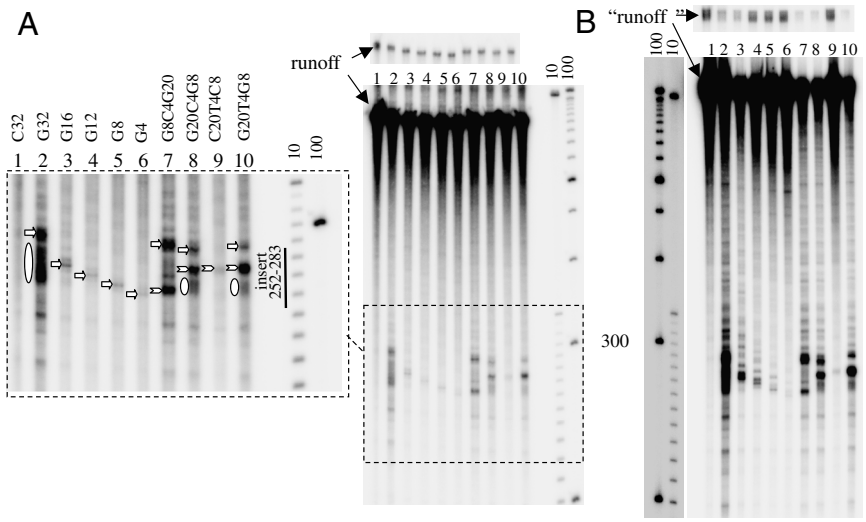

Fig. 1. Transcription blockage by various G-rich inserts. Lanes designated by vertically written "10" and "100" correspond to denatured 10 and 100 nucleotide DNA size markers, respectively. Above the full gel image, lower exposures for the runoff bands are shown. (*A*) Linear DNA template. The dashed-lined box shows a higher exposure for the blockage products. At the top of the box, sequences of the inserts (nontemplate strands) are shown above the corresponding lanes, and within the box, block arrows, chevrons, and ovals show repeat-exiting, interruption, and diffuse blockage bands, respectively. (*B*) The same experiment as in *A*, but for negatively supercoiled DNA templates. Note that for the supercoiled DNA, we put the term "runoff" within quotation marks, because actually it is likely to be a heterogeneous mixture of long transcription products obtained due to spontaneous transcription termination somewhere within the circular plasmid.

blockage signal by itself because of the very low stability of the rU/dA duplex (17, 18). However, this does not seem to be the case, because when the T4 interruption was imbedded in the C32 run (C20T4C8 sequence) instead of the G32 run, it produced little if any blockage signal (Fig. 1*A*, lane 9).

The single-round transcription experiments (see *SI Methods* and Fig. S3) suggest that the blocked RNAP is not able to resume transcription after blockage, at least within the characteristic duration of experiments performed in this work. It remains to be established whether this blockage is accompanied by RNAP dissociation from the template.

The patterns of the blockage signals in the case of the negatively supercoiled (sc) DNA template (Fig. 1*B*) were similar to those for the linear DNA template (Fig. 1*A*). However, in contrast to linear templates, the total yield of "runoff products" was significantly decreased for the plasmids with blocking sequences compared to that for the plasmids without blocking inserts, indicating that the blockage is significantly higher for scDNA than for linear DNA. We estimated that the blockage for the G20T4G8 insert was about 1 order of magnitude greater for scDNA than for linear DNA (Fig. S4).

Both Upstream and Downstream G Stretches Contribute to Blockage at a C4 Interruption. Short Py interruptions imbedded in the G32 sequence produce additional blockage signals, which we call "interruption blockage signals" (Fig. 1, lanes 7, 8, and 10). For sequences G8C4Gn, the interruption blockage signals strongly decrease when the length of the downstream G stretch was reduced from 16 to 4 nt, as well as when the upstream G stretch G8 was "scrambled" by G-to-C inversions (Fig. 2). Thus, both upstream and downstream G stretches contribute to blockage at the C4 interruption.

The Intensity of the Blockage Does not Correlate with the H-DNA-Forming Potential of the G32 Sequence Derivatives. One of the models for transcription blockage by homopurine-homopyrimidine sequences involves the formation of the intramolecular triplex (H-DNA) behind a moving RNA polymerase that causes transcription blockage immediately downstream of the insert (6, 7). To determine whether this mechanism could be responsible for the blockage observed for the G32 sequence, we took advantage of the fact that H-DNA formation requires mirror symmetry of the sequence (reviewed in ref. 19). Thus, a sequence based upon a G32 run containing several G-to-A substitutions, symmetrically positioned relative to its center (G32-S: $G_5A_1G_5A_1G_8A_1G_5A_1G_5$), should form H-DNA much more readily than a sequence with an equal number of asymmetric G-to-A substitutions (G32-AS: $G_5A_1G_5A_1G_6A_1G_2A_1G_{10}$). A restriction protection assay confirmed that the G32-S forms H-DNA more readily than G32-AS (Fig. S5*A*). However, the symmetric and asymmetric substitutions had the same effects on transcription blockage: The overall intensity of the blockage bands within the insert decreased about 3-fold in comparison with the original G32 sequence, whereas the intensity of the repeat-exiting band was

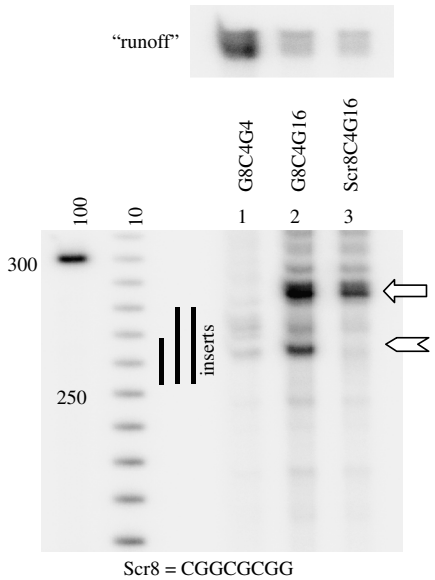

Fig. 2. Both upstream and downstream G stretches contribute to the blockage signal at the C4 interruption. The block arrow and the chevron show the repeat-exiting and interruption blockage bands, respectively. The interruption blockage band is well-pronounced only for the sequence G8C4G16. Presented results are for scDNA.

roughly the same as that for G32 (Fig. S5*B*). Thus, the blockage does not correlate with the H-DNA-forming potential of the sequence.

Transcription Blockage Is not Sensitive to the Type of Monovalent Cations in the Transcription Reaction. Sufficiently long G stretches can form quadruplexes comprised of G quartets (reviewed in ref. 20), which in our case could occur either in the nontemplate strand (8, 13) or in the nascent transcript (21), and could contribute to the transcription blockage. To determine whether this is the case, we used the sensitivity of quadruplex formation to the type of monovalent cation: For example, quadruplexes are much more readily formed in the presence of potassium (K^+) than in the presence of lithium (Li^+) cations [(22), (23), and references therein]. If quadruplex formation contributes to the transcription blockage, then the blockage should be much stronger in the presence of K^+ than in the presence of Li^+ ions. We found that the transcription blockage was similar for reactions performed in the presence of either cation (Fig. 3*A*), suggesting that quadruplex formation does not contribute to the observed blockage. However, this does not exclude the possibility that a quadruplex could form within the nontemplate strand after the blockage occurred, for example, as a consequence of R-loop formation (see below).

Substitution of Inosine, but not 7-Deazaguanosine, for Guanosine in the Transcription Reaction Abrogates Blockage. Inosine and 7-deazaguanosine are both guanosine analogs (i.e., they form Watson–Crick base pairs with cytidine) with impaired quadruplex-forming abilities (see refs. 21 and 22, and references therein). Inosine also forms significantly less stable Watson–Crick base pairing in comparison with guanosine because it forms two hydrogen bonds with cytidine instead of three. In contrast, 7-deazaguanosine forms the same three hydrogen bonds, and, consequently, a Watson–Crick base pair of a similar stability, as guanosine. Thus, the replacement of guanosine by inosine in the transcript would destabilize both Watson–Crick duplexes and quadruplexes containing the nascent transcript, whereas replacing guanosine by 7-deazaguanosine would destabilize only quadruplexes. Fig. 3*B* shows that the transcription blockage completely disappears when guanosine is replaced by inosine, but not with 7-deazaguanosine. These data suggest that some form of Watson–Crick duplex, either RNA/DNA duplex within the R loop, or an RNA/RNA hairpin, must be responsible for the transcription blockage. Because there are no evident self-complementary regions in the transcript in the vicinity of the G32 insert, we conclude that the most likely candidate for the blocking structure is an R loop. In principle, R loops could be additionally stabilized by Hoogsteen base pairing between the displaced DNA strand and RNA within the R loop ["collapsed R loop" (12)]. Because this base pairing involves the N7 nitrogen, it would be destabilized by 7-deazaguanosine (but not by inosine) for guanosine substitution in nascent RNA within the R loop. That can explain the somewhat weaker blockage for 7-deazaguanosine compared with guanosine. Still, complete abrogation of transcription blockage in the presence of inosine, but not 7-deazaguanosine, suggests that Hoogsteen pairing plays only a minor role in the structure stabilization, which is largely stabilized by Watson–Crick base pairing in a RNA-DNA hybrid.

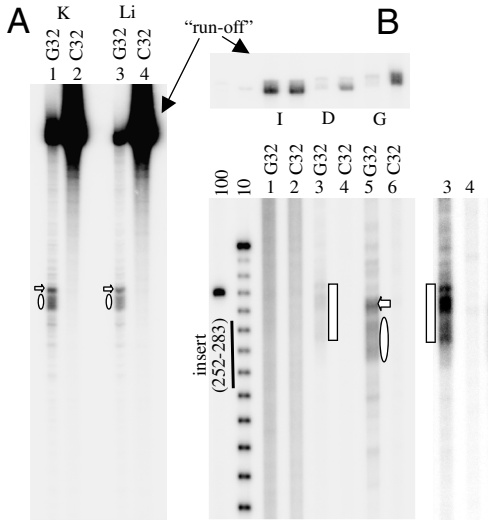

Fig. 3. Effect of monovalent cations and nucleoside substitution on transcription blockage. (*A*) The blockage is not sensitive to monovalent cations. In the standard transcription buffer (see *Materials and Methods*), NaCl was replaced by 83 mM of either KCl or LiCl. The ratio of the intensity of the blockage band (normalized to runoff) for K to the intensity of the blockage band (normalized to runoff) for Li was 0.6. Thus, K does not facilitate blockage in comparison with Li. These results are for scDNA. (*B*) Substitution of ITP, but not 7-deaza-GTP for GTP abolishes blockage. Guanosine (G) was replaced by either inosine (I) or 7-deazaguanosine (D) in transcription reaction with scDNA substrates. It is seen that for inosine there is no detectable difference between blocking insert G32 and control C32. In contrast, in the case of 7-deazaguanosine there is blockage and decrease in runoff product in the case of the G32 insert, though the blockage signal (shown by white rectangle) is somewhat weaker, more smeary, and more shifted downstream from the insert in comparison with guanosine. An additional more contrasted image for lanes 3 and 4 with 7-deazaguanosine is shown at *Right*.

Transcription Facilitates Oligonucleotide Hybridization to DNA near the Blockage Site. Stable local distortions in double-stranded DNA facilitate transient DNA opening and, consequently, oligonucleotide hybridization in the vicinity of the distortions (24). One would therefore anticipate the preferred hybridization of a complementary oligonucleotide to a sequence in the vicinity of stalled transcription complexes and/or stable R loops. Results shown in Fig. S6 (see *SI Methods* for details) confirm this prediction: Hybridization of an oligonucleotide complementary to a template DNA strand of the sequence immediately downstream of the insert is much stronger for the orientation in which transcription blockage occurs. Furthermore, treatment with RNase H prior to adding the oligonucleotide abolishes the hybridization (Fig. S6*D*), showing that a stable RNA-DNA hybrid is necessary for the oligonucleotide hybridization.

Transcription-Dependent Replication Blockage in *E. coli* Correlates with T7 RNAP Transcription Blockage in Vivo. Krasilnikova et al. (16) first observed transcription-dependent DNA replication blockage at G_n/C_n stretches in *E. coli*, which was pronounced only when nontemplate (sense) strand for transcription was G_n. These data are in accord with our results for T7 RNAP transcription blockage in vitro, suggesting the same general mechanism for transcription blockage in these two systems. To further support this idea, we now show that replication is blocked to a similar extent at a noninterrupted G32 stretch, or at the same stretch with the C4 interruption (Fig. 4), which correlates with the results for transcription blockage (Fig. 1).

Discussion

Mechanism of Transcription Blockage by G_n/C_n Sequences and Their Derivatives. We demonstrated strong transcription elongation blockage when T7 RNAP encounters the G_{32}/C_{32} DNA sequence. The blockage occurs only in one orientation, in which extrastable rPu/dPy RNA/DNA duplex (10) is formed. The importance of increased stability of the RNA/DNA duplex for blockage is further supported by results with various nucleoside substitutions in the transcription reaction (Fig. 3).

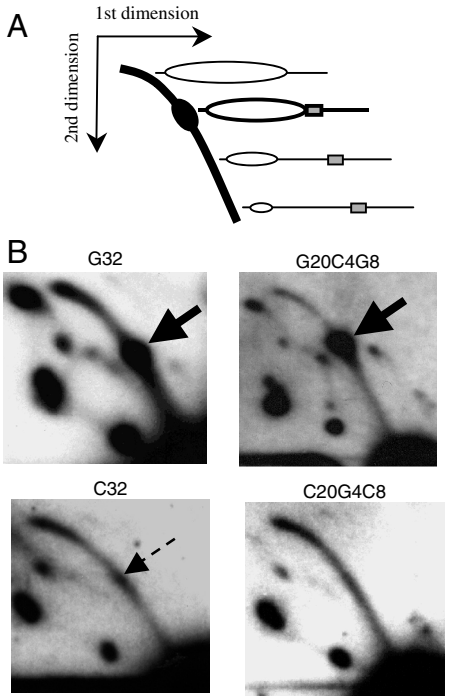

Fig. 4. Transcription-dependent replication blockage in *E. coli* cells. (*A*) A general scheme for the 2D gel electrophoresis of bubble-like replication intermediates. Replication fork moves unidirectionally from left to right producing replication intermediates with gradually decreasing electrophoretic mobility, which together form an arch on the gel. Fork stalling at an obstacle (gray rectangle) leads to accumulation of corresponding replication product, producing a bulge on the arch (black oval), which intensity reflects the "strength" of the replication blockage. (*B*) Gel electrophoresis of replication intermediates. The sequences above the figures correspond to the nontemplate (sense) strand for transcription, which is also the lagging strand template for DNA replication. Bulges on the replication arcs (marked by arrows) indicate replication stall sites. The bulges are equally well-pronounced for the G32 insert and its interrupted derivative, G20C4G8, whereas they are barely detectable for the C32 insert and its interrupted derivative, C20G4C8.

It might seem counterintuitive that an increased stability of the RNA/DNA hybrid would interfere with transcription, because the formation of a more stable base pair between the incoming ribonucleotide and the DNA template should facilitate RNA synthesis. Note, however, that in addition to creating a stable RNA/DNA hybrid during transcription, the RNA polymerase must also unwind this hybrid prior to extruding and releasing the nascent RNA [(25, 26); also reviewed in ref. 27)]. The elevated stability of the RNA/DNA hybrid within the transcription complex could impede its unwinding, thereby preventing the release of RNA from the complex and trapping the RNAP. This model predicts that RNA polymerase should readily enter G_n/C_n stretches, but when the G transcript exceeds the size of the normal RNA/DNA hybrid within the transcription complex [roughly 8 nt (25)], T7 RNAP will fail to efficiently unwind the extrastable RNA/DNA hybrid and stall.

This interpretation is consistent with our experimental data for G32, for which truncated transcription products begin to appear about 10 bp downstream from the start of the G stretch. The "worst" situation for the RNA polymerase might develop when it reaches the downstream end of the G stretch and begins to transcribe a flanking sequence that forms a less stable mixed sequence RNA/DNA hybrid, or an especially unstable rPy/dPu hybrid, while still attempting to unwind the preexisting extrastable rPu/dPy hybrid. This consideration could explain the sharp blockage signals at Py interruptions, or closely downstream from the end of the G stretches. In accord with this idea, one of the natural arrest sites for RNAP II contains a short homopurine sequence closely upstream of the site of the transcript termination (28). Note that this model does not require the blockage to occur exactly at the end of the G stretch: Several non-G bases could still be added to the transcript.

If the mechanism described above were the only explanation for the transcription blockage, the strength of the blockage would not significantly increase when the length of the G stretch exceeds that of the RNA-DNA hybrid in the elongation complex. Apparently, this is not the case: The strength of the "exiting" blockage signal strongly increases when the length of the G stretch is increased from 16 nt to 32 nt, which is far beyond the normal size of the RNA-DNA hybrid in the elongation complex. Thus, some type of structure that extends beyond the transcription complex must additionally contribute to the blockage. A likely candidate for this structure is an extended R loop, in which the nascent RNA rehybridizes with the template DNA strand behind the elongating RNAP. Indeed, evidence for the formation of R loops was obtained for transcripts containing G stretches, i.e., when extrastable RNA/DNA duplex was formed (11). R-loop formation could exacerbate transcriptional blockage by preventing the rewinding of a DNA duplex upstream of the RNA polymerase, which contributes to transcription elongation (29, 30). Also, we predict that stable R-loop formation would inhibit resumption of normal transcription downstream (*SI Discussion* and Fig. S7). The enhanced blockage by negative DNA supercoiling together with the facilitated oligonucleotide hybridization in the vicinity of the blocking sequence strongly support the R-loop model.

An alternative model could be the formation of a triplex (H-DNA) behind the RNAP, which might sterically trap RNAP at the downstream flank of the hPu/hPy insert (6, 7). However, our results with symmetric and asymmetric G-to-A point substitutions argue against this model for the G32 sequence and its derivatives. Note that for other triplex-forming sequences, the results of similar substitutions did implicate triplex formation behind the RNAP (31). It is therefore possible that the outcome of the competition between the R-loop and triplex formation behind the RNAP could be sequence-dependent.

We believe, however, that in our system, a triplex could form in front, rather than behind, the RNAP: The pronounced blockage at an interruption within the Gn stretch depends upon sequences both upstream and downstream from this interruption, suggesting that a structure responsible for the blockage involves both sequences (Fig. 2). A model that could explain these effects was first proposed for replication blockage at hPu/hPy repeats (32). This model, called "suicidal replication," postulates that a displaced nontemplate strand upstream from the blockage site can fold back to form a triplex with the downstream duplex. We believe that similar events could occur during transcription and contribute to the interruption and the "diffused" blockage bands. Fig. 5 summarizes all proposed pathways leading to transcriptional blockage by hPu/hPy sequences.

Biological Relevance. We found that relatively short oligoG stretches present a novel type of blockage signals for transcription. The observed blockage is surprisingly strong; for the G32 sequence there is up to 75% RNAP blockage in negatively supercoiled DNA. For comparison, transcriptional blockage was only 5% at a Z-DNA-forming sequence of comparable length, (CG) 14, in negatively supercoiled DNA (33). Furthermore, interruptions in the G stretch did not reduce the blockage signal, and, moreover, the blockage signal was "concentrated" at interruptions. This tolerance to sequence interruptions greatly increases the number of naturally occurring sequences that could block transcription in vivo by the mechanism described above. Transcription blockage is known to trigger various types of transcription-associated mutagenesis (reviewed in refs. 1, 34, and 35), possibly including "gratuitous" transcription-coupled repair (2). Another interesting possibility is that these sequences might

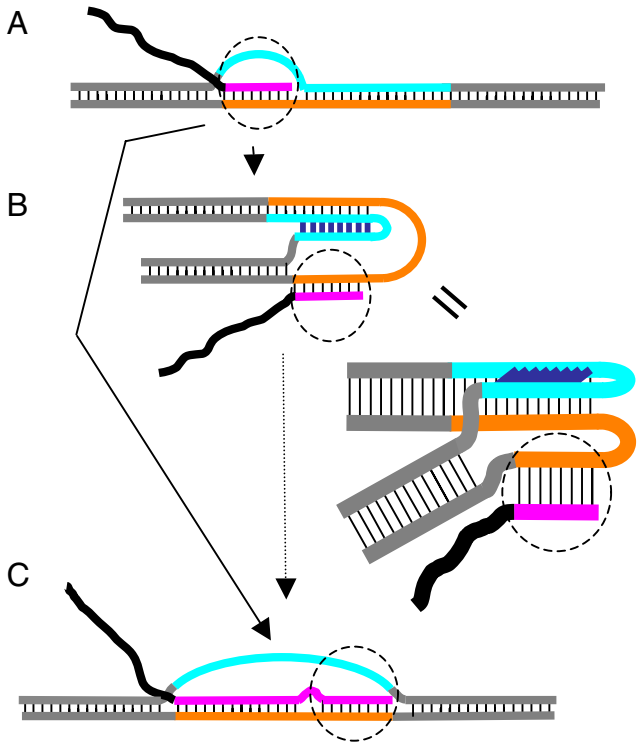

Fig. 5. Suggested model for transcription blockage by G-rich homopurine-homopyrimidine sequences. RNAP is shown as a dashed-line circle, the hPu DNA, hPyDNA, regular DNA, hPuRNA and regular RNA sequences are shown in cyan, orange, gray, magenta, and black, respectively. Watson–Crick and Hoogsteen base pairs are shown as short, thin, black lines, and thicker dark blue lines, respectively. (*A*) RNAP transcribes into the hPu/hPy sequence without any obstacles, until the hPu transcript reaches the size of the RNA/DNA hybrid inside the transcriptional complex, and the extrastable RNA/DNA duplex has to be unwound to provide nascent RNA extrusion. This might cause a transcription impediment or partial blockage, which could be exacerbated if the RNAP encounters an interruption or end of the G stretch, because in that case it begins to synthesize a less stable RNA/DNA duplex, while still unwinding the extrastable RNA/DNA duplex to extrude nascent RNA. (*B*) The blockage could be further exacerbated if the nontemplate strand interacts with the downstream duplex forming a triplex. A more realistic drawing of the triplex is shown at *Right*. (*C*) During transcription of the pyDNA template, the RNA could rehybridize with the DNA template forming an R loop, which also could exacerbate the blockage (see the text). Triplex and R-loop pathways could be either mutually exclusive, each occurring with a certain probability during transcription, or the triplex formation could be reversible and precede the R-loop formation.

exacerbate transcription blockage by DNA lesions that are not per se strong blockage signals for transcription. This could extend the transcription-coupled repair mechanism (36–38) to a wider range of DNA lesions.

Interestingly, we observe some transcription blockage and increased oligonucleotide hybridization for the human telomeric sequence $(TTAGGG)_n/(CCCAAT)_n$, in the orientation in which the G-containing transcript is produced (Fig. S8), suggesting that the human telomeric sequence might be prone to R-loop formation. This is in accordance with the recent discovery that human telomeres are transcribed in the orientation in which G-containing transcripts are produced and that the nascent RNA remains bound to telomeres (39). If telomeric sequences are prone to R-loop formation and transcription blockage, various types of transcription-facilitated mutagenesis within telomeres might occur as well. In general, our model predicts that in biological processes in which R loops are implicated, transcription blockage might occur.

Besides transcription per se, genetic instability at DNA repeats could also be caused by the collisions between replication and transcription, which can lead to replication forks stalling and, consequently, facilitate double-stranded break formation and other instabilities. Freely elongating RNAP blocks replication during "head-on" collisions only, whereas stalled RNAP can block replication approaching it both codirectionally and head-on (reviewed in ref. 40). Replication blockage during codirectional transcription and replication collision was observed within G_n/C_n sequences inserted in the plasmids propagated in *E. coli*, which led to the proposal that these sequences stalled elongating RNA polymerases and that this, in turn, blocked replication forks (16). Correlation between transcription-dependent replication blockage in vivo (Fig. 4) and transcription blockage in vitro (Fig. 1) makes it tempting to suggest that similar structures, i.e., R loops and triplexes, are responsible for these effects. Similar phenomena could also occur in eukaryotic cells; for example, extended R loops were recently shown to cause replication blockage in mammalian cells (41). Since hPu/hPy repeats (42) and G-rich stretches (43) are abundant in eukaryotic genomes, the proposed mechanism of transcription-dependent replication blockage could contribute to gross chromosomal rearrangements at these sequences in various genetic processes.

Materials and Methods

DNA Substrates. All inserts were cloned into the *Bam* HI/*Xho* I sites of the pUCGTG-TS plasmid (44) 252 bp downstream from the T7 promoter. Oligonucleotides used for cloning were purchased either from Integrated DNA Technologies or from Midland Certified. For the pG32 plasmid the sequences of oligonucleotides used for cloning were 5′GATCGGTACCTCTAGA**GGGGGGGGGGGGGGGGGGGGGGGGGGGGGG**TCTGCACCGTGG3′ and 3′CCATGGAGATCT**CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC**AGACGTGGCACCAGCT5′. The first and the second strands are the nontemplate and template strands, respectively, for T7 transcription. The G_{32}/C_{32} insert is shown in bold. All other insert sequences are shown in the *Results* section and in corresponding figure legends. The flanking regions were the same for all inserts. In the sequence designation, the sequence of the nontemplate strand is shown, and regular, instead of subscript font was used to indicate the number of nucleotides. For example, G20T4G8 stands for $G_{20}T_4G_8$. The ribozyme-containing versions of the plasmids were obtained by inserting a ribozyme sequence from the RiboCop plasmid (Sigma) into the corresponding plasmid as described (33). Purification of transcription substrates was as described in ref. 31. Unless otherwise stated, in transcription experiments with linear DNA we used plasmid templates linearized by Dra III restriction digestion 0.96 kb downstream from the promoter.

In Vitro T7 RNAP Transcription. Transcription reactions were performed essentially as described in ref. 31. For the details about the buffer conditions, single-round transcription, and transcription-dependent oligonucleotide hybridization conditions, see *SI Text*.

Analysis of Replication Intermediates. All plasmids constructions and experimental procedures are the same as in ref. 16.

ACKNOWLEDGMENTS. We thank Graciela Spivak for critical reading of the manuscript and all members of the Hanawalt lab for helpful discussions. This work was supported by Grant 2R01 CA077712 from the National Cancer Institute, National Institutes of Heath, to P.C.H. and Grant GM60987 from the National Institute of General Medicine to S.M.M.

1. Aguilera A (2002) The connection between transcription and genomic instability. *EMBO J* 21:195–201.
2. Hanawalt PC (1994) Transcription-coupled repair and human disease. *Science* 266:1957–1958.
3. Mirkin SM (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol* 16:351–358.
4. Wang G, Vasquez KM (2006) Non-B DNA structure-induced genetic instability. *Mutat Res* 598:103–119.
5. Bacolla A, Wells RD (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog* 48:273–285.
6. Grabczyk E, Fishman MC (1995) A long purine-pyrimidine homopolymer acts as a transcriptional diode. *J Biol Chem* 270:1791–1797.

7. 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.
8. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* 18:1618–1629.
9. 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.
10. Roberts RW, Crothers DM (1992) Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science* 258:1463–1466.
11. Roy D, Yu K, Lieber MR (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol* 28:50–60.
12. Reaban ME, Lebowitz J, Griffin JA (1994) Transcription induces the formation of a stable RNA. DNA hybrid in the immunoglobulin alpha switch region. *J Biol Chem* 269:21850–21857.
13. Tornaletti S, Park-Snyder S, Hanawalt PC (2008) G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* 283:12756–12762.
14. Zhao Y, O'Neil NJ, Rose AM (2007) Poly-G/poly-C tracts in the genomes of Caenorhabditis. *BMC Genomics* 8:403.
15. Boyer JC, et al. (2002) Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. *Hum Mol Genet* 11:707–713.
16. Krasilnikova MM, Samadashwily GM, Krasilnikov AS, Mirkin SM (1998) Transcription through a simple DNA repeat blocks replication elongation. *EMBO J* 17:5095–102.
17. Macdonald LE, Zhou Y, McAllister WT (1993) Termination and slippage by bacteriophage T7 RNA polymerase. *J Mol Biol* 232:1030–1047.
18. Martin FH, Tinoco I, Jr (1980) DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res* 8:2295–2299.
19. Mirkin SM, Frank-Kamenetskii MD (1994) H-DNA and related structures. *Annu Rev Biophys Biomol Struct* 23:541–576.
20. Neidle S (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struct Biol* 19:239–250.
21. Kuzmine I, Gottlieb PA, Martin CT (2001) Structure in nascent RNA leads to termination of slippage transcription by T7 RNA polymerase. *Nucleic Acids Res* 29:2601–2606.
22. Williamson JR, Raghuraman MK, Cech TR (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* 59:871–880.
23. Wong A, Wu G (2003) Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5′-monophosphate: A solid-state NMR study. *J Am Chem Soc* 125:13895–13905.
24. Iyer M, Norton JC, Corey DR (1995) Accelerated hybridization of oligonucleotides to duplex DNA. *J Biol Chem* 270:14712–14717.
25. Huang J, Sousa R (2000) T7 RNA polymerase elongation complex structure and movement. *J Mol Biol* 303:347–358.
26. Daube SS, von Hippel PH (1994) RNA displacement pathways during transcription from synthetic RNA-DNA bubble duplexes. *Biochemistry* 33:340–7.
27. Steitz TA (2006) Visualizing polynucleotide polymerase machines at work. *EMBO J* 25:3458–68.
28. Hawryluk PJ, Ujvari A, Luse DS (2004) Characterization of a novel RNA polymerase II arrest site which lacks a weak 3′ RNA-DNA hybrid. *Nucleic Acids Res* 32:1904–1916.
29. Gopal V, Brieba LG, Guajardo R, McAllister WT, 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.
30. Kireeva ML, Komissarova N, Kashlev M (2000) Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J Mol Biol* 299:325–335.
31. 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.
32. Samadashwily GM, Dayn A, Mirkin SM (1993) Suicidal nucleotide sequences for DNA polymerization. *EMBO J* 12:4975–83.
33. 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.
34. Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9:958–970.
35. Saxowsky TT, Doetsch PW (2006) RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? *Chem Rev* 106:474–488.
36. Mellon I, Spivak G, Hanawalt PC (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51:241–249.
37. Mellon I, Hanawalt PC (1989) Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342:95–98.
38. Tornaletti S (2005) Transcription arrest at DNA damage sites. *Mutat Res* 577:131–45.
39. Azzalin CM, Lingner J (2008) Telomeres: The silence is broken. *Cell Cycle* 7:1161–1165.
40. Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71:13–35.
41. Tuduri S, et al. (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* 11:1315–1324.
42. Cox R, Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci USA* 94:5237–5242.
43. Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* 13:1055–1059.
44. Tornaletti S, Patrick SM, Turchi JJ, Hanawalt PC (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.