

The role of G-quadruplex/i-motif secondary structures as *cis*-acting regulatory elements*

Samantha Kendrick¹ and Laurence H. Hurley^{1,2,3,‡}

¹Arizona Cancer Center, University of Arizona, Tucson, AZ 85724, USA; ²College of Pharmacy, University of Arizona, Tucson, AZ 85721, USA; ³BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA

Abstract: The nature of DNA has captivated scientists for more than 50 years. The discovery of the double-helix model of DNA by Watson and Crick in 1953 not only established the primary structure of DNA, but also provided the mechanism behind DNA function. Since then, researchers have continued to further the understanding of DNA structure and its pivotal role in transcription. The demonstration of DNA secondary structure formation has allowed for the proposal that the dynamics of DNA itself can function to modulate transcription. This review presents evidence that DNA can exist in a dynamic equilibrium between duplex and secondary conformations. In addition, data demonstrating that intracellular proteins as well as small molecules can shift this equilibrium in either direction to alter gene transcription will be discussed, with a focus on the modulation of proto-oncogene expression.

Keywords: *c-myc*; G-quadruplex; i-motif; supercoiling; transcription.

INTRODUCTION

For over half a century, DNA has been widely recognized for its central biological role in heredity, but how does DNA accomplish this intricate function so vital for life? The discovery of the molecular double-helix model of DNA by Watson and Crick provided an explanation for the mode of heredity [1]. The Watson–Crick model established that DNA exists in a form, termed “B-form DNA,” where two complementary strands of DNA consist of hydrogen-bonded nucleotide base-pairs: adenine with thymine and guanine with cytosine [1]. Thus, the complementarity of the strands and the distinct base sequence allow for the synthesis and reconstruction of new, identical DNA double helices. In addition, the specific nucleotide sequence serves as a template for transcription that ultimately enables the production of peptides required for cellular function. While the general acceptance is that the majority of DNA within cells consists in the B-form, DNA also has the potential to exist in other conformations.

The continued interest in the structure of DNA and advances in molecular structural techniques have uncovered that the structure of DNA is not limited to the B-form conformation, but extends to both non-B-form and non-helical secondary structures. Alternative helical structures such as the A- and Z-forms of DNA have been shown to exist under different experimental conditions [2–5]. While the presence of these DNA forms within cells remains controversial, the formation of Z-form, or left-handed, DNA has been observed under conditions of negative supercoiling [6]. Interestingly, this is not the only non-B-DNA structure to form in response to negative supercoiling, indicating the important in-

*Paper based on a presentation on the “Chemistry for Health” theme at the 42nd IUPAC Congress, 2–7 August 2009, Glasgow, UK. Other presentations are published in this issue, pp. 1569–1717.

‡Corresponding author: Tel.: (520) 626-5622; Fax: (520) 626-4824; E-mail: hurley@pharmacy.arizona.edu

fluence of topology on DNA conformation. The formation of unconventional DNA secondary structures, including cruciforms [7], H-DNA (triplexes) [8], G-quadruplexes, and i-motifs [9], has demonstrated that negative supercoiling may induce and stabilize less energetically favorable structures. But how can supercoiling alter DNA structure?

The transition of DNA from the typical B-form to a non-B-form requires the local unwinding of the double helix, which occurs to relieve the stress of negative supercoiling [10]. In nature, negative supercoiling can result from nuclear processes such as transcription, replication, recombination, and repair and produce significant torsional stress [11,12]. Specifically, the transient high degree of supercoiling induced during transcription occurs as the RNA polymerase (RNAP) complex moves along the transcriptionally active segment of DNA (Fig. 1). In order to accommodate the RNAP and the nascent RNA, DNA downstream of the transcription complex becomes positively supercoiled, or overwound, whereas DNA upstream is negatively supercoiled, or underwound [13]. Although DNA topoisomerases have been implicated as the predominant mechanism for relieving DNA torsional stress, these enzymes may inadequately alleviate this stress in response to the significant demand produced by transcription [14–16]. Rather, local melting of double-stranded DNA upstream of transcriptionally active promoters has been shown to be a consequence of torsional stress [10,17]. This local unwinding results in an open, single-stranded DNA conformation that presumably is sufficient for the formation of DNA secondary structures.

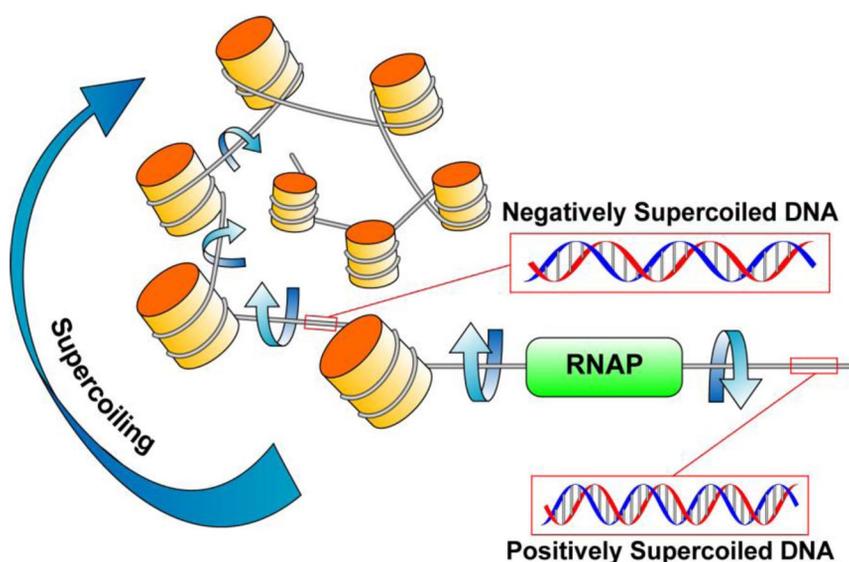


Fig. 1 Model of the transcriptionally induced supercoiling that occurs upstream and downstream of the RNAP. Adapted from Kouzine and Levens [12].

So if DNA supercoiling can induce the formation of a variety of non-B-DNA secondary structures, then what determines whether the DNA will adopt either a cruciform or a G-quadruplex conformation, for example? In addition to topological constraints, these DNA secondary structures exhibit sequence dependency. For example, cruciform structures arise from palindromic sequences within double-stranded DNA [18]. Sequences with contiguous guanine or cytosine runs have the potential to form G-quadruplex or i-motif secondary structures, respectively [19–22]. Specifically, the G-quadruplex building block, the G-tetrad, is formed from a sequence consisting of at least four contiguous runs of two or more guanines (Fig. 2A). The four G-tetrad guanines interact through Hoogsteen base-pairing, and the tetrad is stabilized by the presence of a monovalent cation, as shown in Fig. 2B [19]. The

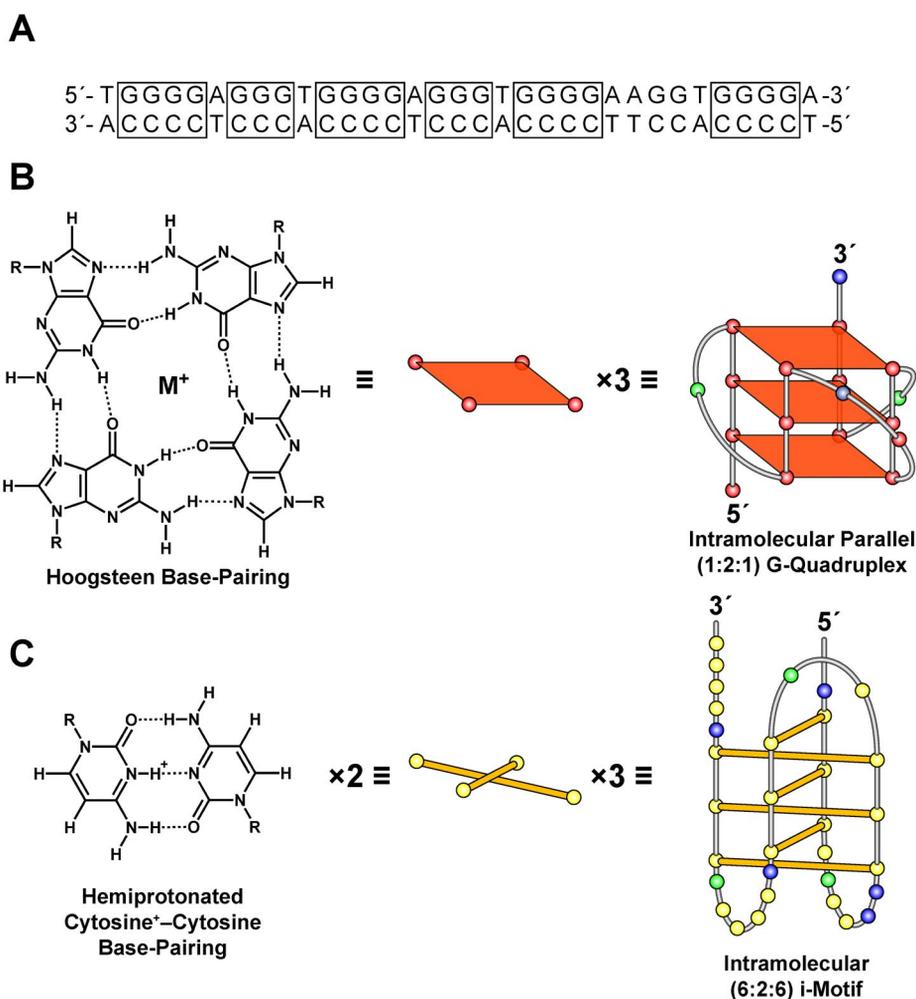


Fig. 2 The *c-myc* GC-rich promoter sequence capable of adopting G-quadruplex and i-motif structures is provided in panel A. The building blocks of these secondary structures are guanine–guanine and cytosine⁺–cytosine base pairings that give rise to the G-quadruplex (B) and i-motif (C) secondary structures. The previously proposed *c-myc* [9] G-quadruplex and i-motif structures formed under conditions of negative supercoiling serve as examples, with the yellow, green, red, and blue circles (online version) representing the nucleobases cytosine, adenine, guanine, and thymine, respectively.

i-motif structure specified by the complementary C-rich sequence comprises two parallel duplexes with intercalated hemiprotonated cytosine⁺–cytosine base pairs (Fig. 2C) [20–22].

The dynamic topology of DNA then poses the question, what is the biological significance of DNA interconverting between duplex and secondary conformations? We, along with many others, have postulated that the ability of duplex DNA to adopt G-quadruplex and i-motif secondary conformations serves as a molecular switch to modulate transcription [23]. If the G-quadruplex and i-motif structures play a role in transcriptional regulation, then, similar to other *cis*-regulatory elements, these secondary structure-forming sequences must be in close proximity to the transcriptional start site (TSS). In addition, for these structures to be effective transcription modulators they would also need to compete with duplex DNA, be recognized by proteins for remodeling, be targeted by small molecules to induce changes in transcription, and exhibit diversity and complexity among promoter regions.

This review will describe the prevalence, diversity, and complexity of G-quadruplex and i-motif structures as well as observed protein and small-molecule interactions to provide evidence for the role of DNA secondary structures in transcriptional regulation. While genome-wide investigations into the occurrences of G-quadruplex-forming sequences will be discussed, emphasis will be placed on the demonstration of stable secondary structure formation within proto-oncogene promoter regions.

G-QUADRUPLEX- AND I-MOTIF-FORMING SEQUENCES EXTENSIVELY FOUND IN CLOSE PROXIMITY TO TRANSCRIPTIONAL START SITES

If the G-quadruplex and the i-motif play a role in transcriptional regulation, then one would expect to find sequences capable of forming these structures near the TSS within the promoter regions of various genes. Indeed, recent genomic analysis has revealed that of promoter sequences within 1 kb upstream of the TSS, 43 % of genes have the potential to form at least one G-quadruplex structure [24]. In consideration with genomic wide-analyses, which have detected G-quadruplex-forming sequences throughout the genome at ~376 000 sites [25,26], the extensive incidence within close proximity to the TSS demonstrates an enrichment for promoter G-quadruplexes. This enrichment has been postulated to reflect an evolutionary selection for the location of secondary structures and thereby supports the biological role of these structures in transcription [24]. Consistent with quadruplex elements functioning as transcription regulatory units, conservation of potential G-quadruplex-forming sequences near the TSS has been observed across related species. A comparative genome-wide analysis among humans, chimpanzees, mice, and rats not only demonstrated at least a 40 % prevalence of promoter regions (within ± 1 kb of the TSS) to potentially adopt G-quadruplex conformations, but that over 700 orthologous promoter regions consisted of conserved DNA secondary structure-forming sequences [27]. While the probing of the human genome for potential sites of G-quadruplex formation provides information on the prevalence and localization of these sequences, these studies do not demonstrate the precise capability of the identified regions to form stable G-quadruplex structures. The defined sequence utilized in these searches consisted of four runs of at least three guanines spaced by one to seven bases ($G_{3+}N_{1-7}G_{3+} N_{1-7}G_{3+} N_{1-7}G_{3+}$). To date, the majority of stable G-quadruplex conformations favor smaller loop sizes [28–31]; however, the loop N_{1-7} search parameters may allow for identification of potential i-motif-forming sequences on the complementary strand since stable i-motif structures tend to exhibit larger loop sizes [9,32]. Although genome investigations have focused on the detection of G-quadruplex motifs, presumably there is an equal prevalence of potential i-motif-forming sequences in the complementary strand. If there is a significant incidence of these DNA secondary structure-forming sequences within promoter regions, then which types of genes contain these elements? Furthermore, can these potential sequences actually form stable structures?

A review of the gene ontology codes of the promoters that were found to contain G-quadruplex motifs revealed a significant occurrence within genes involved in transcription factor and kinase activity, development, and neurogenesis [24,33,34]. Conversely, olfaction, G-protein signaling, immune response, nucleic acid binding, and protein biosynthesis-related genes were significantly less likely to contain G-quadruplex motifs [24,33,34]. In addition, significantly more proto-oncogenes (69 %) were shown to consist of putative G-quadruplex motifs than tumor suppressors [33,34]. Consistent with this prediction, GC-rich elements within several proto-oncogenes and a tumor suppressor gene have been shown to adopt stable G-quadruplex and i-motif conformations. The *c-myc* proto-oncogene promoter has been extensively studied and has led to the identification of G-quadruplex motifs within the *VEGF* [35,36], *c-kit* [37–40], *c-myb* [41], *PDGF-A* [42], *RET* [43], *KRAS* [44], *hif-1 α* [45], *bcl-2* [46–49], *hTERT* [50], and *Rb* [51] promoter regions. The wild-type GC-rich element located –142 to –115 upstream of the *c-myc* P1 promoter has the potential to form an intramolecular parallel G-quadruplex structures with loop sizes of 1:2:1 [9,52–57] and an intramolecular i-motif structure [58], which, under conditions of negative supercoiling, adopts a 6:2:6 loop conformation [9]. To date, C-rich elements within the *VEGF* [36], *RET* [43], *bcl-2* [59], and *Rb* [51] oncogene promoter regions have also been

demonstrated to form i-motif structures. The occurrence of DNA secondary structure motifs within oncogene promoter regions in close proximity to the TSS suggests that these alternative DNA conformations serve a regulatory function in transcription that requires resolution of duplex DNA.

DNA SECONDARY STRUCTURES THAT COMPETE WITH DUPLEX DNA

Cellular DNA is mostly present in the B-form, which seemingly presents an obstacle to secondary structure formation that requires single-stranded DNA. For secondary structures to form within B-DNA, local unwinding or melting of the duplex must occur to provide regions of single-stranded DNA [9] and thus create an equilibrium in which there is active transition between duplex and single-stranded DNA (Fig. 3). If secondary structure conformations efficiently compete with duplex DNA, then there would be a favorable shift in the equilibrium toward G-quadruplex/i-motif formation. So the question remains, can DNA secondary conformations efficiently compete with duplex DNA? Various investigations into this issue have provided evidence of an equilibrium between duplex and secondary DNA conformations that is context- or environment-dependent. For example, G- and C-rich strands of the telomeric sequence were demonstrated to efficiently compete with duplex DNA under conditions that facilitate the

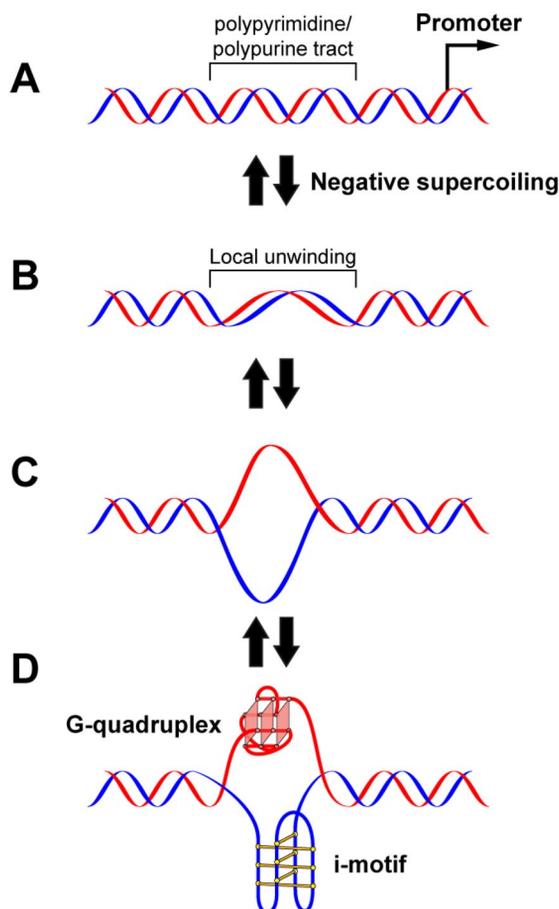


Fig. 3 The proposed model for the dynamic equilibrium of DNA topology induced by negative supercoiling. Negative supercoiling within duplex DNA (A) induces the local unwinding (B), which facilitates the transition from duplex DNA (A) to single-stranded (C) to G-quadruplex and i-motif DNA secondary formation (D).

formation of the G-quadruplex and/or the i-motif ($\text{pH} \leq 4.9$ or high temperatures of 50–64 °C) [60]. Within the pH range of 5.3–7.0, high temperatures (64–70 °C) result in the melting of the duplex DNA and conversion to the G-quadruplex. The unfolding of the G-quadruplex and consequent transition back to duplex DNA is favored at temperatures below 64 °C. In contrast, when pH is less than 5.0, low to moderate temperatures (below 55 °C) favor G-quadruplex formation over duplex [60]. Similarly, Li et al. [61] and Kumar et al. [62] have also confirmed the dependency of lower pH and specific cation species on the efficiency of DNA secondary structure and duplex competition in the telomeric or thrombin-binding aptamer sequences, respectively. However, a closer examination of the unfolding and folding rates of the telomeric [63] and *c-myc* [64] G-quadruplexes in the presence of duplex DNA demonstrated that at low, equivalent molar strand concentrations, duplex conformation is significantly less favorable than G-quadruplex formation, even at neutral pH. While the low, equivalent molar strand concentration achieves more physiological conditions compared to previous studies, these studies did not take into consideration the G-quadruplex/i-motif-flanking sequences or negative supercoiling that would normally be present within the genome.

As mentioned previously, transcriptionally active genes generate both negative and positive supercoiled sections of DNA, with the former acting as the driving force for unwinding DNA. Recent studies demonstrated a significant increase in negative supercoiling with a concurrent increase in transcriptional activity [10,17]. More importantly, melting of a duplex region 1.2 kb upstream of the inducible *c-myc* promoter resulted from the negative supercoiling stress [17]. This may be sufficient to provide single-stranded DNA within the GC-rich element capable of adopting G-quadruplex/i-motif conformations (Fig. 3). Recently, our group utilized the inherent negative supercoiling within plasmid DNA as a model system to test this hypothesis for the *c-myc* GC-rich promoter region. Indeed, the stress induced from negative supercoiling is sufficient to cause unwinding of duplex DNA and allow for G-quadruplex and i-motif formation [9]. Unexpectedly, the i-motif conformation not only formed under neutral pH, but also remained present regardless of G-quadruplex formation [9]. In contrast, single-stranded C-rich DNA requires the pH-dependent protonation of cytosines for i-motif formation [58]. Therefore, studies using plasmid DNA provide a more physiologically relevant system in which to observe the dynamics of DNA topology as consequences of negative supercoiling with respect to the formation of DNA secondary structures. It is important to note that flanking sequence length alone is not sufficient to increase G-quadruplex/i-motif competition with duplex DNA, as was shown within the *c-kit* promoter region [65], further demonstrating the importance of the in vivo dynamics of DNA topology resulting from negative supercoiling. In addition, intracellular proteins may promote and stabilize the G-quadruplex/i-motif conformation and shift the equilibrium away from duplex.

PROTEIN RECOGNITION AND REMODELING OF G-QUADRUPLEX-FORMING REGIONS

Protein recognition of G-quadruplex structures with consequent biological effect provides further evidence for the in vivo existence of DNA secondary structures and their role as *cis*-regulatory elements. Several proteins, including nucleases and helicases, from different organisms have been identified to bind selectively and with high affinity to G-quadruplex structures [66]. These proteins may promote the cleavage, folding/unfolding, or stabilization/destabilization of G-quadruplex structures [66]. More recently, proteins within the heterogeneous ribonucleoprotein (hnRNP) and NM23 families were shown to facilitate the unfolding of the *KRAS* [66] and *c-myc* [68,69] G-quadruplexes, respectively, while nucleolin confers stabilization on the *c-myc* G-quadruplex [70]. Initial affinity column studies using pancreatic nuclear extract and the *KRAS* G-quadruplex structure revealed three proteins, poly[ADP-ribose] polymerase 1 (PARP-1); ATP-dependent DNA helicase 2, subunit 1 (Ku70); and hnRNP A1, that bound specifically to DNA in the secondary conformation [71]. Further analysis of hnRNP A1 and its derivative Up1 demonstrated not only recognition of the *KRAS* G-quadruplex, but also a destabilization of the structure that resulted in transition to duplex DNA [67]. Likewise, NM23-H2 facilitates the unfolding

of the *c-myc* G-quadruplex and sequesters the G-rich strand in single-stranded form [68,69]. The expression of NM23-H2 was demonstrated to be inversely correlated with *c-myc* expression in CaCo-2, HT29, and HCT116 colon cancer cell lines. Furthermore, *c-myc* expression was significantly increased subsequent to knock-down of NM23-H2 in HeLa cells [68]. Conversely, nucleolin substantially decreases *c-myc* transcription through recognition and stabilization of the G-quadruplex [70]. This study also demonstrated that the recognition and binding of nucleolin was highly preferential to the *c-myc* G-quadruplex and, to a lesser extent, the *VEGF*, *RET*, and *PDGF-A* structures. Nucleolin displayed low affinity for other G-quadruplexes, such as those formed in the *bcl-2* and *c-myc* promoters, indicating that nucleolin may specifically recognize parallel G-quadruplex conformations [70]. Collectively, these findings establish protein involvement in the remodeling of DNA topology that allows for the switch from G-quadruplex/i-motif to single- or double-stranded conformations (Fig. 4). In the instance of *c-myc*, the presence of NM23-H2 facilitates unwinding of the G-quadruplex, allowing for transition to duplex DNA, and double-stranded binding transcription factors, such as Sp-1, can bind and initiate transcription (Fig. 4). Alternatively, transcription becomes repressed in the presence of nucleolin, which stabilizes the G-quadruplex and shifts the equilibrium toward secondary structure formation (Fig. 4). This remains consistent with the ability of DNA secondary structures to actively participate in regulation of transcription. As key players in the transcriptional process, targeting of DNA secondary structures with small molecules should also modulate transcription.

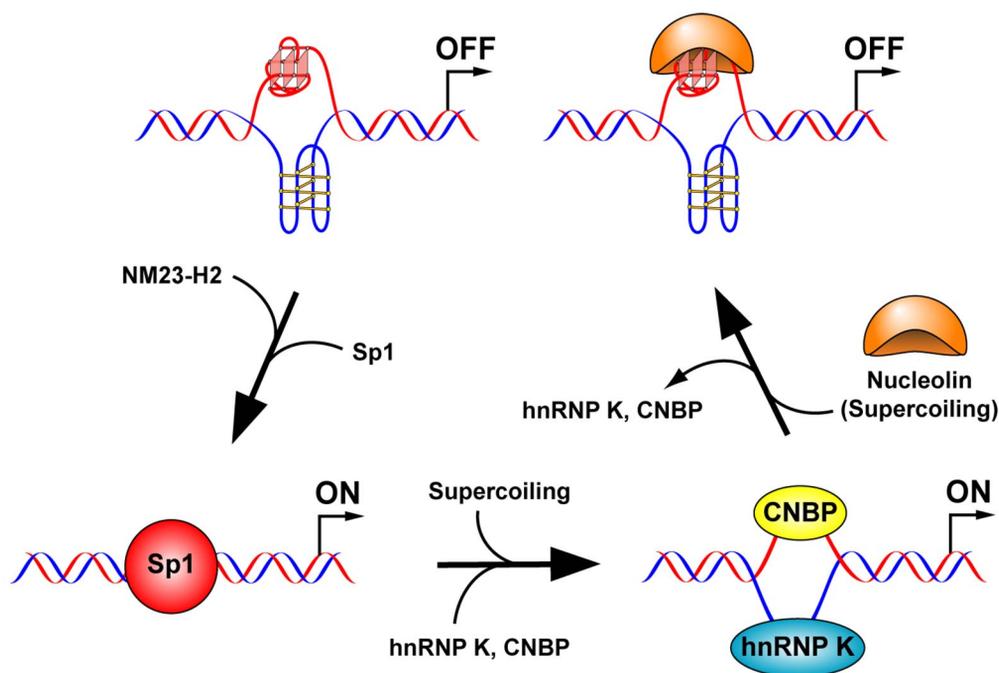


Fig. 4 A proposed model of *c-myc* transcriptional regulation that involves the resolution of the G-quadruplex by NM23-H2 for duplex DNA formation and subsequent transcriptional activation by Sp1. The binding of hnRNP K and CNBP to single-stranded DNA induced by negative supercoiling also leads to *c-myc* transcription activation. The stabilization of the G-quadruplex by nucleolin results in negative regulation of *c-myc* transcription.

SMALL MOLECULES THAT BIND TO G-QUADRUPLEXES AND MODULATE TRANSCRIPTION

The stabilization of DNA secondary structures using small molecules has been well established for both telomeric [72] and *c-myc* [53,73–75] G-quadruplexes and has led to studies investigating the biological consequences of targeting DNA secondary structures. TMPyP4, a cationic porphyrin known to be a selective G-quadruplex stabilizer [76], stabilizes the *c-myc* G-quadruplex in a dose-dependent manner and inhibits *c-myc* promoter transcriptional activity in the Ramos Burkitt's lymphoma cell line, but not in the CA46 cell line, which lacks one of the alleles of the GC-rich promoter element [53]. Similarly, in a *c-kit* gastrointestinal stromal tumor cells model, Gunaratnam et al. have demonstrated that a naphthalene diimide derivative stabilizes the *c-kit* G-quadruplex structures and significantly decreases *c-kit* transcription [77]. In contrast, a destabilizing G-quadruplex ligand would presumably result in an increase in gene transcription. Accordingly, triarylpyridine ligands have been shown to resolve the *c-kit* G-quadruplexes and increase *c-kit* gene expression in the HGC-27 gastric carcinoma cell line [78]. So how does the stabilization/destabilization of DNA in the secondary structure conformation alter gene expression?

Perhaps the formation of DNA secondary structures within promoter regions acts as steric block to the transcriptional machinery. However, for DNA secondary structures to actively participate in transcriptional regulation, either as silencer or activator elements, G-quadruplex/i-motif-forming sequences must localize in close proximity to the TSS (as previously discussed) and also occur in regions of transcription factor and protein recognition sites. The formation of DNA secondary structures would then “mask” these binding sites and prevent their function in transcription activation or repression. It is also likely that there are proteins that recognize the structures themselves to achieve similar biological effects. Recent investigations are beginning to elucidate the mechanism by which G-quadruplex/i-motif elements act as switches to modulate transcription and support this hypothesis. The stabilization of the *KRAS* and *c-myc* G-quadruplexes through the use of TMPyP4 hindered the ability of the previously mentioned proteins, hnRNP A1 [67] and NM23-H2 [68,69], to unwind the secondary structures. In these examples, small-molecule targeting of DNA secondary structures disrupts DNA–protein interactions and results in promoter silencing, suggesting that these particular DNA secondary structures act to negatively regulate their respective promoters.

It is equally likely, however, that in different promoters DNA secondary structures may act as positive regulators. For example, we postulate that the formation of DNA secondary structures within the *bcl-2* promoter region results in displacement of WT-1, a known repressor of *bcl-2* transcription [79], and activates transcription. While studies involving the stabilization of the *bcl-2* G-quadruplex with TMPyP4 and other small molecules have been performed [46,80,81], the effects on transcription have not been assessed. It is important to note that similar molecules, if not the same molecules, have been used to target multiple G-quadruplex structures; however, specificity in molecular targeting will require taking advantage of structural differences among DNA secondary structures. Recent development of a G-quadruplex-specific antibody, hf2, has demonstrated not only selectivity for the *c-kit2* G-quadruplex structure over other G-quadruplexes, but also a 3000-fold lower affinity for duplex DNA [82]. Furthermore, interactions of these G-quadruplex-specific antibodies with promoter G-quadruplex-forming sequences throughout the genome have been shown to either increase or decrease transcription of the adjacent gene [83].

DIVERSITY AND COMPLEXITY IN G-QUADRUPLEX AND i-MOTIF STRUCTURES

As with any *cis*-regulatory element or targetable group of molecules, conserved features essential to structure and overall function must exist in the presence of diversity and complexity to allow for molecular exploitation. DNA secondary structures exhibit conservation in the core sequences that gives rise to either G-tetrads [28,36] or cytosine⁺–cytosine hemiprotonated base pairs [36] for G-quadruplex or

i-motif assembly, respectively. In addition, certain features within the loop regions are also conserved to accommodate specific structural features in the G-quadruplex, such as the double-chain reversal. However, the loop regions as well as the number of G-tetrads and cytosine⁺–cytosine hemiprotonated base pairs that define a given DNA secondary structure provide a source of diversity and contribute to the complexity of these structures. For a comprehensive discussion on the diversity and complexity of G-quadruplexes readers are directed to a recently published review article [28]. Most commonly, the promoter DNA G-quadruplexes are formed from three G-tetrads; however, *PDGF-A* (four G-tetrads), *Rb* (two G-tetrads), and *c-Myb* (two G-tetrads) deviate from this trend (Table 1). Comparison of the three-tetrad G-quadruplex-forming sequences within the *c-myc*, *KRAS*, *bcl-2*, *VEGF*, *hTERT*, *hif-1 α* , *RET*, and *c-kit* and promoter regions reveals the conserved single-nucleotide 3'-end loop region (Fig. 5, Table 1) [28]. While this nucleotide may vary depending on the promoter, it yields the same structural feature, the double-chain reversal. The stability of a given G-quadruplex structure has been attributed mostly to the incorporation of the double-chain reversal [30,31,43,45,46,55]. With the exception of *hTERT* and *bcl-2*, this double-chain reversal also includes the 5'-end loop. The central loop region of these three-tetrad structures exhibits the greatest flexibility in terms of length and provides for the diverse folding patterns. In addition, further diversity is shown in both the *c-Myb* [41] and *hTERT* [50] G-quadruplexes, which consist of two tandem structures separated by a linker sequence.

Table 1 Proto-oncogene promoter G-quadruplex-forming sequences.

	N ₁₋₃			N ₂₋₂₆		N ₁₋₂			Topology	
<i>c-myc</i> :	5'-	GGG	T	GGG	TA	GGG	T	GGG	-3'	Parallel
<i>RET</i> :	5'-	GGG	C	GGG	GCG	GGG	C	GGG	-3'	Parallel
<i>VEGF</i> :	5'-	GGG	C	GGG	CCGG	GGG	C	GGG	-3'	Parallel
<i>c-kit-2</i> :	5'-	GGG	C	GGG	CGCGA	GGG	A	GGG	-3'	Parallel
<i>hif-1α</i> :	5'-	GGG	A	GGG	GAGAGG	GGG	C	GGG	-3'	Parallel
Mouse <i>KRAS</i> :	5'-	GGG	A	GGG	AAGGAGGGGA	GGG	C	GGG	-3'	Parallel
<i>hTERT</i> (1:3:1):	5'-	GGG	A	GGG	GCT	GGG	A	GGG	-3'	Parallel
<i>hTERT</i> (3:26:1):	5'-	GGG	GCT	GGG	C ₃₅ -G ₆₀	GGG	C	GGG	-3'	Parallel/Antiparallel
<i>bcl-2</i> :	5'-	GGG	CGC	GGG	AGGAAGG	GGG	C	GGG	-3'	Parallel/Antiparallel
<i>PDGF-A</i> :	5'-	GGGG	GG	GGGG	GGGCG	GGGG	CG	GGGG	-3'	Parallel
<i>Rb</i> :	5'-	GG	GG	GG	TTTT	GG	CG	GG	-3'	Antiparallel
<i>c-Myb</i> :	5'-	GG	A	GG	A	GG	A	GG	-3'	Parallel

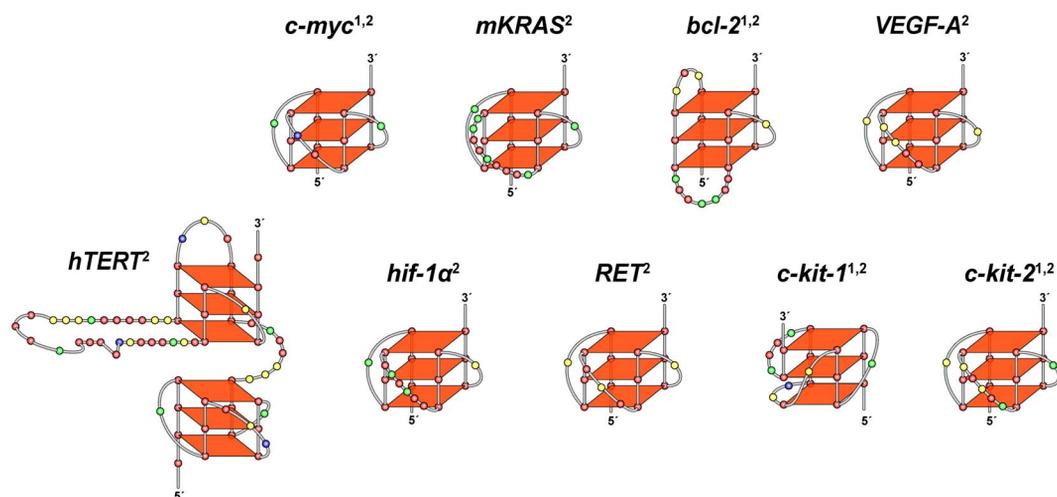


Fig. 5 The proposed folding patterns of the G-quadruplex structures formed within the *c-myc*, *VEGF*, *c-kit-1*, *c-kit-2*, *RET*, *KRAS*, *hif-1α*, *bcl-2*, and *hTERT* promoter regions. Structures were either determined by NMR (1) or biophysical methods (2), such as circular dichroism and DMS footprinting or both.

In contrast, promoter i-motif-forming sequences do not display as much similarity as that observed with G-quadruplexes. For example, i-motifs formed in the *VEGF*, *c-myc*, and *bcl-2* promoters consist of six intercalated cytosine⁺–cytosine hemiprotonated base pairs, while *RET* and *Rb* i-motifs are formed from five and four base pairs, respectively (Table 2). The loop sizes for the i-motifs also vary significantly with the *c-myc* [9] and *bcl-2* [32] i-motifs having the largest lateral loops, with six and eight nucleotides, and the *VEGF* [36], *RET* [43], and *Rb* [51] i-motifs having maximum lateral loop sizes of three nucleotides. The differences in loop sizes of both i-motif and G-quadruplex structures may offer one structurally diverse feature for small-molecular targeting.

Table 2 Proto-oncogene promoter i-motif-forming sequences.

	N ₂₋₈		N ₂₋₅		N ₂₋₇		
VEGF:	5'- CCC	GC	CCC	CGG	CCC	G	CCC -3'
c-myc:	5'- CCC	CACCTT	CCC	CA	CCC	TCCCCA	CCC -3'
RET:	5'- CCC	GC	CC	CGC	CCC	GC	CC -3'
bcl-2:	5'- CCC	GCTCCCGC	CCCC	TTTCT	CCC	GCGCCCG	CCCC -3'
Rb:	5'- CC	GC	CC	AAAA	CC	CC	CC -3'

SUMMARY

The rapid increase in recognition of the prevalence of DNA secondary structures within the genome, demonstration of stable structure formation within several promoter regions, and their ability to compete with duplex DNA is consistent with a *cis*-regulatory function. The close proximity of secondary structure-forming sequences to the TSS allows these regions to be included in the local unwinding that occurs during transcription. This creates a dynamic equilibrium between DNA secondary structures and duplex DNA that can be shifted toward G-quadruplex/i-motif formation upon binding of stabilizing intracellular proteins or small molecules. Likewise, the equilibrium may be shifted in the other direc-

tion in the presence of proteins or small molecules that facilitate the unwinding of the DNA secondary structures. More importantly, shifts in this equilibrium result in modulation of transcription that may be exploited for the development of novel anticancer agents.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (CA95060, CA122952, CA009213, GM085585) and the Leukemia & Lymphoma Society (6225-08). We thank David Bishop for his significant contribution in the preparation and editing of the final version of the text and figures displayed in the article.

REFERENCES

1. J. D. Watson, F. H. C. Crick. *Nature* **171**, 737 (1953).
2. R. E. Franklin, R. Gosling. *Nature* **171**, 740 (1953).
3. F. M. Pohl, T. M. Jovin. *J. Mol. Biol.* **67**, 375 (1972).
4. A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, A. Rich. *Nature* **282**, 680 (1979).
5. H. Drew, T. Takano, T. Tanaka, K. Itakura, R. E. Dickerson. *Nature* **286**, 567 (1980).
6. R. R. Sinden, T. J. Kochel. *Biochemistry* **26**, 1343 (1987).
7. D. M. Lilley, J. R. Hallam. *J. Mol. Biol.* **180**, 179 (1984).
8. H. Htun, J. E. Dahlberg. *Science* **243**, 1571 (1989).
9. D. Sun, L. H. Hurley. *J. Med. Chem.* **52**, 2863 (2009).
10. F. Kouzine, J. Liu, S. Sanford, H. J. Chung, D. Levens. *Nat. Struct. Mol. Biol.* **11**, 1092 (2004).
11. R. R. Sinden. *DNA Structure and Function*, Academic Press, San Diego (1994).
12. F. Kouzine, D. Levens. *Front. Biosci.* **12**, 4409 (2007).
13. L. F. Liu, J. C. Wang. *Proc. Natl. Acad. Sci. USA* **84**, 7024 (1987).
14. Z. Wang, P. Droge. *EMBO J.* **15**, 581 (1996).
15. D. A. Koster, V. Croquette, C. Dekker, S. Shuman, S. N. H. Dekker. *Nature* **434**, 671 (2005).
16. X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, R. H. Singer. *Nat. Struct. Mol. Biol.* **14**, 796 (2007).
17. F. Kouzine, S. Sanford, Z. Elisha-Feil, D. Levens. *Nat. Struct. Mol. Biol.* **15**, 146 (2008).
18. D. M. Lilley. *Proc. Natl. Acad. Sci. USA* **77**, 6468 (1980).
19. J. R. Williamson, M. K. Raghuraman, T. R. Cech. *Cell* **59**, 871 (1989).
20. J. L. Leroy, K. Gehring, A. Kettani, M. Gueron. *Biochemistry* **32**, 6019 (1993).
21. L. Chen, L. Cai, X. Zhang, A. Rich. *Biochemistry* **33**, 13540 (1994).
22. M. Gueron, J. L. Leroy. *Curr. Opin. Struct. Biol.* **10**, 326 (2000).
23. T. A. Brooks, L. H. Hurley. *Nat. Rev. Cancer* **9**, 849 (2009).
24. J. L. Huppert, S. Balasubramanian. *Nucleic Acids Res.* **35**, 406 (2007).
25. J. L. Huppert, S. Balasubramanian. *Nucleic Acids Res.* **33**, 2908 (2005).
26. A. K. Todd, M. Johnston, S. Neidle. *Nucleic Acids Res.* **33**, 2901 (2005).
27. A. Verma, K. Halder, R. Halder, V. K. Yadav, P. Rawal, R. K. Thakur, F. Mohd, A. Sharma, S. Chowdhury. *J. Med. Chem.* **51**, 5641 (2008).
28. Y. Qin, L. H. Hurley. *Biochimie* **90**, 1149 (2008).
29. P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle. *J. Am. Chem. Soc.* **126**, 16405 (2004).
30. A. Risitano, K. R. Fox. *Nucleic Acids Res.* **32**, 2598 (2004).
31. A. Bugaut, S. Balasubramanian. *Biochemistry* **47**, 689 (2008).
32. S. Kendrick, Y. Akiyama, S. M. Hecht, L. H. Hurley. *J. Am. Chem. Soc.* **131**, 17667 (2009).
33. J. Eddy, N. Maizels. *Nucleic Acids Res.* **34**, 3887 (2006).
34. J. L. Huppert. *Biochimie* **90**, 1140 (2008).

35. D. Sun, K. Guo, J. J. Rusche, L. H. Hurley. *Nucleic Acids Res.* **33**, 6070 (2005).
36. K. Guo, V. Gokhale, L. H. Hurley, D. Sun. *Nucleic Acids Res.* **36**, 4598 (2008).
37. S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian, S. Neidle. *J. Am. Chem. Soc.* **127**, 10584 (2005).
38. H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle, S. Balasubramanian. *Biochemistry* **45**, 7854 (2006).
39. A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel. *J. Am. Chem. Soc.* **129**, 4386 (2007).
40. P. S. Shirude, B. Okumus, L. Ying, T. Ha, S. Balasubramanian. *J. Am. Chem. Soc.* **129**, 7484 (2007).
41. S. Palumbo, R. M. Memmott, D. J. Uribe, Y. Krotova-Khan, L. H. Hurley, S. W. Ebbinghaus. *Nucleic Acids Res.* **36**, 1755 (2008).
42. Y. Qin, E. M. Rezler, V. Gokhale, D. Sun, L. H. Hurley. *Nucleic Acids Res.* **35**, 7698 (2007).
43. K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L. H. Hurley. *J. Am. Chem. Soc.* **129**, 10220 (2007).
44. S. Cogoi, L. E. Xodo. *Nucleic Acids Res.* **34**, 2536 (2006).
45. R. De Armond, S. Wood, D. Sun, L. H. Hurley, S. W. Ebbinghaus. *Biochemistry* **44**, 16341 (2005).
46. T. Dexheimer, D. Sun, L. H. Hurley. *J. Am. Chem. Soc.* **128**, 5404 (2006).
47. J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones, D. Yang. *J. Am. Chem. Soc.* **128**, 1096 (2006).
48. J. Dai, D. Chen, R. A. Jones, L. H. Hurley, D. Yang. *Nucleic Acids Res.* **34**, 5133 (2006).
49. H. Li, Y. Liu, S. Lin, G. Yuan. *Chem.—Eur. J.* **15**, 2445 (2009).
50. S. L. Palumbo, S. W. Ebbinghaus, L. H. Hurley. *J. Am. Chem. Soc.* **131**, 10878 (2009).
51. Y. Xu, H. Sugiyama. *Nucleic Acids Res.* **34**, 949 (2006).
52. T. Simonsson, R. Sjoback. *J. Biol. Chem.* **274**, 17379 (1999).
53. A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley. *Proc. Natl. Acad. Sci. USA* **99**, 11593 (2002).
54. A. T. Phan, Y. S. Modi, D. J. Patel. *J. Am. Chem. Soc.* **126**, 8710 (2004).
55. J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, L. H. Hurley. *J. Am. Chem. Soc.* **126**, 8702 (2004).
56. A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang. *Biochemistry* **44**, 2048 (2005).
57. D. Yang, L. H. Hurley. *Nucleosides Nucleotides* **25**, 951 (2006).
58. T. Simonsson, M. Pribylova, M. Vorlickova. *Biochem. Biophys. Res. Commun.* **278**, 158 (2000).
59. N. Khan, A. Avino, R. Tauler, C. Gonzalez, R. Eritja, R. Gargallo. *Biochimie* **89**, 1562 (2007).
60. A. T. Phan, J. L. Mergny. *Nucleic Acids Res.* **30**, 4618 (2002).
61. W. Li, P. Wu, T. Ohmichia, N. Sugimoto. *FEBS Lett.* **526**, 77 (2002).
62. N. Kumar, S. Maiti. *Biochem. Biophys. Res. Commun.* **319**, 759 (2004).
63. Y. Zhao, Z. Y. Kan, Z. X. Zeng, Y. H. Hao, H. Chen, Z. Tan. *J. Am. Chem. Soc.* **126**, 13255 (2004).
64. K. Halder, S. Chowdhury. *Nucleic Acids Res.* **33**, 4466 (2005).
65. A. Arora, D. R. Nair, S. Maiti. *FEBS Lett.* **276**, 3628 (2009).
66. M. Fry. *Front. Biosci.* **12**, 4336 (2007).
67. M. Paramasivam, A. Membrino, S. Cogoi, H. Fukuda, H. Nakagama, L. E. Xodo. *Nucleic Acids Res.* **37**, 2841 (2009).
68. T. S. Dexheimer, S. S. Carey, S. Zuohe, V. M. Gokhale, X. Hu, L. B. Murata, E. M. Maes, A. Weichsel, D. Sun, E. J. Meillet, W. R. Montfort, L. H. Hurley. *Mol. Cancer Ther.* **8**, 1363 (2009).
69. R. K. Thakur, P. Kumar, K. Halder, A. Verma, A. Kar, J. L. Parent, R. Basundra, A. Kumar, S. Chowdhury. *Nucleic Acids Res.* **37**, 172 (2009).
70. V. Gonzalez, K. Guo, L. Hurley, D. Sun. *J. Biol. Chem.* **284**, 23622 (2009).
71. S. Cogoi, M. Paramasivam, B. Spolaore, L. E. Xodo. *Nucleic Acids Res.* **36**, 3765 (2008).

72. J. Cuesta, M. A. Read, S. Neidle. *Mini Rev. Med. Chem.* **3**, 11 (2003).
73. A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel. *Nat. Chem. Biol.* **1**, 167 (2005).
74. J. Seenisamy, S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui-Jain, N. Streiner, K. Shin-ya, E. White, W. D. Wilson, L. H. Hurley. *J. Am. Chem. Soc.* **127**, 2944 (2005).
75. L. H. Hurley, D. D. Von Hoff, A. Siddiqui-Jain, D. Yang. *Semin. Oncol.* **33**, 498 (2006).
76. H. Han, D. R. Langley, A. Rangan, L. H. Hurley. *J. Am. Chem. Soc.* **123**, 8902 (2001).
77. M. Gunaratnam, S. Swank, S. M. Haider, K. Galesa, A. P. Reszka, M. Beltran, F. Cuenca, J. A. Fletcher, S. Neidle. *J. Med. Chem.* **52**, 3774 (2009).
78. Z. A. E. Waller, S. A. Sewitz, S. T. D. Hsu, S. Balasubramanian. *J. Am. Chem. Soc.* **131**, 12628 (2009).
79. C. Heckman, E. Mochon, M. Arcinas, L. M. Boxer. *J. Biol. Chem.* **272**, 19609 (1997).
80. H. Li, Y. Liu, S. Lin, G. Yuan. *Chem.—Eur. J.* **15**, 2445 (2009).
81. M. del Toro, P. Buck, A. Avino, J. Jaumot, C. Gonzalez, R. Eritja, R. Gargallo. *Biochimie* **91**, 894 (2009).
82. H. Fernando, R. Rodriguez, S. Balasubramanian. *Biochemistry* **47**, 9365 (2008).
83. H. Fernando, S. Sewitz, J. Darot, S. Tavares, J. L. Huppert, S. Balasubramanian. *Nucleic Acids Res.* **37**, 6716 (2009).