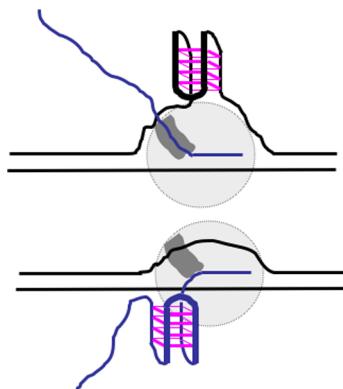


## DNA Sequences That Interfere with Transcription: Implications for Genome Function and Stability

Boris P. Belotserkovskii,<sup>†</sup> Sergei M. Mirkin,<sup>‡</sup> and Philip C. Hanawalt\*,<sup>†</sup>

<sup>†</sup>Department of Biology, Stanford University, Stanford, California 94305, United States

<sup>‡</sup>Department of Biology, Tufts University, Medford, Massachusetts 02155, United States



### CONTENTS

1. Introduction	8620
2. Transcription Blockage Mechanisms	8621
3. Effects of Unusual DNA Structures and DNA/RNA Complexes on Transcription	8622
3.1. General Introduction to DNA Structure and DNA Supercoiling	8622
3.2. Overview of Unusual Structures and Their Effects on Transcription	8623
3.2.1. Branched/Looped Structures	8623
3.2.2. Z-DNA	8625
3.2.3. Triplex Structures	8627
3.2.4. Guanine Quadruplexes	8628
3.2.5. R-Loops	8630
4. More Complex Phenomena Involving Anomalous Transcription	8631
4.1. Transcription–Replication Collisions	8631
4.2. Transcription-Coupled Repair	8631
4.3. Interdependence between Transcription and Unusual DNA Structures Mediated by Chromatin Remodeling	8632
4.4. Nascent RNA Anchoring to DNA and Its Implications	8632
Appendix	8633
R-Loop Stability against Positive Supercoiling	8633
Author Information	8633
Corresponding Author	8633
Funding	8633
Notes	8633
Biographies	8633
Acknowledgments	8634
References	8634

### 1. INTRODUCTION

The primary role of DNA-dependent RNA synthesis, or transcription, is to create components for the cellular machinery. The nascent RNA product of transcription is released from the DNA template and either serves as an intermediate message for protein synthesis or is used directly, as in the case of rRNA, tRNA, and various types of regulatory RNAs. Recently, a growing number of examples suggest that transcription per se, rather than its released product, could play a regulatory role in gene function or as a trigger for genomic modifications.

The latter scenario is commonly attributed to anomalous progression of the RNA polymerase (RNAP), such as pausing or termination and/or retaining rather than releasing the nascent transcript. This has been implicated in class-switch recombination and somatic hypermutation (reviewed in refs 1, 2), telomere maintenance,<sup>3</sup> and replication initiation.<sup>4–9</sup> Anomalous transcription elongation has also been linked to various deleterious phenomena, such as genomic instabilities, transcription–replication collisions, and transcription deficiency in some hereditary human disorders (reviewed in refs 10–14).

Importantly, many of the DNA sequences at which RNAP stumbles are prone to form alternative (non-B form) DNA structures and/or stable RNA/DNA hybrids. Though some sequences that can form these unusual DNA structures are localized within genes (mostly in introns), the majority of them occur in genomic regions that were previously assumed to be nontranscribed.<sup>15</sup> Thus, the recent discovery of transcription in presumed transcriptionally silent regions, such as telomeres,<sup>3</sup> combined with the realization that most of the genome is transcribed (reviewed in ref 16), strongly implies that these sequences could be highly biologically relevant. We might expect to find many more examples in which unusual DNA structures would impact cellular processes as a result of their effects on transcription.

In this review, we discuss possible mechanisms of transcription through DNA sequences with unusual structural properties and their biological implications. We will consider primarily the elongation stage of transcription, since transcription initiation is generally sequence-specific and often involves many other proteins; thus, it is more difficult to interpret the effects in terms of the physical properties of the DNA template. [Note, however, that recent data show that

Special Issue: 2013 Gene Expression

Received: February 6, 2013

Published: August 23, 2013

alternative DNA structures could serve as promoter-like elements.<sup>17]</sup> We will only mention briefly, in the context of the topic, short sequence-specific termination signals and defects in the template strand, such as strand breaks and chemically modified bases.

We will begin by outlining the general possible mechanisms for transcription blockage or stalling. We will then describe various families of alternative DNA structures and their effects on transcription. Finally, we will consider more complex phenomena like transcription-coupled DNA repair, transcription–replication collisions, and protein-mediated interactions between nascent RNA and the DNA template.

This is a new and rapidly emerging field with many models that are still highly hypothetical. We believe that the models and ideas suggested for one class of transcription-related events might be of value for other classes. Thus, we will emphasize general mechanistic aspects of the problem, rather than focusing upon particular biological phenomena.

## 2. TRANSCRIPTION BLOCKAGE MECHANISMS

Before we proceed to review transcription blockage by various structures, it is useful to consider in general the features of DNA or nascent RNA that could cause transcription blockage. During transcription, RNAP translocates along the DNA template strand, synthesizing complementary RNA, and in this process, it must open up or unwind structures in which the template DNA strand is involved. Thus, the most obvious mechanisms would comprise obstacles for transcription localized in the template strand. They could be breaks, gaps, or chemically modified nucleotides, which for RNA polymerase are difficult to overcome (e.g., see refs 18, 19); alternatively, the template DNA strand could be chemically intact but involved in some extrastable structure formation, which for RNA polymerase may be difficult to unwind. These two straightforward mechanisms, in principle, are in common for RNA polymerases and DNA polymerases and probably for some other proteins translocating along the DNA template.

However, RNA polymerase is additionally interacting with the nontemplate DNA strand, nascent RNA, and an RNA/DNA hybrid within the transcription complex; unusual structural properties of any of these nucleic acid moieties could affect transcription and, in particular, facilitate transcription blockage. For example, an oligo-dA DNA template forms an especially weak duplex with the complementary oligo-U RNA, urging RNAP to spontaneously terminate transcription in stretches of oligo-dT/dA, when an especially weak rU/dA duplex is formed within the transcription complex.<sup>20</sup> While these stretches are very weak pausing/termination signals per se, they could be strongly exacerbated by structures formed in the nascent RNA or between RNA and DNA (e.g., see refs 21, 22). Interestingly, not only stability but also the shape of RNA/DNA hybrids within the transcription complex can modulate the propensity of RNA polymerase for pausing.<sup>23</sup>

In contrast to DNA polymerases, RNAP faces the unique challenge to render the nascent RNA available for further transactions, by separating the RNA from its DNA template. During normal transcription, a short (~8–10 bp) RNA/DNA duplex is formed within the transcription complex, and a special wedgelike protein moiety “peels” nascent RNA from the RNA/DNA duplex, followed by RNA extrusion through a positively charged RNA exiting-channel.<sup>24–26</sup> Binding of nascent RNA to the RNA exiting-channel is thought to strongly contribute to the stability of the transcription complex, and structures within

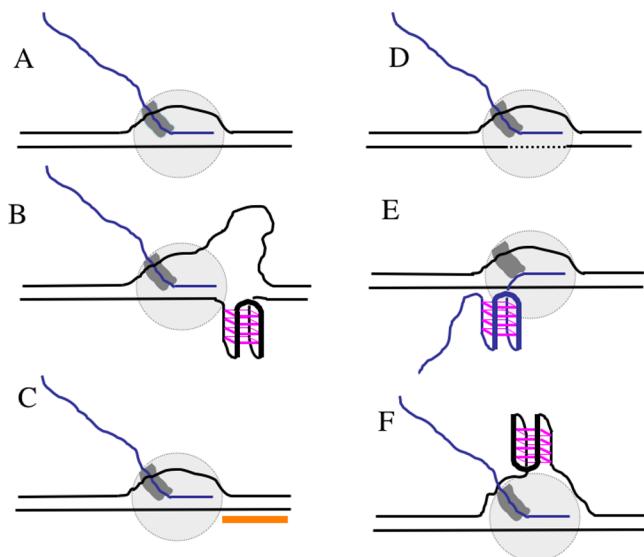
or involving nascent RNA would be expected to compete with this binding, thus destabilizing the transcription complex.<sup>27,28</sup> Alternatively, this secondary structure could “push” RNA polymerase forward without RNA synthesis, thus shortening the RNA/DNA hybrid within the transcription complex instead of disrupting the interaction between the nascent RNA and the exiting channel.<sup>29,30</sup> Interestingly, although this forward translocation mechanism mediated by the nascent RNA secondary structure formation has been suggested for transcription termination at intrinsic terminators,<sup>30</sup> in some cases nascent RNA secondary structures can have the opposite effect: they can inhibit RNAP pausing, thus moving transcription forward.<sup>31</sup> It is important to note that the sequence within the nascent RNA that forms a secondary structure interfering with transcription is not necessarily completely encoded by the DNA template; it could be partially or completely synthesized by reiterative transcription, in which nucleotides are repetitively added to the 3'-end of a nascent RNA due to slippage between the nascent RNA and the DNA template (e.g., refs 32–35). For example, slippage can produce quadruplex-forming oligo-G sequences in nascent RNA, which interfere with further transcription.<sup>36</sup>

Another mechanism of interference with transcription at the stage of RNA extrusion could be that some extra-stable RNA/DNA duplexes resist “peeling out” from the complex, thereby temporarily immobilizing the transcription machine.<sup>37</sup>

In contrast to DNA polymerase, RNAP normally maintains contact with the nontemplate strand during template copying. The nontemplate strand participates in proper RNA displacement and increases the processivity of transcription (see ref 27 and references therein). Thus, it is possible that sequestering of the nontemplate strand by an unusual structure formation could interfere with transcription.<sup>38</sup> On the other hand, transcription elongation can proceed in the absence of a complementary nontemplate strand, and at least in some systems, the lack of the nontemplate DNA strand has only a very weak effect for short stretches of nucleotides with a random sequence.<sup>39</sup> Furthermore, there are no clearly defined blockages at the sites of disruptions in the nontemplate strand, also suggesting that contiguous contact with the nontemplate strand is not an absolute requirement for transcription.<sup>40</sup> However, disruptions in the nontemplate strand might strongly exacerbate the blockages produced by other factors.<sup>40</sup> Another possible mechanism is that nontemplate-strand-mediated interactions could alter the conformation of the neighboring DNA, sterically sequestering the RNAP, thus interfering with transcription.<sup>41,42</sup>

Figure 1 provides examples of various mechanisms of transcription blockage by unusual DNA or RNA structures.

It is important to appreciate that although the mechanisms of interference with transcription described above are applicable to all RNAPs, the details of how a particular RNAP reacts to these impediments may vary, resulting either in reversible pausing or slowing down or in irreversible termination with or without dissociation of the RNAP from the DNA template or in backward sliding (backtracking) along the DNA template. Also, the effects observed for purified RNAPs may be quite different from those occurring in cells or extracts, in which additional proteins may be invoked to resolve particular challenges.



**Figure 1.** Possible mechanisms for sequence-specific transcription elongation blockage/impediment. (A) Normal transcription. DNA is shown in black and RNA in dark blue, and RNA polymerase (RNAP) is shown as a gray oval with a dotted border; an area within RNAP that interacts with nascent RNA is shown by a darker gray patch. (B, C) RNAP encounters a stable structure formed by or with participation of the template strand. For panel B, this structure is a G4-quadruplex stabilized by G-quartets (shown in magenta); for panel C, it is a triplex between the downstream duplex region and a triplex-forming oligo (shown in orange). (D) Defects or sequence features in the template strand (shown in dotted line) that renders it a poor substrate for transcription. (E) Structure formation with participation of the nascent RNA (in this case, G4-quadruplex), which disrupts or weakens an interaction between the nascent RNA and RNAP, thus destabilizing the transcription complex. (F) Structure within the nontemplate strand (in this case, G4-quadruplex) that could also destabilize the transcription complex.

### 3. EFFECTS OF UNUSUAL DNA STRUCTURES AND DNA/RNA COMPLEXES ON TRANSCRIPTION

#### 3.1. General Introduction to DNA Structure and DNA Supercoiling

The most common secondary structure adopted by DNA in living organisms is B-DNA. B-DNA is a right-handed double-helix formed by two complementary DNA strands in an antiparallel orientation, bound to each other through a special pattern of hydrogen bonding between their stacked bases, called Watson–Crick base pairing. B-DNA has no special sequence requirements (except for the complementarity of base pairing between the strands), and its structural parameters are relatively insensitive to the sequence. Under physiological conditions, B-DNA is the most energetically favorable DNA structure, and “energy input” is required to convert it to another structure. One of the most important sources of this energy input is the special type of DNA deformation called supercoiling. DNA is supercoiled *in vivo* as a consequence of the various DNA transactions performed by proteins, and it strongly impacts DNA functioning (reviewed in refs 43, 44). Although DNA supercoiling has been frequently reviewed (e.g., refs 45–47), it would be useful to briefly describe those aspects of this phenomenon that are pertinent for our topic.

In B-DNA, the stereochemical optimal number of base pairs per helical turn is around 10.5. Decreasing (overwinding) or increasing (underwinding) this number increases the free

energy of DNA in the similar way that twisting a rubber tube increases its elastic energy. If the ends of the DNA molecule become constrained, following the over- or underwinding, such that their relative rotation is forbidden (e.g., when each strand becomes covalently closed in circular duplex DNA), the DNA becomes trapped in the deformed over- or underwound state. These DNA molecules are called positively or negatively supercoiled, respectively, while the concomitant DNA deformation is called “superhelical stress”. Similar to a rubber tube under torsional stress, the DNA winding deformation is distributed between twisting around its axis and bending into a “figure-eight-like” or plectonemic structure; the latter is responsible for the term “supercoiled”. For purposes of our brief analysis, however, these effects on shape are not essential. What is important, however, is that although the “total winding” within supercoiled DNA cannot be changed without breaking one of the DNA strands (for rigorous analysis of supercoiling, see ref 45), it could be redistributed over the DNA molecule as a whole. For example, in negatively supercoiled DNA, one short segment could become more strongly (or even completely) unwound, and this local change would relax negative superhelical stress in the rest of the molecule. This is exactly how negative supercoiling facilitates formation of noncanonical DNA structures, most of which, as described below, are topologically equivalent to unwound DNA (i.e., can be formed from a completely unwound DNA region without rotating the flanking regions). For the same reason, negative supercoiling facilitates invasion of single-stranded DNA or RNA (D-loop or R-loop formation) into a DNA duplex. Notably, some DNA invading agents (for example, “peptide nucleic acid” (PNA), an artificial DNA mimic with peptide-like backbone) bind DNA so strongly that they can unwind DNA “against the flow” of superhelical stress, thereby generating positive superhelical stress in the rest of the DNA molecule.<sup>48,49</sup> For some sequences, RNA/DNA hybrids can also be strong enough to render an R-loop stable, even under positive superhelical stress (see the Appendix). In terms of unusual DNA structures that are overwound in comparison with B-DNA, and whose formation would be facilitated by positive supercoiling, we are aware of only one example of a so-called “Pauling-like structure” observed in single-molecule experiments, in which DNA has undergone strong stretching and positive twisting.<sup>50</sup>

In addition to the steady-state DNA supercoiling described above, there are also “dynamic” regions of increased negative and positive supercoiling that transiently appear behind and in front, respectively, of the transcribing RNAP (or other DNA tracking enzymes) due to the frictional resistance against the relative rotation of these molecular machines and DNA (see ref 51; reviewed in refs 43, 44, 52). In addition, we have proposed a hypothetical mechanism for generating transcription-dependent supercoiling driven by anchoring of the nascent RNA to the DNA template.<sup>53</sup>

Another form of dynamic DNA supercoiling was recently proposed to be driven by chromatin remodeling complexes in eukaryotes. Removal of the nucleosome ahead of the translocating RNAP is crucial for its progression. It is observed that SWI/SNF and RSC remodeling complexes induce the formation of negatively supercoiled DNA loops upon their active translocation.<sup>54–56</sup> It is notable that while in most prokaryotes negative supercoiling is created by specially dedicated enzymes, in eukaryotes supercoiling is more likely

to appear transiently, due to various DNA–protein transactions (reviewed in refs 43, 44).

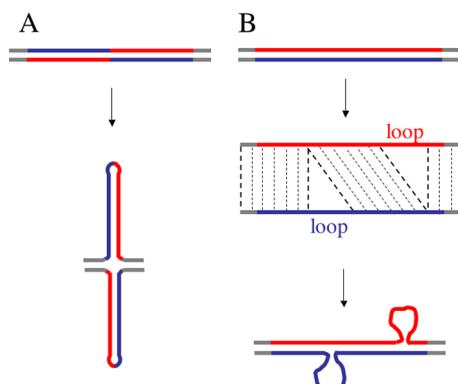
### 3.2. Overview of Unusual Structures and Their Effects on Transcription

In this section, we will discuss unusual DNA structures roughly in order of their increasing deviations from canonical B-DNA. We will start with branched/looped structures (e.g., cruciforms and slippage loops), which comprise regions in Watson–Crick B-form structures or in unstructured single-stranded DNA. Completely denatured DNA regions and partially mismatched Watson–Crick hairpins will be also considered under this category. The recently discovered PX-DNA structure<sup>57</sup> also belongs to this category, but as far as we know, its transcriptional properties have not been investigated. Then we will consider Z-DNA, which has Watson–Crick base pairing, but its left-handed conformation is strikingly different from that of B-DNA. Next, we will consider DNA tripleplexes, which comprise both Watson–Crick and non-Watson–Crick base pairing. Finally, we will consider G-quadruplexes, which comprise only non-Watson–Crick interactions. There are a number of other structures formed by non-Watson–Crick interactions, for example, those formed by partially protonated C-rich sequences, like cytosine–cytosine duplexes and quadruplexes (i-DNA),<sup>58,59</sup> and various purine–purine duplexes (e.g.,<sup>60,61</sup>). It is interesting to note that though i-DNA includes protonated cytosine and is most stable at a mildly acidic pH around 5, it is still detectable at a physiological pH around 7.<sup>62,63</sup> We are not aware of studies related to effects of these structures upon transcription elongation; thus, we do not consider them in detail in this review. However, it has been suggested that i-DNA could be involved in regulation of transcription initiation.<sup>64</sup>

The topic of RNA/DNA hybrids and R-loops will appear frequently in the context of various unusual structures, and in addition, we will devote a separate subsection to R-loops.

#### 3.2.1. Branched/Looped Structures. 3.2.1.1. Cruciforms.

Cruciforms (Figure 2A) can appear at inverted-repeat DNA sequences (palindromes), where two DNA strands contain self-complementary regions and, consequently, can fold back upon

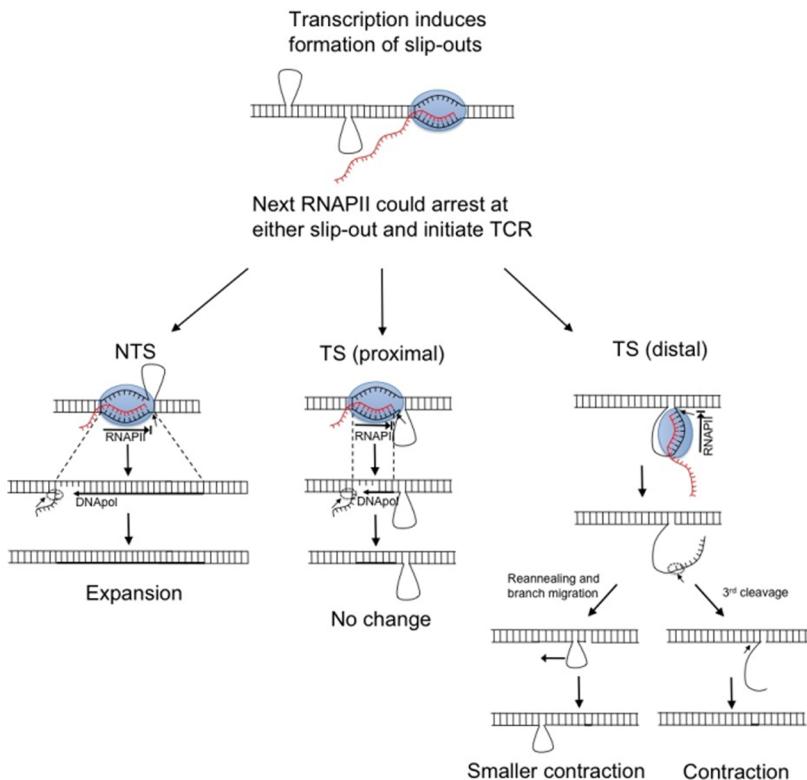


**Figure 2.** Examples of branched DNA structures stabilized by Watson–Crick interactions. Complementary regions within the sequence of interest are shown in blue and red. (A) Cruciform formation by complementary regions within each of the strands. (B) Slippage loops. After separation of complementary DNA strands containing repetitive sequences, some regions within these sequences could reanneal normally (shown by vertical dashed lines), and some regions could reanneal with a “shift” (shown by skewed dashed lines), which leads to slippage and loop formation.

themselves to form hairpins, instead of forming a regular duplex with each other (reviewed in ref 11). In nonsupercoiled DNA, cruciform structures are less energetically favorable than linear B-DNA, because they have a dramatic distortion of base-pairing at the base of their “stems” (called four-way junctions) plus several unpaired bases at the tip of each hairpin. In contrast, in negatively supercoiled DNA, a cruciform can out-compete linear B-DNA, being topologically equivalent to unwound DNA and, therefore, relaxing superhelical stress. The cruciform could also transiently appear in nonsupercoiled DNA, if the DNA strands were temporarily separated (for example, by denaturation *in vitro* or during replication or transcription *in vivo*) and then allowed to reanneal. In this scenario, the cruciform should eventually convert to duplex DNA as a result of the strand-exchange process termed branch migration. It is also worth noting that Holliday junctions, which appear *in vivo* as intermediates in homologous recombination and sometimes are generated at stalled replication forks (through fork regression), are structurally similar to cruciforms and thus might be expected to interact with transcription in a similar manner.

For model *in vitro* systems, stable supercoiling-independent cruciforms can be created by engineering the self-complementary regions of the strands to be noncomplementary to each other within the linear duplex. In this manner, it is also possible to generate the hairpin selectively on the template strand or only on the nontemplate strand to study their effects separately. In a purified transcription system with phage T7 RNAP, a stable cruciform or a hairpin on either strand does not create a detectable obstacle for transcription.<sup>65</sup> In HeLa extracts, however, each of these constructions produced partial blockage for RNAP II transcription, as well as some blockage for T7 RNAP (when T7 transcription is performed in the presence of HeLa extracts), suggesting that proteins bound to the branched structure are at least partially responsible for this effect.<sup>65</sup> Of course, RNA transcripts from the self-complementary DNA regions can also form hairpins. A hairpin in the nascent RNA can then disrupt the interaction between the RNA and the exiting channel of RNAP, thus destabilizing the transcription complex. This effect plays a key role in protein-independent termination signals (e.g., “type-one” signals for T7 RNAP) with a G-rich hairpin and a short dT/dA region immediately downstream from it, forming an unstable rU/dA RNA/DNA hybrid. While the inhibitory effect of each of these two sequences on transcription is almost unnoticeable, together they produce a strong termination signal.<sup>21</sup>

**3.2.1.2. Completely Denatured DNA Regions.** Under physiological conditions in nonsupercoiled double-stranded DNA, even the least stable A/T-rich sequences are in the duplex form. Strong negative supercoiling, however, can “melt” A/T-rich sequences, rendering them single-stranded. Long A/T-rich repeats are present in genomes and can operate as innate DNA unwinding elements in the initiation of DNA replication and other genetic processes (e.g., ref 66). A single-stranded DNA segment could also appear if the complementary strand is bound to proteins, RNA, or another DNA strand. In a model system, these regions could be mimicked by a noncomplementary “bulge”. The difference in behavior of transcription on the single-stranded versus the double-stranded template is defined by the role of the nontemplate strand in transcription. (Here we would remind the reader that in this review we are always considering transcription in the mature elongation mode; the role of the nontemplate strand in the initiation phase



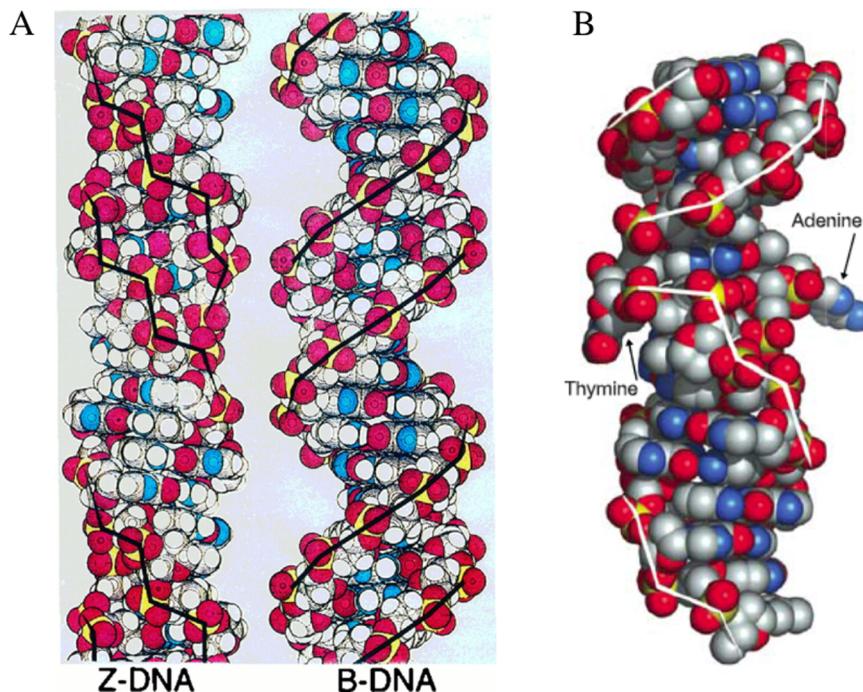
**Figure 3.** Model for transcription-dependent repeat instability mediated by slippage-loop formation. In genes involved in polyglutamine diseases, CTG repeats are present on the transcribed strand (TS) and CAG repeats are on the nontranscribed strand (NTS). After a translocating RNAP II has induced the formation of slipped-strand structures, the next RNAP II can arrest at either a CAG slip-out on the NTS or a CTG slip-out on the TS (at either the proximal or distal position). If the arrest at a CAG slip-out on the NTS triggers TCR and the repair nucleases cut the TS, the CAG slip-out would be copied during repair-patch formation, leading to expansion. TCR triggered by arrest at the proximal end of the CTG slip-out on the TS could lead to no change in repeat number if incisions are made before the slip-out. TCR resulting from arrest at the distal end of the CTG slip-out, which is larger than the transcription bubble, on the TS could lead to either complete contraction of the slip-out or a smaller contraction. If, following excision of the DNA fragment, the remaining slip-out reanneals and the branch migrates, it would escape further contraction. However, complete contraction might still occur if an additional cleavage event removed the remaining part of the slip-out. Reprinted with permission from ref 65. Copyright 2011 Oxford University Press.

could be even more complex and unique for each different RNAP.)

There are at least two contributions from the nontemplate strand to the elongation process. First, it participates in RNA displacement from the RNA/DNA hybrid within the transcription complex.<sup>27</sup> If RNA is not displaced properly, it forms an extended RNA/DNA hybrid with the DNA template, which could disrupt interactions between nascent RNA and the RNA-exiting groove of RNAP, thus destabilizing the transcription complex in a similar manner to that of a hairpin in the nascent RNA. Another contribution is the direct interaction between the nontemplate strand and RNAP, which can additionally stabilize the transcription complex. These contributions of the nontemplate strand are likely to affect the processivity of RNAP, but they are not ultimate requirements for transcription, and the effect of these contributions is practically unnoticeable for a short template of random sequence, where even the lack of the entire nontemplate strand does not affect transcription elongation.<sup>39</sup> It must be noted, however, that in the cited work, multiple-round transcription was performed; thus, after the first round of transcription utilizing the single-stranded template, hybrids between the nascent RNA and the templated DNA strand are likely to have formed, and these would serve as substrates for the following rounds of transcription. Also, under different conditions, the same (T7) RNAP was unable to transcribe single-stranded template.<sup>67</sup> Thus, the importance of

the nontemplate DNA strand for transcription could vary for different systems.

**3.2.1.3. Slipped-Strand DNA Structures.** Slipped-strand DNA structures (Figure 2B) appear when complementary repetitive sequences hybridize with each other out-of-register. In recent years, these structures have received wide attention because they have been implicated in repeat-expansion diseases (i.e., expansions of individual repeats in the genome that are associated with numerous hereditary disorders in humans) (reviewed in refs 11, 68, 69). In model *in vitro* systems, these structures are frequently obtained by thermal denaturation-renaturation of repetitive sequences, and at least for some sequences, they exhibit remarkable kinetic stability, probably due to mismatched hairpin formation in the loop.<sup>70–72</sup> Whether these structures could be thermodynamically stable and spontaneously form in normal duplex DNA without a denaturation step is less clear: it seems that even under conditions of strong negative supercoiling to facilitate DNA opening, it is more energetically favorable for the slippage loops to merge into a single loop-out, since the latter relaxes the same number of negative supercoils and contains the same number of intrastrand interactions, while leaving fewer energetically unfavorable distortions of the duplex caused by loop extrusion. The most likely mechanism for appearance of these structures *in vivo* seems to be replication accompanied by slippage between the nascent and template DNA strands (reviewed in



**Figure 4.** Z-DNA. (A) Z-DNA versus B-DNA. Black lines show the paths of the DNA strands within the structures. Reprinted with permission from ref 206. Copyright 1996 American Society for Biochemistry and Molecular Biology. (B) B-Z junctions (Reprinted with permission from ref 84. Copyright 2005 Nature Publishing Group). White lines show the paths of the Z-DNA (lower part of the DNA duplex) and B-DNA (higher part of the DNA duplex) strands within the structures. Two bases (in this case A and T) at the B-Z junction are flipped out from the helix.

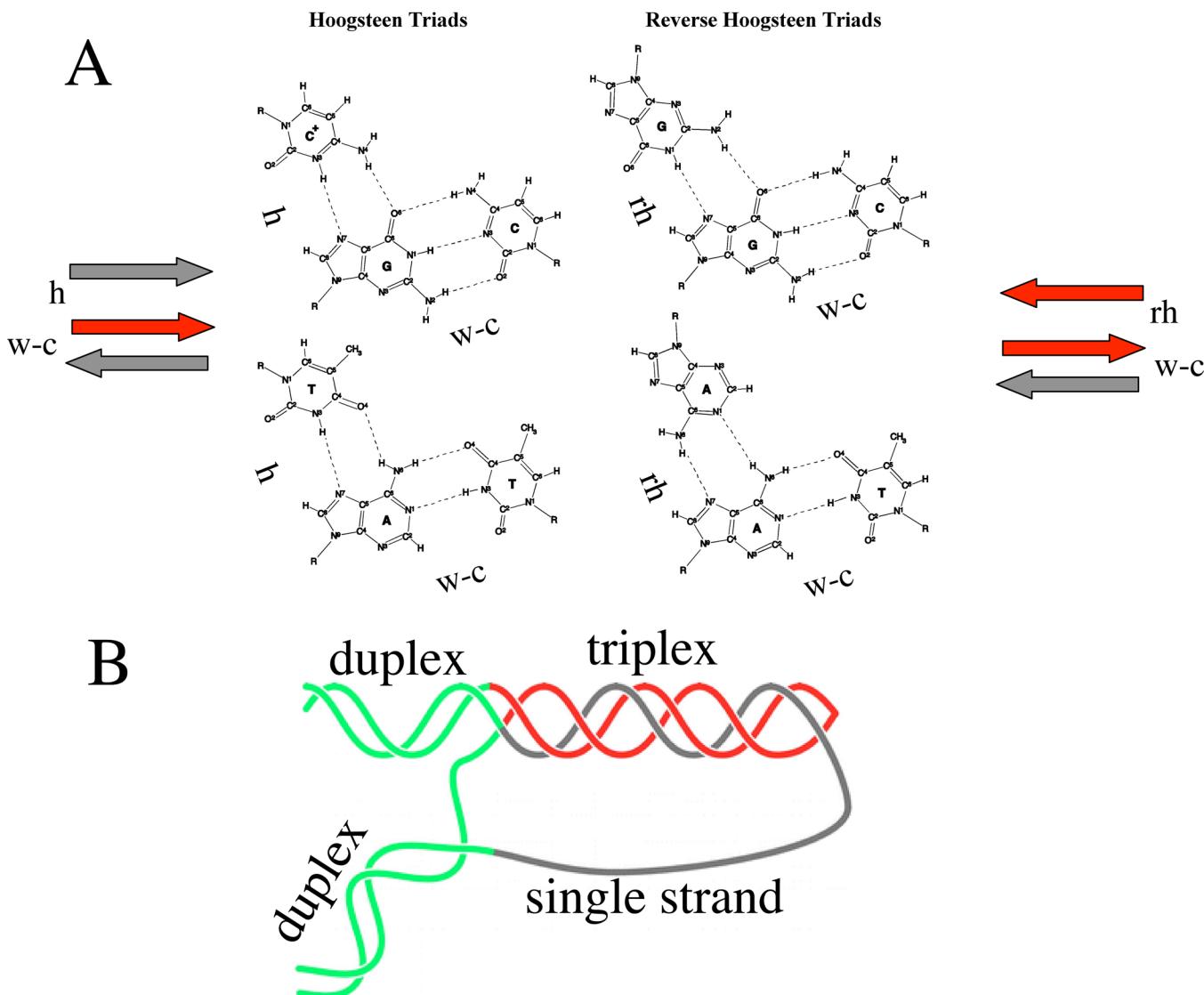
ref 11). However, repeat expansions can also occur in nondividing cells (see ref 73 and references therein). This led to the proposal that slipped-strand structures can be formed during the process of transcription,<sup>74</sup> in particular when it is accompanied by R-loop formation.<sup>75</sup> Furthermore, repeat instability in nondividing cells appears to depend upon gene products involved in transcription-coupled repair (TCR), a pathway of DNA repair triggered by stalled RNAP at a lesion in the transcribed DNA strand.<sup>76</sup> The current model (Figure 3) stipulates that a first round of transcription induces the formation of slipped-strand structures and that these structures (possibly additionally stabilized by mismatch-repair proteins) then stall RNAP during the following rounds of transcription. RNAP stalling attracts the TCR machinery, which is suggested to process these structures in a manner that results in repeat-length changes.<sup>65,76</sup> The idea that slipped-strand structures might block RNAP was examined using their preformed stable analogs *in vitro*.<sup>65</sup> While it appeared that they do not stall T7 RNAP in a purified system, they do stall RNAP II (and to a lesser extent RNAP T7) in HeLa extracts, similar to the results observed for DNA cruciforms and hairpins (see above). That implies that these structures are likely to block transcription *in vivo*, if they are bound to proteins. The nature of the responsible proteins is yet to be established. Interestingly, long CTG repeats in the intact double-stranded form have been reported to temporarily pause RNAP II.<sup>77</sup>

**3.2.2. Z-DNA.** The Z-DNA conformation deviates strikingly from B-form DNA (Figure 4A). Z-DNA is a left-handed helix, with about 12 bp per turn, in contrast to the 10.5 bp per helical turn in the B-form right-handed helix (see ref 78; reviewed in ref 79). This means that if a region of double-stranded DNA in the B-conformation makes a certain number of right-handed turns, in the Z-conformation it will make that number of left-handed turns times  $10.5/12 \approx 0.8$ . In other words, from the

topological point of view, Z-DNA is basically equivalent to a left-handed form of B-DNA. Thus, the local B-to-Z transition within a certain DNA region would be topologically equivalent to complete unwinding of a DNA region almost twice as long as the region of the B-to-Z transition. Consequently, the B-to-Z transition is exceptionally favorable in negatively supercoiled DNA.

The transition from B to Z-DNA generally occurs at sequences with alternating pyrimidines and purines, preferably (CG)<sub>n</sub>. At nonphysiological, high salt concentrations, these sequences adopt the Z- rather than the B-conformation, even in linear polynucleotide sequences. Under physiological ionic conditions, however, Z-DNA per se is less energetically favorable than B-DNA, even for these sequences, and the B-to-Z transition occurs only under the influence of negative supercoiling.<sup>80–82</sup> At a sufficiently high degree of negative supercoiling Z-DNA can be formed even in sequences with strong deviations from the purine-pyrimidine alternation.<sup>83</sup> Typically, the Z-DNA-forming sequences are embedded within “random” DNA sequences that remain in B-form. Because of the very significant structural difference between Z-DNA and B-DNA, there are characteristic distortions in DNA base pairing at the borders between Z-DNA and adjacent B-DNA, called BZ-junctions, in which a single base pairing at each junction is broken, such that the respective bases are flipped out of the double-helix (Figure 4B).<sup>84</sup>

It has been documented that B-Z transitions can occur in living cells and affect mutagenesis, transcription initiation, and recombination, and it was hypothesized that Z-DNA formation in naturally occurring sequences could play a significant role in many biological processes, including regulation of gene expression. This hypothesis is supported by the discovery of several proteins that strongly and specifically bind Z-DNA (reviewed in refs 79, 85). The effect of Z-DNA upon



**Figure 5.** Triplex structures. (A) Triads. Watson–Crick, Hoogsteen, and reverse-Hoogsteen base pairing are designated as w-c, h, rh, respectively. At the left and at the right from respective triads, relative orientations of strands within respective triplexes are shown by block arrows, which point from 5'- to 3'-directions. Homopurine and homopyrimidine sequences are shown in red and gray, respectively. (B) H-DNA structure. The YR<sup>\*</sup>R version of H-DNA (sometimes referred as H-r or H<sup>\*</sup>-DNA) is shown. Homopurine, homopyrimidine, and irregular sequences are shown in red, gray, and green, respectively. It can be seen that when the purine strand (red) folds back to form a triplex, it retraces its path in the duplex, and because of that, one can visualize unwinding this strand from the triplex without disturbing other elements of the structure, which means that H-DNA is topologically equivalent to a completely unwound DNA region.

transcription has been studied in vitro for *Escherichia coli*, T7 RNAP, and wheat germ RNAP II. The mode of interference was evidently different for these respective enzymes and probably also depended upon other conditions: *E. coli* RNAP was completely stalled at the B-Z junction proximal to the promoter.<sup>86</sup> In contrast, most of the T7 RNAP passed through the sequence, and according to one study, the blockage that did occur was at either one of the two B-Z junctions,<sup>87</sup> while another study reported blockage somewhere within the sequence.<sup>88</sup> Wheat Germ RNAP II was shown to transcribe through a poly(CG) sequence when it was in the Z-conformation, though somewhat less efficiently than when that same sequence was in the B-conformation.<sup>89</sup>

The reason why Z-DNA interferes with transcription is not well understood.

A simple energetic consideration<sup>88</sup> suggests that RNAP entering a Z-DNA forming sequence would decrease the length

of the Z-DNA segment, thus increasing negative superhelical stress within the closed circular DNA substrate. This would create an apparent force against RNAP progression through Z-DNA. Moreover, the same pressure would work to expel RNAP as it reaches the distal end of the Z-forming sequence. An alternative hypothesis is that B-Z junctions or other structural features of Z-DNA may inhibit transcription.<sup>87</sup> That would imply that RNAP somehow senses the shape of the downstream duplex DNA template rather than simply pulling the template strand from it.

Note that a priori self-complementary GC motifs might block transcription due to the formation of hairpin or cruciform DNA structures. This question was addressed directly by studying transcription through self-complementary sequences of the same length and G/C content that lacked the ability to adopt Z-DNA. These sequences did not produce significant transcription blockage, suggesting that Z-DNA, rather than

hairpins, caused the blockage.<sup>88</sup> That said, hairpins in the RNA transcript could additionally contribute to the Z-DNA-mediated transcription blockage.

**3.2.3. Triplex Structures.** Within a canonical DNA duplex in which purine and pyrimidine bases are bound via “Watson–Crick base pairing”, purine bases (i.e., G or A) have extra “valences” to bind a third base via another type of hydrogen bonding called Hoogsteen or reverse-Hoogsteen base pairing. The three bases held together are called a triad (Figure 5A), and the regular three-stranded structure comprised of triads is called a triplex (for reviews, see refs 90, 91). At first glance, it would appear that the triplex could be formed at any sequence via base pairing of the third strand with purines in either strand. For a random sequence this would imply that the third strand switches pairing from one strand to the other following the purines. Such strand-switching is, however, sterically difficult to accommodate, and consequently, it is energetically costly. Thus, triplex formation requires sufficiently long homopurine/homopyrimidine (hPu/hPy) sequences, i.e., the sequence composition that would not require frequent strand switching for accommodating the third strand. While strand-switching within the triplex is possible,<sup>92–97</sup> homopurine/homopyrimidine sequences at least several nucleotides long on both sides of the switch are required to form a stable triplex.

While the central strand of the triplex always comprises purines, the third strand comes in various “flavors”. Depending upon the sequence of the third strand, there are two major prototypes of triplexes. The first prototype is pyrimidine–purine–pyrimidine (YR<sup>\*</sup>Y) triplexes (Figure 5A, left), in which the third strand, comprised of pyrimidines (T and C), is oriented parallel to the purine strand of the duplex; T and C in the third strand form Hoogsteen base pairs (symbolized by the asterisk) with A and G, respectively, in the central strand. For Hoogsteen base pairing with G, the C in the third strand has to be protonated; as a consequence, triplexes of this type are stabilized by acidic pH. The second prototype is purine–purine–pyrimidine (YR<sup>\*</sup>R) triplexes (Figure 5A, right), in which the third strand is comprised of purines (A and G) and is oriented antiparallel to the purine strand of the duplex; A and G in the third strand form reverse-Hoogsteen base pairs with A and G (symbolized by the asterisk), respectively, in the central strand. These triplexes do not require acidic pH for stabilization, which makes them more likely candidates for biologically relevant phenomena.

Note that both of these triplex prototypes are often referred to as “anti-parallel” triplexes, because the third strand is antiparallel to the similar (in terms of purine/pyrimidine composition) strand in the duplex. There are numerous variations based upon these two prototypes. For example, in YR<sup>\*</sup>R triplexes, an A in the third strand could be replaced by T, which in this case binds A in the duplex in the antiparallel orientation.<sup>98</sup> Similarly, a G in the third strand could bind G in the central strand, both in parallel and in antiparallel orientations, depending upon the sequence context.<sup>99</sup> There are also some other Hoogsteen-like base pairs, for example, between protonated A and G, which could be readily incorporated into the triplex.<sup>100</sup>

In addition to the triplexes described above, with the central homopurine strand, a model has been suggested for an alternative class of triplexes based upon a different type of base pairing, and presumably formed in arbitrary sequences, often referred to as R-DNA (R in this case stands for recombination, by which the formation of these triplexes was

first proposed).<sup>101–108</sup> In these structures, the third strand is homologous to one of the strands in the duplex and is parallel to it. There are certain general concerns about R-DNA model (e.g., see ref 90); however, several base triads implicated in this model were detected in some DNA<sup>109–111</sup> and RNA<sup>112,113</sup> crystal structures. We will briefly mention these triplexes in the context of “collapsed” R-loops; otherwise, the “triplex” term is used for the structures formed by homopurine/homopyrimidine sequences.

In terms of general organization, triplexes could be either intermolecular (as those formed between a sequence within double-stranded DNA and free single-stranded DNA, or with an RNA oligonucleotide as the third strand) or intramolecular, if formed within one molecule. We are primarily interested in a particular case of intramolecular triplexes, called H-DNA (Figure 5B), since it is readily formed by sequences that are overrepresented in eukaryotic genomes and that play important roles in various genetic transactions. These structures are formed via “disproportionation” between the two adjacent homopurine/homopyrimidine stretches or the two halves of a contiguous homopurine/homopyrimidine stretch, such that one donates a third strand to the other, rendering its complement unpaired. Thus, two duplex regions within the same DNA molecule can convert into a triplex region and a single-stranded region of the same length. Although this is somewhat counterintuitive, it can be shown that the resulting H-DNA structure is topologically equivalent to the complete unwinding of both duplexes, rather than just the one which appears unwound. Because of this unwinding, H-DNA is strongly stabilized by negative superhelical stress (e.g., refs 114, 115).

Depending upon the conditions, either the pyrimidine or the purine strand could be donated as a third strand, resulting in either the YR<sup>\*</sup>Y or YR<sup>\*</sup>R triplex, respectively. Historically, the intramolecular triplex containing the protonated, homopyrimidine third strand (YR<sup>\*</sup>Y triplex) was discovered first,<sup>114</sup> hence the name H-DNA. This structure is now called H-y DNA. The YR<sup>\*</sup>R version of intramolecular triplex,<sup>116–118</sup> commonly referred as H-r DNA, was described soon thereafter. Remarkably, under specific ambient conditions, a symmetrical composite structure that contains both of these triplexes and no single-stranded regions (except for short loops) can be formed.<sup>119,120</sup>

A stable triplex is built of isomorphous base triads: CG<sup>\*</sup>C+ and TA<sup>\*</sup>T for H-y DNA, and CG<sup>\*</sup>G and TA<sup>\*</sup>A for H-r DNA. This limitation leads to a peculiar sequence requirement for H-DNA formation: the identical bases in the donor and in the acceptor parts of a triplex-forming sequence must be positioned symmetrically relative to the center of this sequence. Homopurine/homopyrimidine sequences that possess this symmetry are called H-palindromes [these of course include simple homopurine/homopyrimidine repeats like (GA)<sub>n</sub>]. When the halves of the H-palindrome are either immediately adjacent to each other or are separated by only a few bases, H-DNA readily forms within relatively short sequences under appropriate conditions. Similar reactions between two distant homopurine/homopyrimidine stretches separated by a long duplex region have been suggested,<sup>121</sup> but those may run into sterical problems. Formation of a composite structure built from two distant homopurine/homopyrimidine stretches (referred to as sticky DNA) within a circular plasmid has been well-documented,<sup>122,123</sup> but the model for the resulting structure is yet to be established. It was suggested that all four,

rather than only three, strands might be intertwined within sticky DNA.<sup>124,125</sup> This arrangement would take care of the sterical problems, but the nature of the interaction between the fourth strand with the other strands remains unknown.

What effects are triplexes expected to have upon transcription elongation?

It seems natural to assume that transcription would delay or halt upon running into a triplex. The triplex is additionally stabilized compared to the duplex by the third-strand binding, making it more difficult for the transcription machinery to unwind it. If this is the case, a truncated transcription product produced by blocked RNAP should correspond to the promoter-proximal (upstream) flank of the triplex-forming region (and perhaps several bases into the triplex-forming region). Indeed, this pattern of blockage was observed when an oligonucleotide formed a triplex with the DNA segment downstream from the promoter.<sup>126,127</sup> Surprisingly, a completely different pattern of blockage was observed for transcription of the H-DNA-forming sequence.<sup>41,42</sup> In the latter case, blockage occurred primarily at the promoter-distant (downstream) flank of the H-DNA-forming sequence, i.e., when RNAP had just passed the sequence. Moreover, the blockage occurred only when the pyrimidine strand, but not the purine strand, served as template. This again seems counterintuitive, since the purine strand of a triplex is bound to two other strands, which should make it more difficult to pull that strand over from the triplex than the pyrimidine strand. To explain this paradox, it was suggested that while RNAP is transcribing the homopurine/homopyrimidine sequence, a transiently unpaired region of the nontemplate strand (in close vicinity to the transcription bubble) can fold back to form a triplex with the duplex upstream of RNAP. As transcription continues, the nontemplate strand will continue to spool onto the upstream duplex, effectively forming H-DNA behind the RNAP. This process is likely facilitated by two additional factors: (i) the local negative supercoiling behind the translocating RNAP and (ii) stabilization of H-DNA by binding the nascent transcript to its single-stranded portion. That process would continue until the RNAP reached the downstream end of the hPu/hPy sequence, where it becomes sterically sequestered. At neutral pH and in the presence of magnesium in the standard transcription reaction, H-r DNA, rather than H-y DNA, is expected to form, explaining why the blockage occurs when the Py strand serves as template.

Another possible model for triplex-mediated transcription blockage assumes that the displaced nontemplate homopurine strand can form a triplex with the downstream, rather than the upstream, duplex region. This model was initially proposed to explain triplex-mediated blockage of DNA replication<sup>128</sup> and, more recently, transcription of imperfect hPu/hPy sequences.<sup>37</sup>

Recently, transcription stalling at hPu/hPy sequences has received special attention because of the clinically relevant repeat, (GAA).<sup>n</sup> These repeats are localized in the first intron of the frataxin gene, which plays an important role in iron metabolism in mitochondria. When the total length of repeats is below about 100 nt, the organism functions normally, but significant repeat expansion above this value causes a severe neurodegenerative disease, Friedreich's ataxia (see ref 129; reviewed in refs 11, 13, 130, 131). This disease is likely caused by transcription inhibition in this gene when it contains expanded GAA repeats; their ability to interfere with transcription has been well-documented *in vitro*, and both H-DNA and sticky DNA have been invoked to explain this

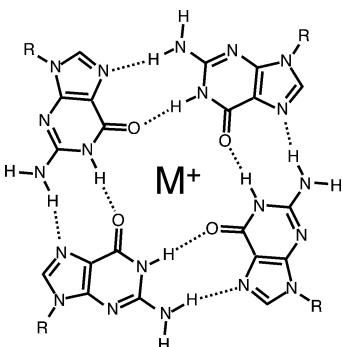
inhibition.<sup>42,132–134</sup> Furthermore, agents that impair the formation of these structures also alleviate transcriptional inhibition, suggesting a potential therapeutic approach for this disease.<sup>135,136</sup>

Another important class of triplex-forming sequences are those located in human gene promoters such as c-Myc,<sup>137</sup> dihydrofolate reductase,<sup>138</sup> and other genes. These are usually short, imperfect, G-rich H-palindromes, which, in addition to triplexes, are also capable of forming quadruplexes (see below). Thus, both multistranded DNA structures have been implicated in models for regulation of gene expression. In the case of the human c-Myc gene, triplex-forming sequences were implicated in genomic instability leading to cancer (see ref 139 and references therein). When transcription proceeds from the major promoter, these sequences are not transcribed. However, these genes also have secondary promoter(s) localized upstream from the major promoter,<sup>140,141</sup> and transcription from these secondary promoters, which are likely to have a regulatory role, does pass through the triplex-forming sequences. For the dihydrofolate reductase gene, transcription from the secondary upstream promoter has been shown to inhibit transcription from the major promoter, and various models, including triplex<sup>138</sup> and quadruplex<sup>142</sup> formation (see below), have been implicated. For T7 RNAP, the triplex-forming sequence from the c-Myc promoter results in a weak but well-defined transcription blockage signal, and effects of various sequence substitutions upon the blockage suggest that H-DNA-like triplex formation contributes to the blockage.<sup>143</sup> Since blocked transcription could lead to DNA rearrangements (see below), one might speculate that it contributes to the genomic instability.

Yet another class of models involving triplexes in transcription regulation proposes direct triplex formation between the nascent RNA and duplex DNA.<sup>138,144</sup> For example, this mechanism has been implicated in the inhibition of the dihydrofolate reductase gene.<sup>138</sup> The nature of this hypothetical triplex remains to be established. With regard to triplex formation between the RNA and DNA chains, one should be aware of the strong effect of the backbone character on triplex stability. For example, while YR\*Y triplexes are readily formed by RNA or a mixture of RNA and DNA chains,<sup>145</sup> stable YR\*R triplexes were not detected when either of the three strands was RNA.<sup>146,147</sup> These results do not completely exclude the possibility that YR\*R triplexes containing RNA chains can be formed transiently. We are not aware of biophysical studies of parallel triplexes (either homopurine/homopyrimidine or “R-DNA type”) containing RNA chains, though there is evidence supporting their formation in the process of transcription.<sup>148,149</sup> In particular, parallel triplex with a central RNA strand (“collapsed R-loop”) has been suggested as a model for stable RNA/DNA hybrid formed in the immunoglobulin  $\alpha$  switch region.<sup>149</sup>

**3.2.4. Guanine Quadruplexes.** Guanine quadruplexes (sometimes referred to as “G4-DNA”) (reviewed in refs 1, 150–152) are comprised of guanine tetrads in which the rings of four guanines are connected via Hoogsteen hydrogen bonds (Figure 6). They are additionally stabilized by a monovalent ion in the center of each tetrad, if it is of an appropriate size. The latter fact leads to a dependence of quadruplex stability on the type of monovalent metal ions: the best stabilizer is potassium, while the worst is lithium.

Guanine quadruplexes were initially discovered in solutions of free guanosine, in which they form tetrads that are stacked



**Figure 6.** G-Quadruplex structure. The G-quartet (tetrad) is stabilized by a monovalent metal ion ( $M^+$ ) in the center.

one above the other.<sup>153</sup> Stacked G4 structures were later observed for single-stranded oligonucleotides containing guanine stretches.<sup>154–157</sup> Usually, three guanines in a row are required,<sup>158</sup> but occasionally, two suffice.<sup>159</sup> Quadruplexes are very versatile: they can be inter- or intramolecular, may have parallel or antiparallel orientation and various topologies of backbone folds, and may include nonguanine bases. They also can be extremely stable. Regions that are capable of G-quadruplex formation (G4-motifs) are ubiquitous elements of the genome (a most important example is found in telomeric sequences), and quadruplexes that could form within these sequences have been implicated in many biological processes.

In contrast to cruciforms, Z-DNA, and H-DNA, quadruplexes do not seem to form spontaneously when the quadruplex-forming sequence is embedded in double-stranded DNA with a random composition, even under negative superhelical stress (e.g., ref 160); as far as we know, their formation from G-rich strands in double-stranded DNA requires DNA unwinding by other factors. In vitro, this can be achieved upon denaturing/renaturing of the dsDNA. Quadruplex formation can also be induced by PNA binding to the complementary C-rich DNA strand.<sup>161</sup> Also, quadruplex could form within the single-stranded region of protonated H-DNA-like structures.<sup>162</sup> In vivo, quadruplex formation could be triggered by DNA unwinding during replication or transcription, as well as by specific protein binding. G-Quadruplexes have recently been detected in human cells using highly potent anti-G4 antibodies. Remarkably, these structures are particularly abundant in the S-phase of the cell cycle, likely forming in the lagging strand during DNA replication.<sup>163</sup>

The inability of a quadruplex to overcompete duplex DNA, even under negative superhelical stress, suggests that the general stability of a quadruplex under usual buffer conditions is lower than that of the corresponding duplex. Quadruplexes appear to be thermodynamically more favorable than the duplex in the presence of a high concentration of polyethylene glycol (PEG) in solution,<sup>164</sup> but even in this case, their formation requires duplex denaturation/renaturation when the quadruplex-forming sequence is flanked by regions with ordinary sequences.<sup>165</sup> It is important to realize, however, that even under conditions in which the overall thermodynamic stability of the Watson–Crick duplex is higher than that of the quadruplex, the latter might still present a greater challenge for DNA unwinding enzymes like helicases (reviewed in ref 166) or the RNAP (see below) than the former. First, quadruplexes usually comprise very energetically favorable guanine tetrads and energetically unfavorable loops. The total energetic stability of the entire structure could be less than that of the

corresponding Watson–Crick duplex, but because the kinetics of stable structure disruption is mostly defined by the largest energetic barrier, the rate of the quadruplex unwinding by enzymes would be mostly defined by the energy of the guanine tetrads, rather than by the whole energy of the structure. Second, even under conditions in which the average energy of interaction per single base pairing within the guanine tetrad is less than that in the Watson–Crick duplex, the “pulling out” of the first guanine from the G-tetrad could comprise a larger energetic barrier than disruption of one Watson–Crick base pair, because, in the former case, two base-pairings instead of one must be disrupted simultaneously, and that could be a limiting factor for structure unwinding. Since only one nucleic acid strand is sufficient for quadruplex formation, it can be formed in nascent RNA or in either strand of the duplex DNA template during transcription. All three scenarios have been considered to explain the effect of quadruplexes on transcription.

**3.2.4.1. Quadruplexes within the Nascent RNA.** Quadruplexes in nascent RNA could result in transcription blockage or stalling by competing with RNA binding to the exiting groove of RNAP and destabilizing the transcription complex, i.e., similar to RNA hairpins (see above). This mechanism was invoked to explain premature termination of transcription by T7 RNAP at a short G-stretch in the nontemplate strand positioned immediately downstream of the T7 promoter.<sup>36</sup> However, quadruplex formation in nascent RNA did not seem to contribute to T7 RNAP blockage at a longer G-rich sequence in the nontemplate strand positioned further downstream from the promoter.<sup>37,143</sup> Finally, quadruplex formation in either the nascent RNA or hybrid DNA/RNA or quadruplex formation between the nascent RNA and the nontemplate DNA strand is likely to be responsible for the specific blockage of mitochondrial RNAP that is required to create an RNA primer to initiate mitochondrial replication.<sup>6,167</sup>

It was also proposed that some proteins can mediate binding of quadruplexes in the nascent RNA to duplex DNA. This model was used to explain inhibition of expression of the dihydropholate reductase gene by a transcript from the upstream secondary promoter<sup>142</sup> and to account for the sequestering of the so-called telomeric RNA (TERRA) within the telomeres.<sup>168</sup> We will discuss those mechanisms below in the context of RNA anchoring.

**3.2.4.2. Quadruplexes within the Nontemplate DNA Strand.** Quadruplexes in the nontemplate strand have been shown to form during transcription in immunoglobulin class switch regions. In this case, nascent RNA forms an extended duplex with the template strand (like in R-loops), while the nontemplate strand folds into a G-quadruplex. Thus, the resulting structures were named G-loops.<sup>169</sup> These G-loops specifically bind activation-induced cytidine deaminase (AID), an enzyme that initiates class switch recombination and somatic hypermutation in B cells by targeted deamination of cytosines in transcribed genes.<sup>170</sup> Both T7 RNAP and RNAP II transcription in vitro through this region exhibit partial blockage in the orientation in which G-loops are formed, suggesting that G-quadruplex formation is responsible for the blockage.<sup>38</sup> Notably, G-rich sequences at the immunoglobulin switch recombination sites are rather long (hundreds of nucleotides). Our studies with much shorter G-rich sequences show that quadruplex formation in the nontemplate strand does not contribute to transcription blockage.<sup>37</sup> Thus, the effect seems to depend upon the length of the G4-DNA sequence,

and its detailed mechanism requires further investigation. In any case, the ability of these sequences to interfere with transcription might contribute to important mechanisms that regulate gene expression; for example, promoter-proximal transcription pausing.<sup>171</sup>

#### 3.2.4.3. Quadruplexes within the Template Strand.

Transcription blockage by quadruplexes in the template strand were studied *in vitro* using either partially double-stranded substrates designed to contain a single-stranded G4-motif triggering quadruplex formation or perfect duplexes that had been denatured, followed by renaturing under conditions favoring quadruplex formation.<sup>172</sup> In both systems, a partial blockage of transcription was observed roughly at the beginning of the quadruplex-forming region, consistent with RNAP being unable to efficiently unwind the quadruplex. G4 motifs in the template strand of the perfect duplex did not block transcription, consistent with the notion that DNA opening is required for quadruplex formation. It is interesting to note, in this regard, that many DNA regions are transcribed *in vivo* in both directions. Thus, a quadruplex formed by transcription in one direction could serve as an obstacle for transcription running in the opposite direction.

#### 3.2.5. R-Loops.

As we have stated above, RNA polymerases need to be able to physically separate the nascent RNA from the template DNA strand in the course of transcription elongation. This is necessary, because the RNA/DNA duplex is more stable than the corresponding DNA/DNA duplex for most natural sequences. Thus, although the nontemplate strand is longer than the nascent RNA and is attached to DNA duplex at both ends, its competition with the nascent RNA for hybridization with the template strand might be insufficient for the separation of nascent RNA from its DNA template.

During transcription, only a short (around 8–10 bp) RNA/DNA duplex is formed within the transcription complex, while a special protein wedgelike moiety “peels” the nascent RNA off the RNA/DNA duplex, and RNA is extruded from the transcription complex via the positively charged RNA exiting channel.<sup>24–26</sup> This emerging nascent RNA chain is instantly covered by proteins *in vivo*, which further prevents its reannealing with the template DNA. Yet despite all efforts to prevent RNA reannealing, it occasionally happens, resulting in the formation of R-loops, which are usually deleterious in cells (reviewed in ref 173) and must be removed by special helicases<sup>174–176</sup> or by RNase H digestion.<sup>177</sup> In some cases, however, R-loops have important regulatory functions (e.g., refs 169, 170, 178, 179), and they also participate in replication initiation in some replicons (e.g., refs 4, 5, 180).

According to current models for R-loop formation, the nascent RNA either could be extruded from the transcription complex in a single-stranded form followed by invasion of the upstream DNA duplex (threading-back model)<sup>181,182</sup> or it could exit the transcription complex in an “unusual” manner as a contiguous RNA–DNA duplex, with no separation or only minor separation from the DNA template strand.<sup>27,148,183</sup> Several factors increase the probability of R-loop formation: (i) negative supercoiling, which facilitates (both kinetically and thermodynamically) DNA strand separation (see ref 182; for review see ref 184); (ii) nicking of the nontemplate strand, which allows the transiently opened nontemplate DNA strand to diffuse further away from the template, thus decreasing its propensity to compete with nascent RNA for hybridization with the template DNA strand;<sup>182</sup> (iii) sequences with increased stability of the RNA/DNA versus the DNA/DNA

duplex;<sup>185</sup> and (iv) sequences in which the nontemplate DNA strand is able to form unusual DNA structures, which preclude its rehybridization with the template DNA strand, interfering with RNA displacement.<sup>169,186</sup>

The effect of factors ii and iii was shown to be much stronger close to promoters, likely because a shorter transcript creates less sterical problems for RNA invasion.<sup>182</sup>

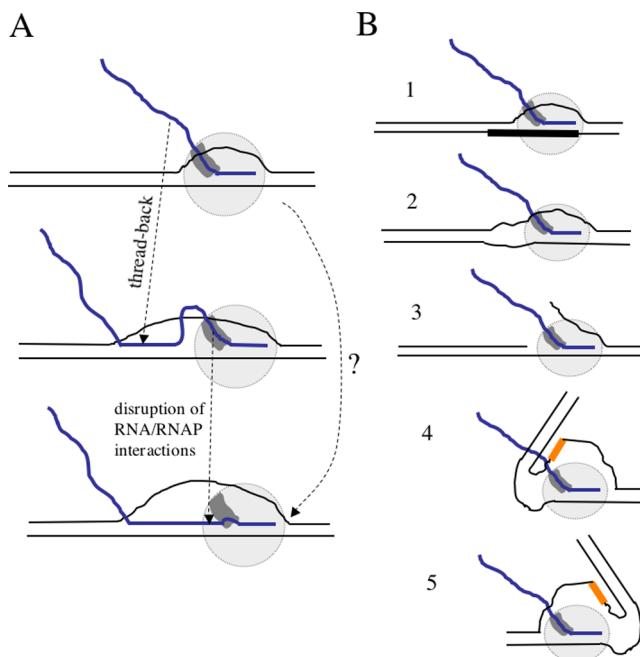
As for factors iii and iv, the families of DNA sequences that satisfy these requirements partially overlap, complicating the interpretation of experimental results. In fact, differential stabilities of RNA/DNA and DNA/DNA duplexes are maximal for G-rich purine RNA/C-rich pyrimidine DNA duplexes, promoting R-loop formation in DNA sequences with G-rich homopurine runs in the nontemplate strand.<sup>145,187</sup> The very same sequences have a propensity to form G-quadruplexes and, in some cases, intramolecular tripleplexes. These structures could additionally stabilize R-loops. However, in the Appendix, we show that for a sufficiently G-rich sequence, the stability of the DNA/RNA hybrid alone, without invoking an unusual DNA structure, could be sufficient to account for RNA sequestration within the DNA duplex under positive superhelical stress.<sup>149</sup> Additional analyses involving modification of nucleotide sequences and ambient conditions are required to distinguish between structures that might be responsible for transcription inhibition by a given sequence. Interpretation of biological effects is additionally complicated by the fact that poly-dG sequences also bind certain transcription regulatory factors,<sup>188</sup> which might also contribute to their effects on transcription.

R-Loops have been implicated in numerous biological processes, and there are abundant studies related to their formation *in vitro* and *in vivo*, as well as their biological roles (e.g., ref 173). There are relatively few studies, however, in which connections between R-loop formation and transcription impediments have been directly investigated. The simplest mechanism for R-loop interference with transcription is that resulting when an RNAP collides with an R-loop formed by the preceding RNAP. Because of the higher stability of RNA/DNA hybrids and the structural differences between them and DNA/DNA hybrids, it could be difficult for RNAP to unwind those structures, resulting in impeded or blocked transcription. In accord with this mechanism, a 2-fold decrease has been observed *in vitro* in the total yield of transcription through a long (300 nt), artificially preformed R-loop.<sup>189</sup>

Our recent studies of transcription blockage by relatively short G-rich homopurine/homopyrimidine sequences have suggested another mechanism for R-loop interference with transcription.<sup>37,40</sup> In this case, the blockage was observed only when the purine-rich RNA was the product, and the blockage sites were mapped to the promoter-distal part of the sequence and even farther downstream into flanking sequences. In other words, blockage occurred predominantly when the RNAP had already passed the causative sequence, in contrast to the expected result in the case of collisions with preformed R-loops. The effect of various solution conditions and nucleotide substitutions suggests that R-loop formation per se is sufficient to cause this blockage. This conclusion was further supported by the observation that nicks in the nontemplate strand (which are known to facilitate R-loop formation) dramatically increase the blockage when they are located in close proximity to the causative sequence.<sup>40</sup>

Remarkably, a nick in the nontemplate strand result in the appearance of weak blockage signals downstream from the nick, even in random DNA sequences.

This reveals a fundamental mechanism for transcription blockage by R-loops that does not require specific sequences. One such mechanism could be that R-loop formation disrupts nascent RNA interactions with the RNAP, in a somewhat similar manner to that due to hairpin formation within the nascent RNA, thus destabilizing the transcription complex and increasing the probability of blockage (Figure 7). In support of



**Figure 7.** Possible mechanism for R-loop interference with transcription. (A) The basic mechanism. Normally, the nascent RNA (shown by a dark blue line) interacts with a certain area (shown by darker gray patch) of RNA polymerase (shown by a gray oval with dotted line border). R-loop formation (either via a thread-back mechanism or some other mechanism shown by a dashed line with a question mark) disrupts this interaction, thus destabilizing the elongation complex and making it more prone to stalling and/or dissociation. (B) Factors that exacerbate the blockage by facilitating R-loop formation: (1) a sequence that forms an extrastable RNA/DNA hybrid (shown by a thick black line); (2) negative supercoiling that increases transient opening of DNA; (3) a nick in the nontemplate strand, which decreases propensity of that strand to displace RNA; and involvement of a part of the nontemplate DNA strand (shown by the thick orange line) in triplex formation with the DNA duplex upstream (4) or downstream (5) of the transcription complex; which would sequester the nontemplate DNA strand and thus decrease its propensity to displace RNA. In addition to facilitating R-loop formation, some of these factors could additionally exacerbate blockage by other mechanisms. For example, an extrastable RNA/DNA hybrid inside the transcription complex could interfere with nascent RNA separation, and triplets could create obstacles for RNAP movement.

this interpretation, R-loop formation has been shown to exacerbate weak transcription termination in stretches of oligo-dT/dA.<sup>22</sup> In general, exacerbation of very weak pausing or termination signals, which should be present in any “random” sequence, by R-loop formation is likely to be the source of the minor blockage signals that extend downstream from the causative sequence.

#### 4. MORE COMPLEX PHENOMENA INVOLVING ANOMALOUS TRANSCRIPTION

Anomalous transcription can lead to transcription-associated mutagenesis (TAM), transcription associated recombination (TAR), and other phenomena. Because there are many reviews related to this topic (for example, refs 173, 190–192), we will only briefly discuss two phenomena that are directly related to our research: transcription collisions with replication and transcription-coupled repair, and the interplay among transcription, unusual structures, and chromatin remodeling. In the last subsection, we consider hypothetical protein-mediated anchoring of nascent RNA to DNA and its possible consequences for transcription.

##### 4.1. Transcription–Replication Collisions

Collisions between replication and transcription can lead to stalled replication forks and consequently facilitate double-strand-break formation and other destabilizing events. Importantly, a stalled RNAP is able to block replication both for codirectional and “head-on” collisions, while the normally elongating RNAP blocks replication only in “head-on” collisions (reviewed in ref 193). This means that while genomic instability due to transcription–replication collisions might be alleviated by placing genes in orientations in which their transcription is codirectional with replication, for stalled transcription that would not solve the problem; thus, stalled replication at an arrested RNAP would be more “mutagenic” than that at an actively translocating RNAP.

Replication blockage during codirectional collision of replication with transcription was observed within G<sub>n</sub>/C<sub>n</sub> sequences inserted into plasmids propagated in *E. coli*, leading to the proposal that these sequences stalled elongating RNA polymerases and that this, in turn, blocked replication forks.<sup>194</sup> The replication blockage was pronounced only when the nontemplate (sense) strand for transcription was G<sub>n</sub>, in accordance with results for T7 RNAP transcription blockage *in vitro*,<sup>37,40</sup> implying the same general mechanism for transcription blockage in these two systems. It is tempting to suggest, therefore, that similar structures, i.e., R-loops and triplets, are responsible for transcription blockage *in vitro* and transcription-mediated replication blockage *in vivo*. Similar phenomena could also occur in eukaryotic cells. Since hPu/hPy repeats and G-rich stretches are abundant in eukaryotic genomes, the proposed mechanism of transcription-dependent replication blockage could contribute importantly to gross chromosomal rearrangements at these sequences in various genetic processes.

##### 4.2. Transcription-Coupled Repair

Transcription-coupled repair (TCR) is the specialized subpathway of nucleotide excision repair that targets blocked RNA polymerases to efficiently remove the responsible lesions from the transcribed DNA strands (reviewed in refs 195–198). It was hypothesized that if transcription were blocked in undamaged DNA, TCR might occasionally activate futile cycles of DNA repair, eventually leading to mutagenesis and destabilizing the genome.<sup>199</sup> That suggests that DNA sequences that block transcription would be prone to various forms of genetic instability. Importantly, some enzymes required for the TCR pathway are specific for TCR and are not involved in general repair pathways. Thus, the role of TCR in some forms of genetic instability might be deduced not only from the transcription dependence of this instability (which could

involve different mechanisms) but also from the dependence of the instability upon TCR-specific enzymes. An important recent example of this kind is the implication of TCR in trinucleotide repeat instability.<sup>76</sup> There is also evidence that DNA triplets can trigger gratuitous TCR.<sup>200</sup> Finally, Z-DNA-induced mutagenesis in certain systems exhibits some dependence upon transcription, leading to speculation that TCR might be involved in this process.<sup>12,201</sup>

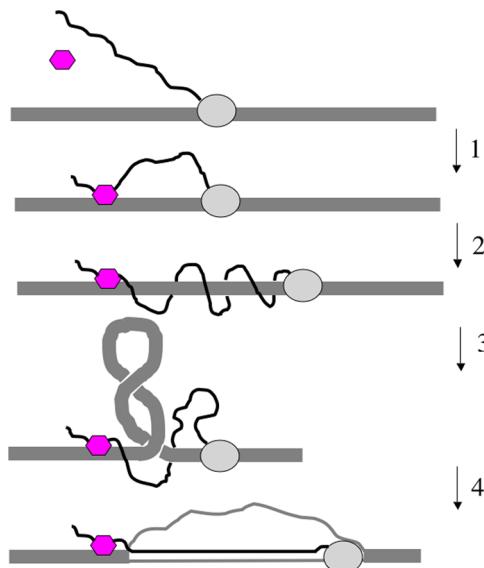
#### 4.3. Interdependence between Transcription and Unusual DNA Structures Mediated by Chromatin Remodeling

In eukaryotic cells DNA is organized in ordered chromatin structures, which are comprised of nucleoprotein particles called nucleosomes, formed by wrapping the DNA around histone protein octamers; displacement or redistribution of nucleosomes within chromatin is called chromatin remodeling. Nucleosomes present obstacles for transcription initiation and elongation, and they need to be displaced to make transcription possible. Importantly, in the elongation stage, various RNAPs are capable of displacing nucleosomes (reviewed in ref 202). Unusual DNA structures, in general, are likely to interfere with nucleosome formation, and vice versa, either due to their intrinsic physical characteristics (e.g., increased stiffness, which disfavors DNA wrapping) (reviewed in ref 203) or possibly due to structure-specific proteins, which could compete with binding of histones. Thus, unusual structure formation could affect transcription indirectly by inhibiting nucleosome formation. This mechanism has been implicated in the positive regulation of transcription by Z-DNA (see refs 203–205 and references therein). Moreover, RNAP in the elongation mode could displace nucleosomes from sequences prone to form unusual structures, thus promoting unusual structure formation in those regions. That would be expected to interfere with nucleosome reformation, thus facilitating further rounds of transcription or transcription initiation from other promoters within the transcribed region (this of course implies that the nucleosome presents a stronger obstacle for transcription than the unusual DNA structure). In fact, this might represent an additional mechanism for regulation of transcription.

#### 4.4. Nascent RNA Anchoring to DNA and Its Implications

What would happen if during normal transcription a region within the nascent RNA became rigidly attached (anchored) to the DNA duplex somewhere upstream from the transcribing RNAP (Figure 8)? Since RNAP follows a helical path along the DNA duplex, its rotation relative to the duplex should cause nascent RNA wrapping around the duplex. This wrapping results in a decrease in the conformational freedom of nascent RNA, creating an entropic force that attempts to unwrap RNA from the duplex. As RNA is rigidly bound to DNA at two positions, one at the anchoring point and the other at the transcription complex, unwrapping of RNA from the duplex could only occur at the expense of DNA unwinding. In other words, the unwrapping force would generate negative supercoiling in the wrapped part of the DNA duplex. According to our estimates,<sup>53</sup> this negative supercoiling becomes stronger as transcription continues, ultimately leading to RNA invasion into the duplex (i.e., R-loop formation). Prior to the RNA invasion, this supercoiling could also facilitate the formation of unusual DNA structures in the wrapped DNA region.

Anchoring might occur either via direct interaction between RNA and DNA (like triplex or R-loop formation) or by proteins, which could simultaneously bind double-stranded DNA and single-stranded RNA. For example, the strongest Z-



**Figure 8.** Nascent RNA anchoring and its consequences. DNA and RNA are shown as dark gray and black lines, respectively. RNA polymerase (RNAP) and an anchoring agent are shown as a light gray oval and as a magenta hexagon, respectively. After the anchoring agent binds the nascent RNA to DNA (1), further transcription causes RNA wrapping around DNA (2), which generates negative superhelical stress in the wrapped region of DNA (3), and this eventually leads to RNA invasion of the DNA duplex (4).

DNA binding protein ADAR is also an RNA editing enzyme (reviewed in ref 206). It was suggested that it binds Z-DNA-forming sequences in the promoter regions and modifies nascent RNA while transcription is taking place. It is tempting to speculate that transient binding of the nascent RNA to the DNA-bound protein during RNA editing might cause RNA wrapping and consequently generate negative superhelical stress in the DNA, promoting Z-DNA formation. Another model that implicates protein-mediated anchoring of the nascent RNA is the regulation of the dihydrofolate reductase gene by transcription from the secondary promoter.<sup>142</sup> According to this model (which is an alternative to the triplex-based model for the same gene,<sup>138</sup> mentioned above), a G-rich region of the transcript forms a quadruplex, which is recognized by a protein(s) that could also bind DNA, thus causing RNA anchoring. A similar quadruplex–protein mediated RNA-to-DNA anchoring was suggested for telomeric repeat-containing RNA (TERRA).<sup>168</sup> Yet another example is the transcription factor YY1, which can simultaneously bind RNA and DNA; according to the model, it tethers nascent RNA to DNA cotranscriptionally, and this tethering plays an important role in X-chromosome silencing.<sup>207,208</sup>

According to our hypothesis, anchoring that occurs cotranscriptionally should ultimately lead to R-loop formation, which might contribute to transcription inhibition and enhance RNA sequestration. Generally speaking, protein-mediated cotranscriptional RNA anchoring should potentially lead to efficient R-loop formation at arbitrary sequences, including those that are normally not prone to R-loop formation. This could explain the numerous R-loops experimentally detected in yeast genomes for sequences without purine/pyrimidine biases.<sup>209</sup> We believe that protein-regulated R-loops could have numerous regulatory consequences, for example, in

replication initiation, regulation of gene expression, and various kinds of genome instability.

## APPENDIX

### R-Loop Stability against Positive Supercoiling

As mentioned before, R-loop formation relaxes negative supercoiling; thus, R-loops are much more efficiently formed and much more stable in negatively supercoiled DNA in comparison with linear DNA. In contrast, R-loop formation in closed circular relaxed DNA would generate positive supercoiling, which would tend to “expel” the R-loops from DNA. However, for some sequences, RNA/DNA hybrids could be sufficiently stable to allow RNA to invade closed circular relaxed DNA and resist the displacement facilitated by positive supercoiling. To illustrate this statement, we will estimate the stability of the R-loop formed by the homopurine sequence from the immunoglobulin switch region (AGGAG).<sup>28</sup> As usual, the sequence of the nontemplate strand (same sequence as the RNA transcript) is shown. Transcription of this sequence embedded into relaxed closed circular plasmid DNA produces positively supercoiled RNA-containing species, indicating that some RNA-containing structures are formed that are stable enough to resist displacement by positive supercoiling.<sup>149</sup> The purpose of these estimates is to answer the question, is Watson–Crick RNA/DNA hybrid formation for this particular system sufficient to explain this stability?

The positive superhelical density,  $\sigma$ , which is generated by unwinding of  $x$  base pairs (e.g. by R-loop formation) in relaxed circular closed plasmid with a total length of  $N$  base pairs is

$$\sigma = \frac{x}{N}$$

The superhelical energy is

$$\Delta G_{sc} = 10NRT\sigma^2 = \frac{10RTx^2}{N}$$

where  $RT$  is the product of the gas constant and the absolute temperature (reviewed in ref 45).

Substituting  $x = 140$  bp and  $N = 5.5$  kbp (from refs 149, 210), we obtain

$$\Delta G_{sc} = 21 \text{ kcal/mol}$$

The difference in free energies between RNA/DNA and DNA/DNA hybrids for the sequence (AGGAG)<sub>28</sub> as calculated using the HyTher program (Peyret, N., & SantaLucia, J., Jr., Wayne State University)<sup>211,212</sup> for 30 mM monovalent cations at 25°C, similar to the TBE buffer for agarose electrophoresis used to analyse these structures in refs 149, 210, is

$$\Delta G_{RNA/DNA-DNA/DNA} = -42 \text{ kcal/mol}$$

Thus, the total energy

$$\Delta G = \Delta G_{sc} + \Delta G_{RNA/DNA-DNA/DNA} = -21 \text{ kcal/mol}$$

is negative, i.e., the extra stability of the RNA/DNA hybrid overcompensates the energetically unfavorable increase in supercoiling. Note that here we have assumed that the length of the R-loop corresponds to the entire (AGGAG)<sub>28</sub> insert, but in principle, for some values of parameters it might be more favorable only for the part of the insert that will be invaded by RNA; thus, the energy could be further optimized.

This energy would provide practically unlimited kinetic stability of R-loops under these conditions: using the analogy

with the DNA melting theory,<sup>213</sup> one can estimate the time required to displace this R-loop via branch migration

$$\tau = \frac{\tau_0 e^{-\Delta G/RT}}{x}$$

(where  $\tau_0$  is the step-time for strand displacement, about  $12 \times 10^{-6} \text{ s}^{214}$ ), which is on the order of  $10^4$  h. Note that our estimations were performed for electrophoretic TBE buffer, rather than transcription buffer, because the former lacks magnesium ions. Magnesium ions are very strong stabilizers of nucleic acids structures; thus, in the transcription buffer in which magnesium ions are present, the stability of R-loops is expected to be even higher. Of course, this result does not exclude the possibility of formation of other structures, and it should be considered as a “proof-of-concept” estimate to demonstrate the capabilities of the R-loops alone.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: hanawalt@stanford.edu.

### Funding

This work was supported by NIH grant CA077712 from the National Cancer Institute to P.C.H., and GM60987 from the National Institute of General Medicine to S.M.M.

### Notes

The authors declare no competing financial interest.

### Biographies



Boris P. Belotserkovskii received his B.S. and M.S. in Biotechnology from Mendeleev Institute of Chemical Technology (Moscow, Russia) in 1989 and obtained his Ph.D. in Biophysics from Moscow Physical–Technical Institute (Moscow, Russia) in 1992 for research on unusual DNA structures that appear under acidic pH and negative superhelical stress, which was done at the Institute of Molecular Genetics (Moscow, Russia) in the laboratory of Prof. M. D. Frank-Kamenetskii. Upon his arrival in the United States in 1993, he first worked as a postdoc at Stanford Research Institute (SRI International) in the laboratory of Dr. B. H. Johnston, primarily researching DNA triplets, and then at Pangene Corp, exploring RecA protein mediated DNA targeting. After that he worked as an associate specialist at the University of California—Berkeley in the laboratory of Prof. N. R. Cozzarelli, primarily investigating DNA topoisomerases. Currently, he is working at Stanford University in the laboratory of Prof. P. C. Hanawalt studying transcription through various DNA sequences with unusual structural properties. He has 27 publications in peer-reviewed journals related to various aspects of DNA structures and physical behavior, as well as various DNA/protein interactions.



Prof. Sergei Mirkin was born in Moscow, USSR, in 1956. He received his B.S. and M.S. in Genetics from the Moscow State University in 1978, followed by a Ph.D. in Molecular Biology from the Institute of Molecular Genetics, Russian Academy of Science, in 1983. During his graduate studies under the supervision of Prof. Roman B. Khesin, he isolated one of the first conditionally lethal mutants of DNA gyrase and found that this enzyme is essential for both DNA replication and transcription in *E. coli*. He then carried out his postdoctoral studies under the supervision of Prof. Maxim D. Frank-Kamenetskii, studying conformational transitions of DNA repeats in superhelical DNA. His research led to the discovery of the first multistranded DNA structure, called H-DNA, which is formed by homopurine/homopyrimidine mirror repeats. This pioneering study triggered a worldwide interest in triplex DNA and other multistranded DNA structures. He was appointed a Group Leader at the Institute of Molecular Genetics in Moscow in 1988. Anticipating the demise of the scientific funding upon the collapse of the Soviet Union, he moved to the United States in 1989 to become an Assistant Professor at the University of Illinois at Chicago, College of Medicine, in 1990. During his years at UIC, he rose in ranks to the Full Professor, establishing himself as a leader in the field of DNA structure and functioning, broadly defined. One of his major achievements was unraveling the replication mechanism for the expansion of triplet repeats—a phenomenon responsible for more than 30 hereditary diseases in humans. In 2007, he moved to Tufts University to become Professor and to hold the White Family Chair in Biology. During his career, he has published over 80 scientific papers, including numerous book chapters and scientific reviews. He is an Editor-in-Chief of *Current Opinion in Genetics and Development*, a Managing Editor for *Frontiers in Biosciences*, and an Advisory Board Member for *Molecular Biology*.



Phil Hanawalt is the Morris Herzstein Professor of Biology at Stanford University, where he has served as Department Chair and as Director of the Biophysics Graduate Program. He received his B.A. from

Oberlin College and earned an M.S. in Physics and Ph.D. in Biophysics at Yale University (with R. B. Setlow), where he initiated studies on the recovery of DNA synthesis in UV-irradiated bacteria that led to his co-discovery of DNA excision-repair in 1964. He carried out postdoctoral research at the University of Copenhagen (with O. Maaløe), on regulation of the bacterial DNA replication cycle, and at Caltech (with R. L. Sinsheimer) before joining the Stanford faculty in 1961. He has mentored 29 Ph.D. students and over 60 postdoctoral scholars from over 30 countries. Hanawalt and his students studied intragenomic heterogeneity of DNA repair in mammalian cells and discovered the pathway of transcription-coupled repair (TCR) in the early 1980s. He is currently interested in the precise signals that initiate TCR to overcome transcription blockage at lesion sites and the possibility of gratuitous TCR in undamaged DNA that may contribute to genomic instability. Dr. Hanawalt is a member of the National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences, the American Academy of Microbiology, and the American Association for the Advancement of Science, as well as a Foreign Associate of the European Molecular Biology Organization. He has served as President of the Environmental Mutagen Society (EMS) and on the Board of Directors for the American Association for Cancer Research (AACR). He has won annual research awards from the American Society for Photobiology and the EMS, the International Mutation Research Award in 1987, and the AACR–Princess Takamatsu Award/Lectureship in 2011. He was co-founding Editor of the journal *DNA Repair:Mutation Research*, a member of the Board of Reviewing Editors for *Science*, and a Senior Editor for *Cancer Research*, and he currently serves on the Editorial Board for the *Proceedings of the National Academy of Sciences of the United States of America*.

## ACKNOWLEDGMENTS

We thank Graciela Spivak for critical reading of the manuscript.

## REFERENCES

- (1) Maizels, N. *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055.
- (2) Larson, E. D.; Maizels, N. *Genome Biol.* **2004**, *5*, 211.
- (3) Azzalin, C. M.; Lingner, J. *Cell Cycle* **2008**, *7*, 1161.
- (4) Masukata, H.; Dasgupta, S.; Tomizawa, J. *Cell* **1987**, *51*, 1123.
- (5) Baker, T. A.; Kornberg, A. *Cell* **1988**, *55*, 113.
- (6) Wanrooij, P. H.; Uhler, J. P.; Shi, Y.; Westerlund, F.; Falkenberg, M.; Gustafsson, C. M. *Nucleic Acids Res.* **2012**, *40*, 10334.
- (7) Masukata, H.; Tomizawa, J. *Cell* **1990**, *62*, 331.
- (8) Masai, H.; Arai, K. *Biochimie* **1996**, *78*, 1109.
- (9) Carles-Kinch, K.; Kreuzer, K. N. *J. Mol. Biol.* **1997**, *266*, 915.
- (10) Bacolla, A.; Wells, R. D. *Mol. Carcinog.* **2009**, *48*, 273.
- (11) Mirkin, S. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 351.
- (12) Wang, G.; Vasquez, K. M. *Mutat. Res.* **2006**, *598*, 103.
- (13) Greene, E.; Handa, V.; Kumari, D.; Usdin, K. *Cytogenet. Genome Res.* **2003**, *100*, 65.
- (14) Wells, R. D. *Trends Biochem. Sci.* **2007**, *32*, 271.
- (15) Cox, R.; Mirkin, S. M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 5237.
- (16) Kapranov, P.; St. Laurent, G. *Front. Genet.* **2012**, *3*, 60.
- (17) Zhang, Y.; Shishkin, A. A.; Nishida, Y.; Marcinkowski-Desmond, D.; Saini, N.; Volkov, K. V.; Mirkin, S. M.; Lobachev, K. S. *Mol. Cell* **2012**, *48*, 254.
- (18) Zhou, W.; Doetsch, P. W. *Biochemistry* **1994**, *33*, 14926.
- (19) Neil, A. J.; Belotserkovskii, B. P.; Hanawalt, P. C. *Biochemistry* **2012**, *51*, 8964.
- (20) Martin, F. H.; Tinoco, I., Jr. *Nucleic Acids Res.* **1980**, *8*, 2295.
- (21) Wilson, K. S.; von Hippel, P. H. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8793.
- (22) Tomizawa, J.; Masukata, H. *Cell* **1987**, *51*, 623.

- (23) Bochkareva, A.; Yuzenkova, Y.; Tadigotla, V. R.; Zenkin, N. *EMBO J.* **2012**, *31*, 630.
- (24) Daube, S. S.; von Hippel, P. H. *Biochemistry* **1994**, *33*, 340.
- (25) Yin, Y. W.; Steitz, T. A. *Cell* **2004**, *116*, 393.
- (26) Jiang, M.; Ma, N.; Vassylyev, D. G.; McAllister, W. T. *Mol. Cell* **2004**, *15*, 777.
- (27) Gopal, V.; Brieba, L. G.; Guajardo, R.; McAllister, W. T.; Sousa, R. *J. Mol. Biol.* **1999**, *290*, 411.
- (28) Mentesana, P. E.; Chin-Bow, S. T.; Sousa, R.; McAllister, W. T. *J. Mol. Biol.* **2000**, *302*, 1049.
- (29) Yarnell, W. S.; Roberts, J. W. *Science* **1999**, *284*, 611.
- (30) Santangelo, T. J.; Roberts, J. W. *Mol. Cell* **2004**, *14*, 117.
- (31) Zamft, B.; Bintu, L.; Ishibashi, T.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 8948.
- (32) Turnbough, C. L., Jr. *Curr. Opin. Microbiol.* **2011**, *14*, 142.
- (33) Strathern, J.; Malagon, F.; Irvin, J.; Gotte, D.; Shafer, B.; Kireeva, M.; Lubkowska, L.; Jin, D. J.; Kashlev, M. *J. Biol. Chem.* **2012**, *288*, 2689.
- (34) Strathern, J. N.; Jin, D. J.; Court, D. L.; Kashlev, M. *Biochim. Biophys. Acta* **2013**, *1819*, 694.
- (35) Zhou, Y. N.; Lubkowska, L.; Hui, M.; Court, C.; Chen, S.; Court, D. L.; Strathern, J.; Jin, D. J.; Kashlev, M. *J. Biol. Chem.* **2013**, *288*, 2700.
- (36) Kuzmine, I.; Gottlieb, P. A.; Martin, C. T. *Nucleic Acids Res.* **2001**, *29*, 2601.
- (37) Belotserkovskii, B. P.; Liu, R.; Tornaletti, S.; Krasilnikova, M. M.; Mirkin, S. M.; Hanawalt, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12816.
- (38) Tornaletti, S.; Park-Snyder, S.; Hanawalt, P. C. *J. Biol. Chem.* **2008**, *283*, 12756.
- (39) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783.
- (40) Belotserkovskii, B. P.; Neil, A. J.; Saleh, S. S.; Shin, J. H.; Mirkin, S. M.; Hanawalt, P. C. *Nucleic Acids Res.* **2013**, *41*, 1817.
- (41) Grabczyk, E.; Fishman, M. C. *J. Biol. Chem.* **1995**, *270*, 1791.
- (42) Grabczyk, E.; Usdin, K. *Nucleic Acids Res.* **2000**, *28*, 2815.
- (43) Baranello, L.; Levens, D.; Gupta, A.; Kouzine, F. *Biochim. Biophys. Acta* **2012**, *1819*, 632.
- (44) Roca, J. *Chromosoma* **2011**, *120*, 323.
- (45) Vologodskii, A. V. *Topology and Physics of Circular DNA*; CRC Press: Boca Raton, FL, 1992.
- (46) Vologodskii, A. V.; Anshelevich, V. V.; Lukashin, A. V.; Frank-Kamenetskii, M. D. *Nature* **1979**, *280*, 294.
- (47) Vologodskii, A. V.; Levene, S. D.; Klenin, K. V.; Frank-Kamenetskii, M.; Cozzarelli, N. R. *J. Mol. Biol.* **1992**, *227*, 1224.
- (48) Cherny, D. Y.; Belotserkovskii, B. P.; Frank-Kamenetskii, M. D.; Egholm, M.; Buchardt, O.; Berg, R. H.; Nielsen, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 1667.
- (49) Belotserkovskii, B. P.; Liu, R.; Hanawalt, P. C. *Mol. Carcinog.* **2009**, *48*, 299.
- (50) Allemand, J. F.; Bensimon, D.; Lavery, R.; Croquette, V. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14152.
- (51) Liu, L. F.; Wang, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7024.
- (52) Nelson, P. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14342.
- (53) Belotserkovskii, B. P.; Hanawalt, P. C. *Biophys. J.* **2011**, *100*, 675.
- (54) Zhang, Y.; Smith, C. L.; Saha, A.; Grill, S. W.; Mihardja, S.; Smith, S. B.; Cairns, B. R.; Peterson, C. L.; Bustamante, C. *Mol. Cell* **2006**, *24*, 559.
- (55) Strick, T. R.; Quessada-Vial, A. *Nat. Struct. Mol. Biol.* **2006**, *13*, 948.
- (56) Strick, T.; Quessada-Vial, A. *Curr. Biol.* **2006**, *16*, R287.
- (57) Wang, X.; Zhang, X.; Mao, C.; Seeman, N. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12547.
- (58) Leroy, J. L.; Gehring, K.; Kettani, A.; Gueron, M. *Biochemistry* **1993**, *32*, 6019.
- (59) Gehring, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561.
- (60) Panyutin, I. G.; Kovalsky, O. I.; Budowsky, E. I. *J. Biomol. Struct. Dyn.* **1991**, *8*, 967.
- (61) Arimondo, P. B.; Garestier, T.; Helene, C.; Sun, J. S. *Nucleic Acids Res.* **2001**, *29*, E15.
- (62) Lacroix, L.; Mergny, J. L.; Leroy, J. L.; Helene, C. *Biochemistry* **1996**, *35*, 8715.
- (63) Mergny, J. L.; Lacroix, L. *Nucleic Acids Res.* **1998**, *26*, 4797.
- (64) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863.
- (65) Salinas-Rios, V.; Belotserkovskii, B. P.; Hanawalt, P. C. *Nucleic Acids Res.* **2011**, *100*, 675.
- (66) Miller, C. A.; Umek, R. M.; Kowalski, D. *Nucleic Acids Res.* **1999**, *27*, 3921.
- (67) Zhou, W.; Reines, D.; Doetsch, P. W. *Cell* **1995**, *82*, 577.
- (68) Pearson, C. E.; Sinden, R. R. *Curr. Opin. Struct. Biol.* **1998**, *8*, 321.
- (69) Sinden, R. R.; Potaman, V. N.; Oussatcheva, E. A.; Pearson, C. E.; Lyubchenko, Y. L.; Shlyakhtenko, L. S. *J. Biosci.* **2002**, *27*, 53.
- (70) Pearson, C. E.; Sinden, R. R. *Biochemistry* **1996**, *35*, 5041.
- (71) Pearson, C. E.; Wang, Y. H.; Griffith, J. D.; Sinden, R. R. *Nucleic Acids Res.* **1998**, *26*, 816.
- (72) Sinden, R. R.; Pytlos-Sinden, M. J.; Potaman, V. N. *Front. Biosci.* **2007**, *12*, 4788.
- (73) Lin, Y.; Hubert, L., Jr.; Wilson, J. H. *Mol. Carcinog.* **2009**, *48*, 350.
- (74) Lin, Y.; Dion, V.; Wilson, J. H. *Nat. Struct. Mol. Biol.* **2006**, *13*, 179.
- (75) Lin, Y.; Dent, S. Y.; Wilson, J. H.; Wells, R. D.; Napierala, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 692.
- (76) Lin, Y.; Wilson, J. H. *Mol. Cell. Biol.* **2007**, *27*, 6209.
- (77) Parsons, M. A.; Sinden, R. R.; Izban, M. G. *J. Biol. Chem.* **1998**, *273*, 26998.
- (78) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680.
- (79) Herbert, A.; Rich, A. *Genetica* **1999**, *106*, 37.
- (80) Singleton, C. K.; Klysik, J.; Stirdvant, S. M.; Wells, R. D. *Nature* **1982**, *299*, 312.
- (81) Nordheim, A.; Lafer, E. M.; Peck, L. J.; Wang, J. C.; Stollar, B. D.; Rich, A. *Cell* **1982**, *31*, 309.
- (82) Haniford, D. B.; Pulleyblank, D. E. *Nature* **1983**, *302*, 632.
- (83) Johnston, B. H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 153.
- (84) Ha, S. C.; Lowenhaupt, K.; Rich, A.; Kim, Y. G.; Kim, K. K. *Nature* **2005**, *437*, 1183.
- (85) Rich, A.; Zhang, S. *Nat. Rev. Genet.* **2003**, *4*, 566.
- (86) Peck, L. J.; Wang, J. C. *Cell* **1985**, *40*, 129.
- (87) Droege, P.; Pohl, F. M. *Nucleic Acids Res.* **1991**, *19*, 5301.
- (88) Ditlevsen, J. V.; Tornaletti, S.; Belotserkovskii, B. P.; Teijeiro, V.; Wang, G.; Vasquez, K. M.; Hanawalt, P. C. *Nucleic Acids Res.* **2008**, *36*, 3163.
- (89) Durand, R.; Job, C.; Zarling, D. A.; Teissere, M.; Jovin, T. M.; Job, D. *EMBO J.* **1983**, *2*, 1707.
- (90) Frank-Kamenetskii, M. D.; Mirkin, S. M. *Annu. Rev. Biochem.* **1995**, *64*, 65.
- (91) Mirkin, S. M.; Frank-Kamenetskii, M. D. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 541.
- (92) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320.
- (93) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976.
- (94) Jayasena, S. D.; Johnston, B. H. *Nucleic Acids Res.* **1992**, *20*, 5279.
- (95) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800.
- (96) Olivas, W. M.; Maher, L. J., III. *Biochemistry* **1994**, *33*, 983.
- (97) Balatskaya, S. V.; Belotserkovskii, B. P.; Johnston, B. H. *Biochemistry* **1996**, *35*, 13328.
- (98) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360.
- (99) Giovannangeli, C.; Rougee, M.; Garestier, T.; Thuong, N. T.; Helene, C. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8631.
- (100) Malkov, V. A.; Voloshin, O. N.; Veselkov, A. G.; Rostapshov, V. M.; Jansen, I.; Soyfer, V. N.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1993**, *21*, 105.
- (101) Kaluzhny, D. N.; Timoshin, V. V.; Borisova, O. F.; Zhurkin, V. B.; Florentiev, V. L.; Shchyolkina, A. K. *J. Biomol. Struct. Dyn.* **2008**, *26*, 301.

- (102) Shchylkina, A. K.; Kaluzhny, D. N.; Arndt-Jovin, D. J.; Jovin, T. M.; Zhurkin, V. B. *Nucleic Acids Res.* **2006**, *34*, 3239.
- (103) Shchylkina, A. K.; Kaluzhny, D. N.; Borisova, O. F.; Hawkins, M. E.; Jernigan, R. L.; Jovin, T. M.; Arndt-Jovin, D. J.; Zhurkin, V. B. *Nucleic Acids Res.* **2004**, *32*, 432.
- (104) Malkov, V. A.; Panyutin, I. G.; Neumann, R. D.; Zhurkin, V. B.; Camerini-Otero, R. D. *J. Mol. Biol.* **2000**, *299*, 629.
- (105) Dagneaux, C.; Porumb, H.; Liquier, J.; Takahashi, M.; Taillandier, E. *J. Biomol. Struct. Dyn.* **1995**, *13*, 465.
- (106) Dagneaux, C.; Liquier, J.; Taillandier, E. *Biochemistry* **1995**, *34*, 14815.
- (107) Kim, M. G.; Zhurkin, V. B.; Jernigan, R. L.; Camerini-Otero, R. D. *J. Mol. Biol.* **1995**, *247*, 874.
- (108) Zhurkin, V. B.; Raghunathan, G.; Ulyanov, N. B.; Camerini-Otero, R. D.; Jernigan, R. L. *J. Mol. Biol.* **1994**, *239*, 181.
- (109) Van Meervelt, L.; Vlieghe, D.; Dautant, A.; Gallois, B.; Precigoux, G.; Kennard, O. *Nature* **1995**, *374*, 742.
- (110) Vlieghe, D.; Van Meervelt, L.; Dautant, A.; Gallois, B.; Precigoux, G.; Kennard, O. *Science* **1996**, *273*, 1702.
- (111) Vlieghe, D.; Van Meervelt, L.; Dautant, A.; Gallois, B.; Precigoux, G.; Kennard, O. *Acta Crystallogr. Sect. D* **1996**, *52*, 766.
- (112) Conn, G. L.; Draper, D. E.; Lattman, E. E.; Gittis, A. G. *Science* **1999**, *284*, 1171.
- (113) Wimberly, B. T.; Guymon, R.; McCutcheon, J. P.; White, S. W.; Ramakrishnan, V. *Cell* **1999**, *97*, 491.
- (114) Lyamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *J. Biomol. Struct. Dyn.* **1986**, *3*, 667.
- (115) Htun, H.; Dahlberg, J. E. *Science* **1989**, *243*, 1571.
- (116) Kohwi, Y.; Kohwi-Shigematsu, T. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 3781.
- (117) Bernues, J.; Beltran, R.; Casasnovas, J. M.; Azorin, F. *EMBO J.* **1989**, *8*, 2087.
- (118) Panyutin, I. G.; Kovalsky, O. I.; Budowsky, E. I. *Nucleic Acids Res.* **1989**, *17*, 8257.
- (119) Kohwi-Shigematsu, T.; Kohwi, Y. *Nucleic Acids Res.* **1991**, *19*, 4267.
- (120) Panyutin, I. G.; Wells, R. D. *J. Biol. Chem.* **1992**, *267*, 5495.
- (121) Christophe, D.; Cabrer, B.; Bacolla, A.; Targovnik, H.; Pohl, V.; Vassart, G. *Nucleic Acids Res.* **1985**, *13*, 5127.
- (122) Lee, J. S.; Ashley, C.; Hampel, K. J.; Bradley, R.; Scraba, D. G. *J. Mol. Biol.* **1995**, *252*, 283.
- (123) Sakamoto, N.; Chastain, P. D.; Parniewski, P.; Ohshima, K.; Pandolfo, M.; Griffith, J. D.; Wells, R. D. *Mol. Cell* **1999**, *3*, 465.
- (124) Vetcher, A. A.; Napierala, M.; Iyer, R. R.; Chastain, P. D.; Griffith, J. D.; Wells, R. D. *J. Biol. Chem.* **2002**, *277*, 39217.
- (125) Vetcher, A. A.; Wells, R. D. *J. Biol. Chem.* **2004**, *279*, 6434.
- (126) Rando, R. F.; DePaolis, L.; Durland, R. H.; Jayaraman, K.; Kessler, D. J.; Hogan, M. E. *Nucleic Acids Res.* **1994**, *22*, 678.
- (127) Giovannangeli, C.; Perrouault, L.; Escude, C.; Gryaznov, S.; Helene, C. *J. Mol. Biol.* **1996**, *261*, 386.
- (128) Samadashwily, G. M.; Dayn, A.; Mirkin, S. M. *EMBO J.* **1993**, *12*, 4975.
- (129) Campuzano, V.; Montermini, L.; Molto, M. D.; Pianese, L.; Cossee, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A.; Zara, F.; Canizares, J.; Koutnikova, H.; Bidichandani, S. I.; Gellera, C.; Brice, A.; Trouillas, P.; De Michele, G.; Filla, A.; De Frutos, R.; Palau, F.; Patel, P. I.; Di Donato, S.; Mandel, J. L.; Cocozza, S.; Koenig, M.; Pandolfo, M. *Science* **1996**, *271*, 1423.
- (130) Grabczyk, E.; Kumari, D.; Usdin, K. *Brain Res. Bull.* **2001**, *56*, 367.
- (131) Wells, R. D. *FASEB J.* **2008**, *22*, 1625.
- (132) Bidichandani, S. I.; Ashizawa, T.; Patel, P. I. *Am. J. Hum. Genet.* **1998**, *62*, 111.
- (133) Krasilnikova, M. M.; Kireeva, M. L.; Petrovic, V.; Knijnikova, N.; Kashlev, M.; Mirkin, S. M. *Nucleic Acids Res.* **2007**, *35*, 1075.
- (134) Sakamoto, N.; Ohshima, K.; Montermini, L.; Pandolfo, M.; Wells, R. D. *J. Biol. Chem.* **2001**, *276*, 27171.
- (135) Grabczyk, E.; Usdin, K. *Nucleic Acids Res.* **2000**, *28*, 4930.
- (136) Burnett, R.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11497.
- (137) Marcu, K. B.; Bossone, S. A.; Patel, A. J. *Annu. Rev. Biochem.* **1992**, *61*, 809.
- (138) Martjanov, I.; Ramadass, A.; Serra Barros, A.; Chow, N.; Akoulitchev, A. *Nature* **2007**, *445*, 666.
- (139) Wang, G.; Vasquez, K. M. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 13448.
- (140) Masters, J. N.; Attardi, G. *Mol. Cell. Biol.* **1985**, *5*, 493.
- (141) Eick, D.; Polack, A.; Kofler, E.; Lenoir, G. M.; Rickinson, A. B.; Bornkamm, G. W. *Oncogene* **1990**, *5*, 1397.
- (142) Blume, S. W.; Meng, Z.; Shrestha, K.; Snyder, R. C.; Emanuel, P. D. *J. Cell Biochem.* **2003**, *88*, 165.
- (143) Belotserkovskii, B. P.; De Silva, E.; Tornaletti, S.; Wang, G.; Vasquez, K. M.; Hanawalt, P. C. *J. Biol. Chem.* **2007**, *282*, 32433.
- (144) Schmitz, K. M.; Mayer, C.; Postepska, A.; Grummt, I. *Genes Dev.* **2010**, *24*, 2264.
- (145) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463.
- (146) Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Helene, C. *Nucleic Acids Res.* **1993**, *21*, 5547.
- (147) Semerad, C. L.; Maher, L. J., 3rd *Nucleic Acids Res.* **1994**, *22*, 5321.
- (148) Karamychev, V. N.; Panyutin, I. G.; Neumann, R. D.; Zhurkin, V. B. *J. Biomol. Struct. Dyn.* **2000**, *17*, 155.
- (149) Reaban, M. E.; Lebowitz, J.; Griffin, J. A. *J. Biol. Chem.* **1994**, *269*, 21850.
- (150) Gros, J.; Webba da Silva, M.; De Cian, A.; Amrane, S.; Rosu, F.; Bourdoncle, A.; Sacca, B.; Alberti, P.; Lacroix, L.; Mergny, J. L. *Nucleic Acids Symp. Ser.* **2005**, *61*.
- (151) Huppert, J. L. *FEBS J.* **2010**, *277*, 3452.
- (152) Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat. Rev. Genet.* **2012**, *13*, 770.
- (153) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013.
- (154) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364.
- (155) Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825.
- (156) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871.
- (157) Panyutin, I. G.; Kovalsky, O. I.; Budowsky, E. I.; Dickerson, R. E.; Rikhirev, M. E.; Lipanov, A. A. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 867.
- (158) Maizels, N. *Ann. N.Y. Acad. Sci.* **2012**, *1267*, 53.
- (159) Matsugami, A.; Ouhashi, K.; Kanagawa, M.; Liu, H.; Kanagawa, S.; Uesugi, S.; Katahira, M. *J. Mol. Biol.* **2001**, *313*, 255.
- (160) Lyamichev, V. I.; Mirkin, S. M.; Danilevskaya, O. N.; Voloshin, O. N.; Balatskaya, S. V.; Dobrynin, V. N.; Filippov, S. A.; Frank-Kamenetskii, M. D. *Nature* **1989**, *339*, 634.
- (161) Onyshchenko, M. I.; Gaynutdinov, T. I.; Englund, E. A.; Appella, D. H.; Neumann, R. D.; Panyutin, I. G. *Nucleic Acids Res.* **2009**, *37*, 7570.
- (162) Voloshin, O. N.; Veselkov, A. G.; Belotserkovskii, B. P.; Danilevskaya, O. N.; Pavlova, M. N.; Dobrynin, V. N.; Frank-Kamenetskii, M. D. *J. Biomol. Struct. Dyn.* **1992**, *9*, 643.
- (163) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nature Chem.* **2013**, *5*, 182.
- (164) Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Biochimie* **2009**, *91*, 1104.
- (165) Zheng, K. W.; Chen, Z.; Hao, Y. H.; Tan, Z. *Nucleic Acids Res.* **2009**, *38*, 327.
- (166) Sissi, C.; Gatto, B.; Palumbo, M. *Biochimie* **2011**, *93*, 1219.
- (167) Wanrooij, P. H.; Uhler, J. P.; Simonsson, T.; Falkenberg, M.; Gustafsson, C. M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16072.
- (168) Biffi, G.; Tannahill, D.; Balasubramanian, S. *J. Am. Chem. Soc.* **2012**, *134*, 11974.
- (169) Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N. *Genes Dev.* **2004**, *18*, 1618.
- (170) Duquette, M. L.; Pham, P.; Goodman, M. F.; Maizels, N. *Oncogene* **2005**, *24*, 5791.

- (171) Eddy, J.; Vallur, A. C.; Varma, S.; Liu, H.; Reinhold, W. C.; Pommier, Y.; Maizels, N. *Nucleic Acids Res.* **2011**, *39*, 4975.
- (172) Broxson, C.; Beckett, J.; Tornaletti, S. *Biochemistry* **2011**, *50*, 4162.
- (173) Aguilera, A.; Garcia-Muse, T. *Mol. Cell* **2012**, *46*, 115.
- (174) Skourtis-Stathaki, K.; Proudfoot, N. J.; Gromak, N. *Mol. Cell* **2011**, *42*, 794.
- (175) Alzu, A.; Bermejo, R.; Begnis, M.; Lucca, C.; Piccini, D.; Carotenuto, W.; Saponaro, M.; Brambati, A.; Cocito, A.; Foiani, M.; Liberi, G. *Cell* **2012**, *151*, 835.
- (176) Yuce, O.; West, S. C. *Mol. Cell. Biol.* **2013**, *33*, 406.
- (177) Chon, H.; Sparks, J. L.; Rychlik, M.; Nowotny, M.; Burgers, P. M.; Crouch, R. J.; Cerritelli, S. M. *Nucleic Acids Res.* **2013**, *41*, 3130.
- (178) Daniels, G. A.; Lieber, M. R. *Nucleic Acids Res.* **1995**, *23*, 5006.
- (179) Yu, K.; Chedin, F.; Hsieh, C. L.; Wilson, T. E.; Lieber, M. R. *Nat. Immunol.* **2003**, *4*, 442.
- (180) Xu, B.; Clayton, D. A. *Mol. Cell. Biol.* **1995**, *15*, 580.
- (181) Roy, D.; Yu, K.; Lieber, M. R. *Mol. Cell. Biol.* **2008**, *28*, 50.
- (182) Roy, D.; Zhang, Z.; Lu, Z.; Hsieh, C. L.; Lieber, M. R. *Mol. Cell. Biol.* **2010**, *30*, 146.
- (183) Reddy, K.; Tam, M.; Bowater, R. P.; Barber, M.; Tomlinson, M.; Nichol Edamura, K.; Wang, Y. H.; Pearson, C. E. *Nucleic Acids Res.* **2011**, *39*, 1749.
- (184) Drolet, M. *Mol. Microbiol.* **2006**, *59*, 723.
- (185) Roy, D.; Lieber, M. R. *Mol. Cell. Biol.* **2009**, *29*, 3124.
- (186) Grabczyk, E.; Mancuso, M.; Sammarco, M. C. *Nucleic Acids Res.* **2007**, *35*, 5351.
- (187) Lesnik, E. A.; Freier, S. M. *Biochemistry* **1995**, *34*, 10807.
- (188) Gowher, H.; Brick, K.; Camerini-Otero, R. D.; Felsenfeld, G. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *109*, 2370.
- (189) Tous, C.; Aguilera, A. *Biochem. Biophys. Res. Commun.* **2007**, *360*, 428.
- (190) Aguilera, A. *EMBO J.* **2002**, *21*, 195.
- (191) Kim, N.; Jinks-Robertson, S. *Nat. Rev. Genet.* **2012**, *13*, 204.
- (192) Gaillard, H.; Herrera-Moyano, E.; Aguilera, A. *Chem. Rev.* **2013**, DOI: 10.1021/cr400017y.
- (193) Mirkin, E. V.; Mirkin, S. M. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 13.
- (194) Krasilnikova, M. M.; Samadashwily, G. M.; Krasilnikov, A. S.; Mirkin, S. M. *EMBO J.* **1998**, *17*, 5095.
- (195) Hanawalt, P. C. *Oncogene* **2002**, *21*, 8949.
- (196) Hanawalt, P. C.; Spivak, G. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 958.
- (197) Gaillard, H.; Aguilera, A. *Biochim. Biophys. Acta* **2013**, *1829*, 141.
- (198) Tornaletti, S. *Cell. Mol. Life Sci.* **2009**, *66*, 1010.
- (199) Hanawalt, P. C. *Science* **1994**, *266*, 1957.
- (200) Wang, G.; Seidman, M. M.; Glazer, P. M. *Science* **1996**, *271*, 802.
- (201) Wang, G.; Christensen, L. A.; Vasquez, K. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 2677.
- (202) Studitsky, V. M.; Walter, W.; Kireeva, M.; Kashlev, M.; Felsenfeld, G. *Trends Biochem. Sci.* **2004**, *29*, 127.
- (203) van Holde, K.; Zlatanova, J. *Bioessays* **1994**, *16*, 59.
- (204) Wong, B.; Chen, S.; Kwon, J. A.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2229.
- (205) Liu, H.; Mulholland, N.; Fu, H.; Zhao, K. *Mol. Cell. Biol.* **2006**, *26*, 2550.
- (206) Herbert, A.; Rich, A. *J. Biol. Chem.* **1996**, *271*, 11595.
- (207) Jeon, Y.; Lee, J. T. *Cell* **2011**, *146*, 119.
- (208) Lee, J. T. *Science* **2012**, *338*, 1435.
- (209) Gomez-Gonzalez, B.; Garcia-Rubio, M.; Bermejo, R.; Gaillard, H.; Shirahige, K.; Marin, A.; Foiani, M.; Aguilera, A. *EMBO J.* **2011**, *30*, 3106.
- (210) Reaban, M. E.; Griffin, J. A. *Nature* **1990**, *348*, 342.
- (211) SantaLucia, J., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460.
- (212) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1999**, *38*, 3468.
- (213) Anshelevich, V. V.; Vologodskii, A. V.; Lukashin, A. V.; Frank-Kamenetskii, M. D. *Biopolymers* **1984**, *23*, 39.
- (214) Radding, C. M.; Beattie, K. L.; Holloman, W. K.; Wiegand, R. C. *J. Mol. Biol.* **1977**, *116*, 825.