A TRIPLEX-FORMING SEQUENCE FROM THE HUMAN c-MYC PROMOTER INTERFERES WITH DNA TRANSCRIPTION.

Boris P. Belotserkovskii, Erandi de Silva, Silvia Tornaletti, Guliang Wang, Karen M. Vasquez and Philip C. Hanawalt.

1Department of Biological Sciences, Stanford University, Stanford, CA, USA, 94305.
2Department of Carcinogenesis, University of Texas, M. D. Anderson Cancer Center, Science Park-Research Division, Smithville, TX, USA, 78957
3Current address: Department of Anatomy and Cell Biology, University of Florida, USA, 32610

Running title: Transcription in triplex-forming DNA sequence.
Address correspondence to: Philip C. Hanawalt, Department of Biological Sciences, Stanford University, Stanford, CA, USA, 94305. Fax: 650-725-1848; E-mail: hanawalt@stanford.edu

Naturally occurring DNA sequences that are able to form unusual DNA structures have been shown to be mutagenic, and in some cases the mutagenesis induced by these sequences is enhanced by their transcription. It is possible that transcription-coupled DNA repair induced at sites of transcription arrest might be involved in this mutagenesis. Thus, it is of interest to determine whether there are correlations between the mutagenic effects of such non-canonical DNA structures and their ability to arrest transcription. We have studied T7 RNA polymerase transcription through the sequence from the nuclease-sensitive element of the human c-MYC promoter, which is mutagenic in mammalian cells (Wang and Vasquez, PNAS (2004), 101, 13448-13453). This element has two mirror-symmetric homopurine-homopyrimidine blocks which potentially can form either DNA triplex (H-DNA) or quadruplex structures. We detected truncated transcription products indicating partial transcription arrest within and closely downstream of the element. The arrest required negative supercoiling and was much more pronounced when the Py-rich strand of the element served as the template. The exact positions of arrest sites downstream from the element depended upon the downstream flanking sequences. We made various nucleotide substitutions in the wild-type sequence from the c-MYC nuclease sensitive element that specifically destabilize either the triplex or the quadruplex structure. When these substitutions were ranked for their effects on transcription, the results implicated the triplex structure in the transcription arrest. We suggest that transcription-induced triplex formation enhances pre-existing weak transcription pause sites within the flanking sequences by creating steric obstacles for the transcription machinery.

In addition to its primary function in protein synthesis, transcription plays an important role in gene regulation and genome modification. Transcription through a particular DNA region can also increase the rate of mutation in this region (transcription assisted mutagenesis, or TAM) or create a hot spot for homologous recombination (transcription assisted recombination, or TAR) (reviewed in (1)). Both of these phenomena can occur as a consequence of the
DNA opening during transcription, since single-stranded DNA is more sensitive than double-stranded DNA to attack from a number of agents, including some DNA modifying enzymes. In some cases, transcription induced mutagenesis is enhanced by the formation of an unusually stable RNA-DNA hybrid and/or secondary structure in the non-template strand, both of which stabilize R-loops, thus prolonging DNA opening (for example, (2,3)). These two pathways, however, are not necessarily associated with transcription pausing or arrest.

A special pathway of DNA processing associated with transcription arrest is transcription-coupled DNA repair (TCR)(4,5), reviewed in (6-9). TCR manifests itself as a preferential repair of DNA lesions (for example pyrimidine photodimers) in the template strand versus the non-template strand. This sub-pathway of nucleotide excision repair involves dedicated enzymatic machinery to displace the RNA polymerase and to provide an accelerated recovery of functionally important genes as well as clearing the DNA template for replication and other transactions. According to our current model for TCR, when RNA polymerase becomes arrested upon encountering a lesion in the template strand, the arrested RNA polymerase interacts with specific TCR factors, and this interaction serves as a signal for excision of the lesion and subsequent DNA repair synthesis (reviewed in (6,7)). It was suggested that not only template DNA lesions, but also other factors which arrest transcribing RNA polymerase, might initiate TCR, leading to DNA excision and repair synthesis, resulting in mutations near the area of arrest (6). The report that mutagenesis induced by triplex-forming DNA oligonucleotides, which are known to arrest transcription, appeared to be dependent upon TCR factors, supports this hypothesis (10). The “gratuitous” TCR hypothesis predicts that naturally occurring DNA sequences which present an obstacle for transcription could be hot spots for mutations. Among the sequences that can be obstacles for transcription are those which can form unusual (i.e. non-B-form) DNA structures. These structures are known to interfere with a number of the enzymes that operate on DNA and they cause genomic instability in many systems (reviewed in (11-13)). Our present study focuses upon unusual DNA structure(s), which can be formed in the sequence from the nuclease sensitive element of the human c-MYC promoter (14). This element was shown to be mutagenic in mammalian cells, and it was suggested that formation of an unusual DNA structure, intramolecular triplex or H-DNA, could be responsible for this effect (15).

It is important to point out that although the nuclease-sensitive element from c-MYC gene is localized upstream of c-MYC promoters P1 and P2, which are most commonly used in normal cells in non-translocated alleles of the gene (reviewed in (16)), it is localized downstream of an alternative promoter, P0; consequently the element would be transcribed when transcription is initiated from the P0 promoter. Transcription from the P0 promoter comprises about 1% in normal cells in non-translocated alleles, but it is more abundant in translocated alleles of the gene or in some cancer cells (17,18). Thus, in addition to serving as a model prototype of the effect on transcription of naturally occurring H-DNA-forming sequences, the transcription through this element might have direct biological relevance.

H-DNA forms within homopurine-homopyrimidine mirror-symmetrical repeats or H-palindromes (Fig.1A) (reviewed in (19)). Upon H-DNA formation, the DNA double helix within one half of the H-palindrome denatures into two single strands, and one of these complementary strands folds back to form a DNA triplex with the non-denatured half of the H-palindrome, while the other
strand remains denatured. The presence of denatured regions makes this structure hypersensitive to single-strand specific endonucleases, like S1, which are commonly used for detection of H-DNA. Because H-DNA is topologically equivalent to complete unwinding of the entire region that participates in the structure formation, it is stabilized by negative supercoiling. Within H-DNA, the triplex is stabilized by either Hoogsteen (PyPuPy-triplex) or reverse-Hoogsteen (PuPuPy-triplex) interactions, depending upon whether the pyrimidine or the purine-rich strand participates in the triplex formation. (Sometimes the PyPuPy triplex-containing structure is referred to as H-DNA or H-y-DNA, and the PuPuPy triplex-containing structure is referred to as H*-DNA or H-r-DNA. For simplicity, we will refer to all of these structures as H-DNA). Because the PyPuPy triplex requires cytosine protonation in the third strand, and consequently is stabilized by low pH, the PuPuPy triplex is usually considered to be the more favored structure under physiological conditions, especially for G-rich sequences. Fig. 1B illustrates possible PuPuPy triplex-containing H-DNA structures within the nuclease-sensitive element of the c-MYC promoter.

However, H-DNA is not the only unusual structure able to form at this sequence. If a DNA strand contains more than four closely-localized guanine blocks which have three or more guanines each, this strand can fold into another unusual DNA structure, a DNA quadruplex stabilized by G-quartets (reviewed in (20)) (Fig. 1C). DNA quadruplexes are known to form in many biologically important sequences, including telomeres (reviewed in (20)) and regulatory elements of the c-MYC promoter similar to the one used in the present study (21,22).

Therefore, one has to determine whether one or both of these structures, triplex or quadruplex, is responsible for the effects caused by the G-rich H-palindrome sequence. Here we have studied the effect on T7 RNA polymerase transcription elongation of the wild-type sequence from the nuclease-sensitive element from the c-MYC promoter and similar sequences with various nucleotide substitutions, which allow us to examine the contributions from the triplex and the quadruplex to the observed effects.

**Experimental Procedures**

*Transcription substrates*-All inserts containing the sequence from nuclease-sensitive element of the c-MYC promoter and its various substitutions (Fig.2) were cloned into Bam HI / Xho I sites of pUCGTG-TS plasmid (23) 252 bp downstream from the T7 promoter. Plasmids pWT-ribo and pWT-Rev-ribo were obtained by inserting a ribozyme sequence from RiboCop plasmid (Sigma, Saint Louis, MO) into the corresponding plasmid 0.3 kb downstream from the c-MYC sequence. Plasmids were propagated in DH5α E. coli cells (Invitrogen, Carlsbad, CA), and purified using Qiagene maxi-prep kit (Qiagen Sciences, MD) according to the manufacturer’s instructions, except that the time of the cell lysis step was reduced to several seconds.

When substrates were gel-purified, the Qiaquick gel purification protocol (Qiagen Sciences, MD) was used with the following modifications: (i) To avoid ethidium bromide (EtBr) contamination or UV exposure of the transcription substrates, preparative gels were run without EtBr. The DNA bands of interest were localized by comparison with a gel slice containing small amount of the sample that was cut from the same gel, stained with EtBr and UV irradiated. Thus, the DNA actually used in our experiments was neither stained, nor UV-irradiated. (ii) The agarose-dissolving step of purification was performed for 40 min at room temperature instead of 10 min at 50°C.

To obtain plasmids with various degrees of supercoiling, 2 μg of plasmid DNA
were incubated with 2 units of Wheat Germ Topoisomerase I (Sigma, Saint Louis, MO) and various concentrations of EthBr during incubation for 2 hours in 200 µl of the buffer containing 35 mM TrisHCl, 1 mM EDTA, 2 mM DTT, 1.8 mM spermidine, 6.5% glycerol and 50 µg/ml BSA. Then proteins and EtBr were removed by phenol-chloroform-isoamyl alcohol (25:24:1) extraction, followed by chloroform extraction and ethanol precipitation. Then samples were rinsed twice with 100 µl 70% ethanol, dried and dissolved in 12 µl of TE buffer (pH 7.9). Supercoiling of the samples was analyzed on agarose/TAE gels with various concentrations of chloroquine.

In vitro T7 RNA polymerase transcription - 10 ng of DNA template was incubated with 20 units of T7 RNA polymerase (Promega corporation, Madison, WI) and 16 units of RNasin (Promega, Madison, WI) in 12 µl of buffer containing 33 mM TrisHCl (pH 7.9), 5 mM MgCl₂, 8.3 mM NaCl, 1.7 mM spermidine, 8.3 mM DTT, 0.17 mM of each ATP, CTP, UTP, 0.017 mM of GTP, 10 µCi of α³²P GTP for 30 min at 37°C. The reaction was stopped by adding 94 µl of buffer containing 1% SDS, 35 mM TrisHCl (pH 7.6), 12.5 mM EDTA, 150 mM NaCl, 0.27 mg/ml tRNA (Invitrogen, Carlsbad, CA) and 0.11 mg/ml ProteinaseK (Invitrogen, Carlsbad, CA) and incubated at room temperature for 15 min. Next, 11 µl of 3 M NaOAc (pH 5.27) and 300 µl of ethanol (pre-cooled at -20°C) were added, the mixture was incubated on dry ice for 20 min and centrifuged for 20 min at 14,000 rpm at 4°C. The supernatant was carefully removed, then 75% ethanol (pre-cooled at -20°C) was added, and the mixture was centrifuged for 5 min under the same conditions. The supernatant was carefully removed, and the pellets were dried on SpeedVac for 10 min and dissolved in 8 µl of the formamide loading solution (94% formamide, 2 mM EDTA, 0.05% bromophenol blue, 0.05 % xylene cyanol). Then, 2 µl of the obtained solution were mixed with 2 µl of the formamide loading solution, heated at 85°C for 3 min, quickly chilled on ice or a pre-cooled rack and loaded on 5% sequencing gel (acrylamide:bis-acrylamide 29:1) containing 8 M urea and run at 70V/cm. As size markers, a 5’-end labeled denatured DNA ladder consisting of DNA fragments of sizes increasing in steps of 100 bases or 10 bases was used. Then, the gel was dried and exposed to a phosphoimager screen.

Quantitation of the data - Intensities of the bands were quantitated by phosphoimaging using ImageQuant software. We measured the volume of the signal from the rectangle drawn closely around a band of interest. The local background was calculated by placing same size adjacent rectangles above and below the one surrounding the band and averaging the signals from these adjacent rectangles. This amount was subtracted from the signal obtained for the band of interest. In the case of WT-C and WT-Rev sequences, the arrest site bands were poorly distinguishable from background and their positions often had to be extrapolated from neighboring samples. These values were normalized to the signal from the major transcription product (Fig.3). Because the transcription was performed on a circular DNA template, the major product was expected to be heterogeneous. However, it formed a relatively narrow band instead of either smearing or trapping in wells. Thus, most of the transcripts from the circular plasmid were sufficiently long to be beyond the resolution capability of the gel, but not too long to prevent their entering the gel. Estimation from the ribozyme-containing plasmid (see Supplemental Materials) suggests that the average size of these products is roughly twice the size of the plasmid.

RESULTS

A sequence from the nuclease-sensitive
element of the c-MYC promoter causes partial transcription arrest. We found that during T7 RNA polymerase transcription through the sequence from the nuclease-sensitive element of c-MYC promoter (Fig.1A; this sequence is further referred as “wild type” or WT sequence) truncated transcription products appeared indicating partial transcription arrests within and closely downstream of the sequence (the left panel in Fig.3, lane “WT”). These products were barely detectable when the bases within the sequence were replaced by complementary bases (lane “WT-C”), or when the sequence was cloned in the opposite orientation (lane “WT-Rev”). Thus, the inhibitory effect on transcription is sequence-dependant. The main purpose of our study is to determine whether this effect on transcription is due to formation of one or another of the non-canonical DNA structures within this sequence. To study this, we constructed sequences that differed in significant ways relevant to formation of particular secondary DNA structures and tested them for effects on transcription arrest.

**Design of test sequences.** Under physiological pH and ionic conditions, this sequence could form two non-canonical DNA structures, the intramolecular purine-purine-pyrimidine (PuPuPy) DNA triplex (H-DNA) and the DNA quadruplex containing a G-quartet (see Fig.1). The ability to form both these structures is a general feature of sequences containing sufficiently long blocks of guanines, because both of these structures are stabilized by Hoogsteen base pairing between guanines. Additionally, the structure that is more energetically favorable in free DNA is not necessarily the one which is actually formed when DNA is partially bound to protein or to another nucleic acid, as in the case of transcription. Thus, *a priori* it is difficult to say which of these structures (if any) is responsible for the effect of interest (i.e. interference with transcription in our case). However, by making sequence substitutions which are known to selectively destabilize either the triplex or the quadruplex, and by monitoring the effects of these substitutions on the interference with transcription one might be able to discern the contribution of each of these structures to the interference observed.

One of the base substitutions is based on the fact that in the triplex, adenines participate in Hoogsteen base pairing (Fig.1B), while in the G-quartet-stabilized quadruplex only guanines participate in Hoogsteen base pairing, and other bases are presumably “looped out” from the structure (Fig.1C). Consequently, the stability of the G-quadruplexes would depend mostly on the length of uninterrupted blocks of guanines within the sequence and is not expected to be sensitive to nucleotide substitutions between the blocks. Thus, substitution of both adenines by thymines in homopurine blocks within the wild-type (WT) sequence, producing sequence designated TT (this and all other sequences are shown in Fig.2), should selectively destabilize the triplex, but not affect the quadruplex. We also used the *Theromyces* telomeric sequence motif \((G_4T_2)_n\) (sequence Q), which is known to form a stable quadruplex (24). The quadruplex formed by this sequence is expected to be significantly more stable than the potential quadruplexes formed by other sequences tested in this work, because it has four nucleotide-long uninterrupted G-blocks, while the others have only three nucleotide-long or less uninterrupted G-blocks (see Fig.2). Thus this sequence should have the strongest effect on transcription if the quadruplex formation is responsible for this effect. In contrast, the triplex formed by this sequence (25) is expected to be less stable than that formed by the WT sequence due to the presence of thymines, which interrupt the triplex.

The design of two other sequences, designated AG and AA, was based on the fact that the formation of an undisturbed
intramolecular triplex within homopurine-homopyrimidine sequence requires mirror symmetry of the sequence (see Fig.1), and deviation from the mirror symmetry significantly destabilizes this structure (26). The sequence AG contains one A-G permutation in the right half of the wild-type c-MYC sequence which distorts the mirror symmetry of two homopurine-homopyrimidine blocks within the sequence, and the sequence AA contains an additional “compensating” A-G permutation in the left half of the sequence, which restores their mirror symmetry. If the interference with transcription is due to intramolecular triplex formation, then a single A-G permutation in the sequence AG should reduce the effect on transcription in comparison to that with the wild-type (WT) sequence; and the compensating second permutation in the sequence AA would restore the effect on transcription approximately to the same level as for the wild-type sequence. In contrast, the potential quadruplex in both of these sequences should be destabilized in comparison with that in the WT sequence because the A-G permutations disrupt one of four three nucleotide-long G-blocks within the sequence.

In summary, the predicted ranking of the inhibitory effects on transcription for the various test sequences is

Q > (WT, TT) > (AA, AG)

if the effect on transcription is due to quadruplex formation,

and

(WT, AA ) > (AG, TT, Q)

if the effect on transcription is due to triplex formation.

*Intensities of arrest sites downstream of the c-MYC sequence depend upon the sequence orientation and correlate with its triplex-forming potential.* The typical results of experiments are shown in Fig.3. We found two bands corresponding to partial arrest sites localized in close proximity to the downstream flank of the c-MYC sequence for which the intensity correlates with the sequence orientation and composition. The bands were clearly pronounced only when the Pu-rich transcript was formed (i.e. for all sequences except WT-Rev and WT-C). Judging from the DNA ladder marker, a stronger (top) band is localized about 10 bases downstream from the c-MYC sequence, and a weaker (bottom) band is localized within the c-MYC sequence close to its downstream flank. The data for the various sequences are summarized in Fig. 4. As described in the previous subsection, for sequences producing a Pu-rich transcript, an expected ranking of the triplex-forming potential is WT, AA > AG, TT, Q. An expected ranking of quadruplex-forming potential is Q > WT, TT > AA, AG. From Fig.4 it is seen that the ranking of observed band intensities correlates with the triplex-forming potential, rather than the quadruplex-forming potential.

All the experiments described above were performed with circular closed negatively supercoiled DNA as a template for transcription (see next section). Consequently, the band at the top of the gel representing a major (“complete”) transcript might correspond to several rounds of RNA polymerase transcribing around the circular template. The intensity of this major band was used as a normalizing factor to obtain the relative values of the arrested band intensities, which reflect relative probabilities for transcription arrest (see Experimental Procedures). To estimate an absolute value for these probabilities, we applied a ribozyme-based approach similar to one used in (27) (see Supplemental Materials). Using this method, we estimated that for the WT sequence the
probability of arrest at the principal top arrest band position was ~1%.

Intensity of arrest sites increases with negative supercoiling. The absence of an observable effect in linear DNA (data not shown) suggests that negative supercoiling is required for arrest in the vicinity of the sequence of interest, and thus the signal should increase with increased negative supercoiling, consistent with the structure formation induced by negative supercoiling. To test this hypothesis we prepared topoisomer distributions with increasing superhelical densities (Fig. 5A). From Fig. 5B it is seen that the intensities of the arrest bands of interest do indeed increase with increasing negative supercoiling. In most of our experiments, we used DNA of native superhelical density, which is usually distributed around -0.05 and should be similar for the same plasmid vectors propagated in the same bacterial strain under the same conditions. However, various plasmid samples might still differ slightly in topoisomer distribution (especially in the highly supercoiled "tails" of the distribution). In addition, they also might contain various amounts of some minor fractions like partially denatured DNA which might be prone to fold into unusual structures and thus strongly contribute to the observed effects. To exclude any artifacts related to these possible differences in the plasmid samples, we gel-purified the topoisomers from natively supercoiled pWT, pWT-Rev, pAA, and pAG plasmids from the area shown by an arrow in the bottom panel of Fig.5A. The purification produced a narrow distribution of topoisomers, which was virtually identical for pAA, pAG and pWT-Rev, and very slightly shifted (less than one topoisomer) for pWT (Fig.6A). The results for the purified topoisomers (Fig.6B and C) were similar to the results for the initial plasmid samples (Fig.4), confirming that the effects observed really result from differences in the nucleotide sequences, and not from possible differences in the plasmid samples.

Predominant arrest sites are localized downstream of the c-MYC element sequence and their positions depend upon the flanking sequences. As already mentioned, according to the DNA size ladder the band which corresponds to the predominant arrest site is localized about 10 bp downstream from the c-MYC sequence. The comparison with RNA markers produced by run-off transcription from the plasmid linearized by restriction enzymes confirmed this conclusion: accordingly, the most intense top arrest band is localized between eight and eleven bases downstream from the c-MYC sequence (Fig.7A). To check if the position of the arrest band depends upon the flanking sequence, we prepared analogs of the plasmids pWT and pWT-Rev, pWT-D223 and pWT-D223-Rev, in which 223 bp 2 bases immediately downstream from the c-MYC sequence were deleted, leaving a flanking sequence without an apparent similarity to the flanking sequence of the parental plasmids (Fig.2B). Fig.7B shows that the orientation-dependence of arrest intensities in the downstream vicinity of the c-MYC sequence for these two plasmids is similar to that for the parental plasmid; they are pronounced for pWT-D223, but not for pWT-D223-Rev). However, in contrast to pWT, in the case of pWT-D223 there is no single predominant band located 10 bp downstream from the c-MYC sequence. Rather, there are several bands of somewhat lower intensities. Thus, though the orientation of the c-MYC sequence affects the intensity of downstream arrest sites, it does not define their exact positions. As expected, the position of the less intense bottom arrested band, which is within the c-MYC sequence, was the same for both pWT and pWT-D223. The sequence determinants and the one-nucleotide resolution for the top arrest band location in pWT plasmid require further investigation. However, the band position is
not determined by the 6 bp AT-rich block 11 bp downstream from the c-MYC sequence, because its deletion in the pWT-D6 plasmid (Fig.2A) doesn't affect the band location (Fig.7C). In contrast, deletion of 11 bp immediately downstream of the c-MYC sequence (WT-D11, Fig.2A) removes the band, leaving instead a diffuse weakly pronounced arrest area (Fig.7C). Again, the position of the bottom arrest band is not affected by these deletions. Interestingly, in Fig.7C extra bands further downstream from the c-MYC sequence became more pronounced upon moving closer to the c-MYC sequence due to deletions. This is consistent with our proposed model for non-specific exacerbation of weak pause sites downstream from H-DNA (see Discussion).

**DISCUSSION**

The character of the dependence of the effect on transcription on various sequence modifications suggests that this effect is due to the triplex formation. In this work we have studied partial transcription arrest sites in the vicinity of the downstream flank of a homopurine-homopyrimidine sequence from the human c-MYC promoter. There are several possible mechanisms for how such a sequence might interfere with transcription: 1) DNA triplex formation, which is known to interfere with transcription (reviewed in (28)); 2) transcription induced formation of G-quartets in the non-template DNA strand (2,29), which could affect DNA interaction with RNA-polymerase; 3) formation of G-quartets in the nascent RNA, which was shown to stimulate transcription termination (30); 4) formation of unusually stable R-loops within Py-rich DNA template (31,32) which could interfere with separation of the template DNA strand and the nascent transcript, thus facilitating RNA polymerase pausing (33); and 5) some yet unknown pausing or termination signal, which might interact with RNA polymerase in a sequence-specific manner (34).

To distinguish between these possibilities, we have introduced various substitutions in the wild type c-MYC sequence that would selectively destabilize either triplex or quadruplex formation. The ranking of their observed effects on transcription was consistent with the formation of triplex DNA being responsible these effects (Fig.4). Especially informative was the comparison between the AA and the AG sequences (Fig.2A). They have the same base composition and are expected to have the same quadruplex-forming potential. However, in contrast to the AG sequence, the AA sequence has a symmetry that is required for the formation of the intramolecular triplex (26). Thus, the fact that arrest sites are stronger for the AA sequence than for the AG sequence (Figs.4 and 6) strongly supports the involvement of the triplex DNA in the observed effect on transcription. The enhancement of the arrest site intensities with negative supercoiling is also consistent with this hypothesis, because H-DNA is stabilized by negative superhelical stress.

Interestingly, we found that although the intensity of the most pronounced arrest band clearly depends on the orientation and modifications of the c-MYC sequence, the band itself is localized downstream from the sequence, and its exact position varies with the flanking sequence.
Possible mechanism of interference with transcription and its potential biological implications. There are at least two possible mechanisms to explain how H-DNA could interfere with transcription. First, H-DNA could form spontaneously under negative superhelical stress (reviewed in(19)). Transcription elongation is known to be arrested by intermolecular triplexes formed by triplex-forming oligonucleotides (35,36); thus, it also could be arrested by the stable triplex within the H-DNA. However, this mechanism implies that the arrest should occur in the upstream flank of the triplex-forming sequence, or possibly within the sequence, if the single-stranded region of H-DNA serves as a template, but not downstream of the sequence, as is the case in our studies. Thus, even if H-DNA forms spontaneously under our conditions, it is not sufficiently stable to detectably interfere with transcription.

The second mechanism is the inhibition by H-DNA, whose formation is induced by transcription itself, for example by the mechanism suggested by Grabczyk and co-workers (27,37,38): transcription provides DNA opening and "donates" an unpaired non-template strand as a third strand for the triplex formation upstream of the transcription bubble (Fig.8). This process could be facilitated by the transient increase in negative superhelical density immediately upstream of an elongating RNA polymerase (39), and the emerging H-DNA could be further stabilized by base-pairing between the template strand and the nascent transcript (37,40,41). Grabczyk and co-workers (27,37,38) suggest a mechanism to explain how transcription-generated H-DNA might trap RNA polymerase within an H-DNA-forming sequence close to its downstream junction. We suggest that this effect could propagate further downstream of the H-DNA-forming sequence. We propose that the sharp DNA bend produced by H-DNA creates an obstacle for elongating RNA polymerase not only within the H-DNA-forming sequence (which in our case can explain the lower arrest band), but also after it passes the H-DNA-forming sequence, by sterically constraining the mutual rotation of DNA and RNA polymerase during transcription (Fig.8C). In principle, the RNA transcript can maintain its binding with the template strand during transcription producing a long R-loop (29,37,42), or it could first dissociate from the template, and then re-hybridize with the unpaired strand of H-DNA. In the latter case, the "anchoring" of the transcript will cause an entropically unfavorable RNA wrapping around the DNA duplex during transcription (Fig. 8D), which could also impede transcription. In addition, a sharp bend created by H-DNA could impede the transport of transcription-induced superhelical stress along the plasmid (43). We suggest that these steric constraints exacerbate some "hidden" pausing and termination signals in the flanking sequences, which normally are too weak to be detected. This could explain why the exact position and intensity of arrest sites downstream from the H-DNA-forming sequence varies with downstream flanking sequences. The predominant arrest sites downstream from the H-DNA-forming sequence rather than inside this sequence may be a function of the length of the H-palindromes. For longer sequences, other patterns of transcription interference have been observed (for example, see (27,44-46)).

Such a mode of interference of transcription by DNA structure might be relevant to the sequence-context dependence of the biological effect(s) of the non-B structures within the genome: some DNA regions that are able to form the H-DNA structure might have a weak effect on transcription within one sequence context, but could have a more significant effect within another; for example, if they are localized closely upstream to a pre-existing sequence, with an increased probability of transcription pausing or termination. In this context it
would be interesting to investigate the effect of H-DNA on known pausing or termination signals. Also, it would be interesting to determine if the arrest of transcription detected in this work is followed by RNA polymerase dissociation, or whether RNA synthesis can resume after interruption without dissociation.

In the context of the potential biological role, it would be interesting to determine whether the effects of H-DNA-forming sequences are similar for eukaryotic RNA polymerases. However, since our suggested mechanism is based upon steric effects rather than some specific interactions between RNA polymerase and DNA, we suggest that it might be general for various RNA polymerases, and could be even stronger for larger enzymes. A next step will also be to address the question of whether TCR is elicited near the arrest sites.

REFERENCES

1. Aguilera, A. (2002) *Embo J* 21, 195-201
2. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004) *Genes Dev* 18, 1618-1629
3. Duquette, M. L., Pham, P., Goodman, M. F., and Maizels, N. (2005) *Oncogene* 24, 5791-5798
4. Mellon, I., Spivak, G., and Hanawalt, P. C. (1987) *Cell* 51, 241-249
5. Mellon, I., and Hanawalt, P. C. (1989) *Nature* 342, 95-98
6. Hanawalt, P. C. (1994) *Science* 266, 1957-1958
7. Hanawalt, P. C. (2002) *Oncogene* 21, 8949-8956
8. Saxowsky, T. T., and Doetsch, P. W. (2006) *Chem Rev* 106, 474-488
9. Laine, J. P., and Egly, J. M. (2006) *Trends Genet* 22, 430-436
10. Wang, G., Seidman, M. M., and Glazer, P. M. (1996) *Science* 271, 802-805
11. Wang, G., and Vasquez, K. M. (2006) *Mutat Res* 598, 103-119
12. Mirkin, S. M. (2006) *Curr Opin Struct Biol* 16, 351-358
13. Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E., and Wells, R. D. (2006) *DNA Repair (Amst)* 5, 1161-1170
14. Firulli, A. B., Maibenco, D. C., and Kinniburgh, A. J. (1992) *Biochem Biophys Res Commun* 185, 264-270
15. Wang, G., and Vasquez, K. M. (2004) *Proc Natl Acad Sci U S A* 101, 13448-13453
16. Marcu, K. B., Bossone, S. A., and Patel, A. J. (1992) *Annu Rev Biochem* 61, 809-860
17. Eick, D., Polack, A., Kofler, E., Lenoir, G. M., Rickinson, A. B., and Bornkamm, G. W. (1990) *Oncogene* 5, 1397-1402
18. Hoover, R. G., Kaushal, V., Lary, C., Travis, P., and Sneed, T. (1995) *Curr Top Microbiol Immunol* 194, 257-264
19. Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994) *Annu Rev Biophys Biomol Struct* 23, 541-576
20. Davis, J. T. (2004) *Angew Chem Int Ed Engl* 43, 668-698
21. Simonsson, T., Pecinka, P., and Kubista, M. (1998) *Nucleic Acids Res* 26, 1167-1172
22. Yang, D., and Hurley, L. H. (2006) *Nucleosides Nucleotides Nucleic Acids* 25, 951-968
23. Tornaletti, S., Patrick, S. M., Turchi, J. J., and Hanawalt, P. C. (2003) *J Biol Chem* 278, 35791-35797
24. Sundquist, W. I., and Klug, A. (1989) *Nature* 342, 825-829
25. Veselkov, A. G., Malkov, V. A., Frank-Kamenetskii, M. D., and Dobrynin, V. N. (1993) *Nature* 364, 496
26. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987) *Nature* 330, 495-497
We thank Ed Grabczyk and Jennifer VanOverbeke for help with ribozyme constructions, Brian Johnston, Sergei Kazakov and members of Hanawalt’s laboratory for suggestions, and Ann Ganesan and C. Allen Smith for the critical reading of the manuscript. This work was supported by grants CA77712 to P. C. H. and CA93729 to K. M. V. from the National Cancer Institute, and by an Undergraduate Research Program Grant to E. de S.

**FIGURE LEGENDS**

**Fig. 1.** Possible DNA structures formed by the S1-sensitive element of the human c-MYC promoter. A. B-DNA duplex. Mirror symmetric purine (Pu) and pyrimidine (Py) blocks are shown in green and blue, respectively. B. One of the possible H-DNA structures formed by this sequence containing a PuPuPy triplex which is formed under physiological ionic conditions. C.
One of the possible quadruplex structures formed by the Pu-rich strand of the sequence. Dashed lines represent hydrogen bonds between guanines participating in G-quartet formation.

**Fig. 2.** Sequences used in this study. Only the non-template strand is shown (5'-3'). The human *c-MYC* sequence and its various substitutions are shown in capital letters, and the first 20 bases of the downstream sequence are shown in lower case letters. Regions which can participate in H-DNA formation are underlined, and regions which can participate in the quadruplex formation are shown in bold. Arched arrows indicate permutations, the outlined T in the TT sequence shows an A-to-T substitution. In WT-D11 and WT-D6, the regions marked by dotted lines (11 bp long and 6 bp long, respectively) from the downstream flank of the WT sequence were deleted. In sequences shown in B, 223 bp were deleted from the WT (WT-D223) and WT-Rev (WT-D223-Rev) plasmids 2 bases immediately downstream of the *c-MYC* sequences. For flanking regions of WT and WT-D223, the ends of run-off products obtained by transcription from the plasmids cut by *Bgl* II and *Hind* III restriction enzymes, respectively are shown by the vertical arrow.

**Fig. 3.** Truncated transcription products (arrest bands) in the downstream vicinity of the S1-sensitive element of the human *c-MYC* promoter. Arrest bands in the vicinity of the sequence of interest are shown by block arrows. At the top of the gel, an underexposed fragment of the gel with the major transcription product is shown. Lanes designated 100 and 10 show denatured DNA ladders with steps 100 bases or 10 bases, respectively.

**Fig. 4.** Relative intensities of arrest bands for various sequences. All results are normalized on wild-type (WT) sequence.

**Fig. 5.** Arrest bands intensities increase with negative supercoiling. *A.* Analysis of plasmid DNA samples with various supercoiling levels. Samples 1, 2, 3, 4, 5, 6 with increasing negative supercoiling were obtained by relaxing of plasmid DNA by WG Topo I in the presence of 0, 0.5, 1, 2, 3, 4 µg/ml of ethidium bromide, respectively. “N” designates DNA with native supercoiling untreated by topoisomerase. Concentrations of chloroquine during gel electrophoresis were, from top panel to bottom, 0, 3, 10, and 24 µg/ml. In the lowest panel, the arrow indicates the area from which topoisomers were cut (see below Fig. 6). *B.* Relative intensities of the arrest bands for the samples 1-6 from *A*, normalized to the intensity of the corresponding band for natively supercoiled (N) DNA.

**Fig. 6.** Relative intensities of arrest bands for various sequences for gel-purified topoisomers are similar to relative intensities of natively supercoiled DNA. *A.* Topoisomer analysis on an 1% agarose gel with 24 µg/ml chloroquine. SC designates supercoiled topoisomers, which are similar for all plasmids, OC is a small amount of open circular (nicked) DNA which most likely formed during gel-purification. *B* and *C.* Results of transcription experiment and quantitation for gel-purified samples from *A* (Bottom band for WT-Rev were indistinguishable from the background). It is seen that the results are similar to the results obtained for native DNA (Fig. 3 and Fig. 4).
Fig. 7. Most pronounced arrest band(s) are localized downstream the *c-MYC* element and their positions depend on flanking sequences. *A*. The numbers on the gel designate the sizes of corresponding RNAs and DNAs. RNA markers 282 and 285 are obtained by transcription from pWT digested by *Bgl*II and pWT-D223 digested by *Hind*III, respectively (see Fig.2 for the sites positions). It is seen, that for WT sequence, a predominant arrest band (a larger block arrow), is localized in between 8 and 11 bases downstream of the *c-MYC* sequence. The weaker bottom arrest band (a smaller block arrow) is within the insert. *B* and *C*. Altering the sequence immediately downstream the *c-MYC* sequence (WT-D223 and WT-11) results in a loss of the sharp top band, and an occurrence of several weaker bands or a smeary band, while deletions of 6 bp at positions 12-17 bp downstream the *c-MYC* sequence does not disrupt the band pattern. In *C*, it is also seen that other bands downstream of the sequence (double-headed block arrows) became more pronounced upon "moving" closer to the *c-MYC* sequence due to deletions. In all cases, a weaker bottom arrest band remains intact, confirming its localization within the insert.

Fig. 8. Model for creation of steric constrains for transcription downstream of the H-DNA-forming sequence. Homopurine and homopyrimidine regions within the H-DNA-forming sequence are shown in green and blue, respectively. *A*. RNA polymerase is moving along the DNA template, generating extra-negative supercoiling in the DNA behind and extra-positive supercoiling in the DNA in front. *B*. When RNA polymerase passes the first half and middle of the H-palindrome (or probably further), stabilization of the DNA open formation by the transcript together with extra-negative supercoiling facilitate H-DNA formation immediately behind RNA polymerase. During further transcription, a nascent RNA could either continuously hybridize to the template forming an extended R-loop (not shown), or either one or both Pu-blocks within the transcript could re-hybridize (shown by dashed arrows in *C*) to the single-stranded Py-block of the H-DNA, forming either duplex or triplex. *C*. Because of helical structure of the DNA duplex, RNA polymerase rotates with respect to the DNA duplex during transcription (rotation is shown by round block arrow). Due to a sharp DNA bend created by H-DNA formation, RNA polymerase "clashes" with DNA duplex upstream of the H-DNA while rotating around the DNA duplex downstream of the H-DNA. Note that here we imply a relative rotation of RNA polymerase in the frame of reference related to DNA. In reality, both RNA polymerase and DNA are rotating in space in opposite directions, and the absolute rates of their rotation depends on their frictional constants. *D*. If Pu-blocks within the RNA re-hybridize with the single-stranded Py-block in H-DNA, such an RNA "anchoring" will cause wrapping of the nascent RNA around the DNA duplex during RNA synthesis.
Fig. 1

A

5′-GGGAGGGGCCTTATGGGGAGGG-3′
3′-CCCTCCCCCGCGAATACCCTCTCCC-5′

B

5′-CCCTCCCC-3′
3′-GGGAGGGGCATAT-5′

C

A-G-A-G

G-G-G-G

G-G-G-G

G-G-G-G

5′-G-G-G-G-3′

C-T-T
A

WT: GGGAGGGGCGCTTATGGGAGGGrctagatctcgtattaagct

AG: GGGAGGGGCGCTTATGGGAGGGrctagatctcgtattaagct

AA: GGGGGAGGCGCTTATGGGAGGGrctagatctcgtattaagct

TT: GGGGGGGGCCTATGGGGGAGGrrctagatctcgtattaagct

Q: TGGGGTTGGGTTGGGTTGGGrrctagatctcgtattaagct

WT-Rev: CCCTCCCCATAAGCGCCCCTCCCrctagatctcgtattaagct

WT-C: CCCTCCCCGCGAATACCCCTCCCrctagatctcgtattaagct

del. in WT-D6

del. in WT-D11

B

WT-D223: GGGAGGGGCGCTTATGGGAGGGrctagcaagcttggcactgg

WT-D223-Rev: CCCTCCCCATAAGCGCCCCTCCCrctagcaagcttggcactgg
Fig. 3
Fig. 4
negative supercoiling chloroquine concentration during gel electrophoresis

**Fig. 5**

A

negative supercoiling

chloroquine concentration during electrophoresis

B

top band

bottom band
Fig. 6

A

B

C

Top band

Bottom band
Fig. 7

A

WT RNA markers DNA 10 b ladder

WT-D223

WT-D223-Rev

WT-Rev

WT-D11

WT-D6

B

C

c-MYC (252-274)
Fig. 8

A

Extra-negative supercoiling

Extra-positive supercoiling

B

D

C
SUPPLEMENTAL DATA

Supplemental Fig. 1. Analysis of the transcription products using ribozyme-containing plasmids. It these plasmids, the transcript cleaves itself with high efficiency at a defined position due to the presence of a self-cleaving ribozyme sequence within the transcript. The ribozyme-based approach is designed to overcome the difficulty in calculating the absolute probability of the transcription termination for a circular DNA template due to absence of the defined run-off length for the transcript (Grabczyk and Usdin (2000), Nucleic Acid Res, 28, 2815-2822). A. The scheme of the plasmids pWT-ribo and pWT-Rev-ribo: Points a, s, b, c correspond to position of the T7 promoter, dominant top arrest band downstream of the insert, ribozyme cleavage site, and restriction site used to obtain a linear substrate, respectively. The numbers indicate distances from the T7 promoter. The total length of the plasmid is 3.7 kb. B. Results of transcription from supercoiled (SC) and linear (L) pWT-ribo and pWT-Rev-ribo substrates. "M" designates a "major" transcription product, which in the case of ribozyme-containing plasmids comprises mostly plasmid-size transcripts (see below). From the results obtained from linear (L) substrates, the percentage of the ribozyme cleavage (f) can be estimated as

\[ f = \frac{I_{ab} + I_{bc}}{I_{ab} + I_{bc} + I_{ac}} \]

where \( I \) designates intensities of corresponding bands. We obtained \( f \) value around 80%, i.e. most of the substrates are cleaved. Thus, for our rough estimations below we will assume 100% cleavage. Because the circular plasmid does not have any strong termination site for T7 RNA polymerase, the transcription from the circular substrate likely proceeds, on average, for several cycles around the plasmid before spontaneous termination. Only the first cycle of transcription produces, after ribozyme cleavage, the product ab. The following cycles of transcription will produce multimer transcription products converted to the plasmid-length monomers (except the very last cycle before termination) upon ribozyme cleavage. Thus, the average number of transcription cycles, \( n \), around the plasmid before spontaneous termination could be estimated from the equation

\[ n = \frac{L_{ab} I_M}{L_M I_{ab}} \]

where \( L \) designates length of corresponding products, and M is the "major" transcription product. We obtained \( n \) value around 2. If we assume, that the arrest bands which we observe in the vicinity of the c-MYC sequence correspond to transcription termination, we can estimate probability \( r \) of this termination as

\[ r = \frac{L_{ab} I_{as}}{L_{as} I_{ab}} \]

For the top arrest band we estimated this probability around 1%.
Supplemental Fig. 1

A

- T7
- 0.56 kb
- Ribozyme cleavage site
- c-MYC
- top stop band (0.28 kb)
- 3.7 kb
- Dra III

B

Template:

| SC | L |
|----|---|
| WT | WT-Rev |
| WT-Rev | |

- M
- ab
- ac
- bc
- as