A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the *MYC* Promoter NHE III$_1$

Caleb Sutherland,[†] Yunxi Cui,[‡] Hanbin Mao,[‡] and Laurence H. Hurley*[†,§,‖]

[†]University of Arizona Cancer Center, 1515 North Campbell Avenue, Tucson, Arizona 85724 USA

[‡]Department of Chemistry and Biochemistry and School of Biomedical Sciences, Kent State University, Kent, Ohio, 44242 USA

[§]University of Arizona, College of Pharmacy, 1703 East Mabel Street, Tucson, Arizona 85721 USA

[‖]BIO5 Institute, 1657 East Helen Street, Tucson, Arizona 85721 USA

*Corresponding Author

Tel: 520-626-5621

Email: hurley@pharmacy.arizona.edu

Keywords: MYC, hnRNP K, SP1, transcription control, negative superhelicity

ABSTRACT: *MYC* is overexpressed in many different cancer types and is an intensively studied oncogene because of its contributions to tumorigenesis. The regulation of *MYC* is complex, and the NHE III_1 and FUSE elements rely upon noncanonical DNA structures and transcriptionally induced negative superhelicity. In the NHE III_1 only the G-quadruplex has been extensively studied, whereas the role of the i-motif, formed on the opposite C-rich strand, is much less understood. We demonstrate here that the i-motif is formed within the 4CT element and is recognized by hnRNP K, which leads to a low level of transcription activation. For maximal hnRNP K transcription activation, two additional cytosine runs, located seven bases downstream of the i-motif–forming region, are also required. To access these additional runs of cytosine, increased negative superhelicity is necessary, which leads to a thermodynamically stable complex between hnRNP K and the unfolded i-motif. We also demonstrate mutual exclusivity between the *MYC* G-quadruplex and i-motif, providing a rationale for a molecular switch mechanism driven by SP1-induced negative superhelicity, where relative hnRNP K and nucleolin expression shifts the equilibrium to the on or off state.

INTRODUCTION

MYC is overexpressed in many different cancer types and is an intensively studied oncogene because of its contributions to tumorigenesis.[1] The MYC protein is a basic helix-loop-helix leucine zipper transcription factor that can dimerize with MAX and activate transcription of enhancer box (E-box)–containing genes.[2] MYC is utilized in many oncogenic pathways that are involved in apoptosis,[3] cell cycle progression,[4] cell adhesion,[5] metabolism,[6] and growth and differentiation.[7] Silencing of *MYC* has been extensively investigated as a potent anticancer strategy. As a consequence of a phenomenon referred to as "oncogene addiction,"[8] brief inactivation of *MYC* can cause preferential killing of tumor cells[9]; however, inhibitors have not reached the clinic due to the difficulty of targeting MYC, which lacks common druggable domains and has a short half-life.[10] Therefore, finding a way to inhibit MYC is an important clinical problem. The inhibitor JQ1, which targets the bromodomains of proteins associated with acetylation of chromatin, decreases *MYC* expression by an indirect mechanism involving the interaction with BRD4 of the IgH enhancer translocated upstream of *MYC*.[11] A more direct mechanism for lowering *MYC* transcription may be achieved by targeting the G-quadruplexes. These four-stranded DNA secondary structures, which are formed from a minimum of four runs of guanines, are found frequently in nuclease-sensitive regions that are located upstream of the transcription start site in many genes involved in cancer, such as MYC, BCL2, PDGFR-β, c-KIT, k-RAS, VEGF, and hTERT, that regulate gene expression.[12-15] While the vast majority of the published work has been around the role of the G-quadruplex as a silencer element, more recently the role of the i-motif, formed on the opposite strand to the G-quadruplexes, has come into focus. In the case of BCL2, the i-motif has been shown to be involved in activation of transcription mediated through binding of heterogeneous nuclear ribonucleoprotein L-like (hnRNP LL), and like the G-quadruplex, it can be targeted with small molecules to modulate gene expression.[16,17] Furthermore, the mutual exclusivity of the G-quadruplex and the i-motif in promoter elements of a number of genes, including hINS and hTERT,[18] adds to the premise that

these structures may act as molecular switches in controlling gene expression. Moreover, the opposite roles of these structures in the control of gene expression provide a means not only to silence *MYC* expression but to increase gene expression, which has been reported to be toxic to tumor cells.[19,20]

The nuclease hypersensitive element III_1 (NHE III_1) is part of a larger proximal control element consisting of a far upstream element (FUSE) located 1.7 kb upstream of the P2 promoter, which is 0.16 kb downstream of the P1 promoter (Figure 1). Negative superhelicity initiated at the proximal promoter travels upstream to melt the FUSE mechanosensor element,[21] which then allows the activating protein, the FUSE-binding protein (FBP), to bind to and interact with the TFIIH-enhancing element. The FUSE/FBP/TFIIH system forms a topologically closed loop that accumulates additional superhelicity and increases transcription, further melting the FUSE and allowing the repressive FBP-interacting repressor (FIR) to interact with and eject FBP, relaxing the DNA and slowing transcription.[22] The FUSE/FBP/FIR/TFIIH system cannot itself initiate transcription, which first requires activation by SP1 binding to five duplex-binding sites upstream of P1 (CT-I to CT-V) and one site upstream of P2 (CT-I_2).[23] More immediately upstream of P1 is the NHE III_1, located between 101 and 147 bases upstream of the promoter and containing eight runs of three or more guanines/cytosines, where SP1 binds (Figure 1). This element has been extensively characterized for its ability to enhance and repress transcription through the formation of single-stranded DNA (ssDNA) and two DNA secondary structures, the G-quadruplex and i-motif.[24-28] The torsional stress induced by negative superhelicity generated behind RNAP can melt the NHE III_1, giving rise to regulatory non–B-DNA structures.[27,29,30] A single predominant G-quadruplex forms, utilizing only the four central runs of the NHE III_1. The *MYC* G-quadruplex can be bound by the protein nucleolin or by small molecules that mimic the stabilizing effect of nucleolin to repress transcription.[31-34] On the strand opposite the G-quadruplex are the eight complementary runs of cytosines that form the i-motif. As we

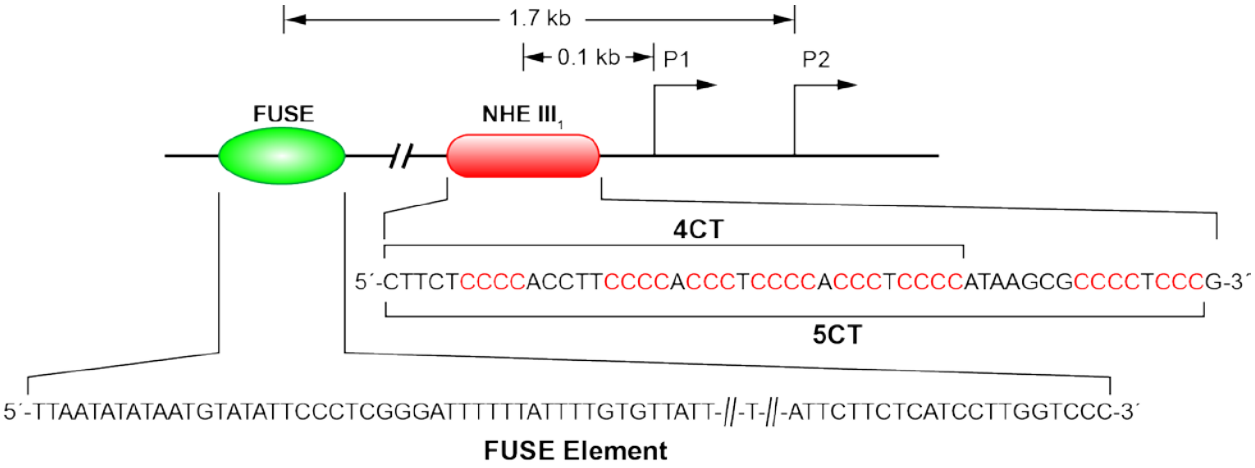


Figure 1. Cartoon showing the proximal promoter elements of *MYC* and the downstream P1 and P2 promoters. *MYC* promoter regulation is controlled by two noncanonical *cis*-regulatory elements induced by negative superhelicity resulting from transcription activation. Two regulatory elements exist upstream of the P1 promoter: the FUSE 1.7 kb upstream of P2 and the NHE III_1 located between 101 and 147 bases upstream of P1. The FUSE element is largely AT-rich, whereas the NHE III_1 element is GC-rich. The 4CT element contains six runs of three or more cytosines and the 5CT element contains eight (shown in red).

demonstrate in this contribution, the i-motif on the C-rich strand has an opposite function to the G-quadruplex and can be bound by hnRNP K to activate transcription.

In a previous mutational study of the *MYC* NHE III_1 it was shown that all eight cytosine runs that make up five tandem CT (5CT) elements (CCCTCCCC) must be completely intact and are required for maximal transcription activation from the P1 and P2 promoters.[23] While hnRNP K is much better known as an RNA-binding protein, there is convincing evidence that this protein has an important role in the control of MYC gene expression through binding to the C-rich strand of DNA.[28] In fact, hnRNP K binds to DNA much better than to RNA,[35] and there is a published crystal structure showing that the KH3 domain of hnRNP K binds specifically to a CCCT ssDNA sequence.[36] The implications of the involvement of the i-motif and the associated 5CT element and hnRNP K have not been investigated. Previous studies using plasmid DNA into which the promoter sequence was inserted demonstrated that the first four CT elements (4CT) could fold into an i-motif with a 2:6:2 loop configuration.[27,37]

Many other G-quadruplexes have been targeted to inhibit transcription, including *hTERT,*[38] *VEGF*,[39] *PDGFA*,[40] and *BCL2*.[41] The *MYC* G-quadruplex was the first to be targeted by the small molecule TMPyP4 to downregulate gene transcription, using two different Burkitt's lymphoma cell lines.[34] In subsequent work the G-quadruplex–selective compound GQC-05 was shown to specifically target the G-quadruplex in the *MYC* promoter, using an exon-specific assay in CA46 lymphoma cells.[42] In more recent work we reported that we can activate or repress *BCL2* expression by stabilizing either the *BCL2* i-motif or its equilibrating unfolded hairpin species.[17] The hairpin-stabilizing compound IMC-76 represses *BCL2* transcription by shifting the dynamic equilibrium from the i-motif to the hairpin species, whereas the i-motif-stabilizing compound IMC-48 shifts the equilibrium to the i-motif to produce overexpression of *BCL2*.[17] In parallel studies we demonstrated that hnRNP LL specifically binds to the lateral loops of the *BCL2* i-motif and unfolds this species to activate transcription.[16] IMC-48 and IMC-76 increase and decrease the *BCL2* promoter occupancy to mediate transcription activation or

repression. The results of these studies with the *BCL2* i-motif suggested that hnRNP K might also recognize the *MYC* i-motif and that it would be possible to identify compounds that acted in a similar manner to IMC-76 and IMC-48.[16]

In this contribution we show that hnRNP K recognizes the i-motif in the 4CT element but requires additional negative superhelicity and the 5CT element for full transcription activation. We also show the antagonistic effects of nucleolin and hnRNP K, which bind to the G-quadruplex and i-motif.

RESULTS

The Same Predominant i-Motif Forms in Both the 4CT and 5CT Sequences in the NHE III_1. The full-length cytosine-rich sequence contains eight runs of three or four consecutive cytosines that can be used for i-motif formation (Figure 2A). Previous chemical and biophysical studies on the C-rich strand of the NHE III_1 have failed to determine the importance of all eight runs by not taking into account the two most 3′ runs (VII and VIII).[43,44] Thus our initial experiments were performed to determine if different i-motif structures would result from using either the 4CT or 5CT sequence. The importance of an acidic pH for i-motif formation in a single-stranded oligonucleotide is well established and can be directly related to the thermostability of the i-motif.[45] To test the relative stability of the 4CT compared to the 5CT sequence, circular dichroism (CD) was used to measure pH dependence. The i-motif has a signature spectrum with a molar ellipticity maximum peak near 286 nm and a minimum peak near 260 nm.[46] Analysis of the 4CT (Figure 2B) and 5CT (Figure 2C) sequences showed that for both sequences the molecular ellipticity decreased and shifted to a lower wavelength as the pH became less acidic, indicative of a less stable structure. This effect was more pronounced in the 5CT sequence. The transitional pHs for the 4CT (6.89) and the 5CT (6.53) sequences were quantitatively measured by plotting the maximum molar ellipticity at 286 nm versus the pH. Similarly, i-motif stability is also directly related to temperature,[47] and the melting temperatures

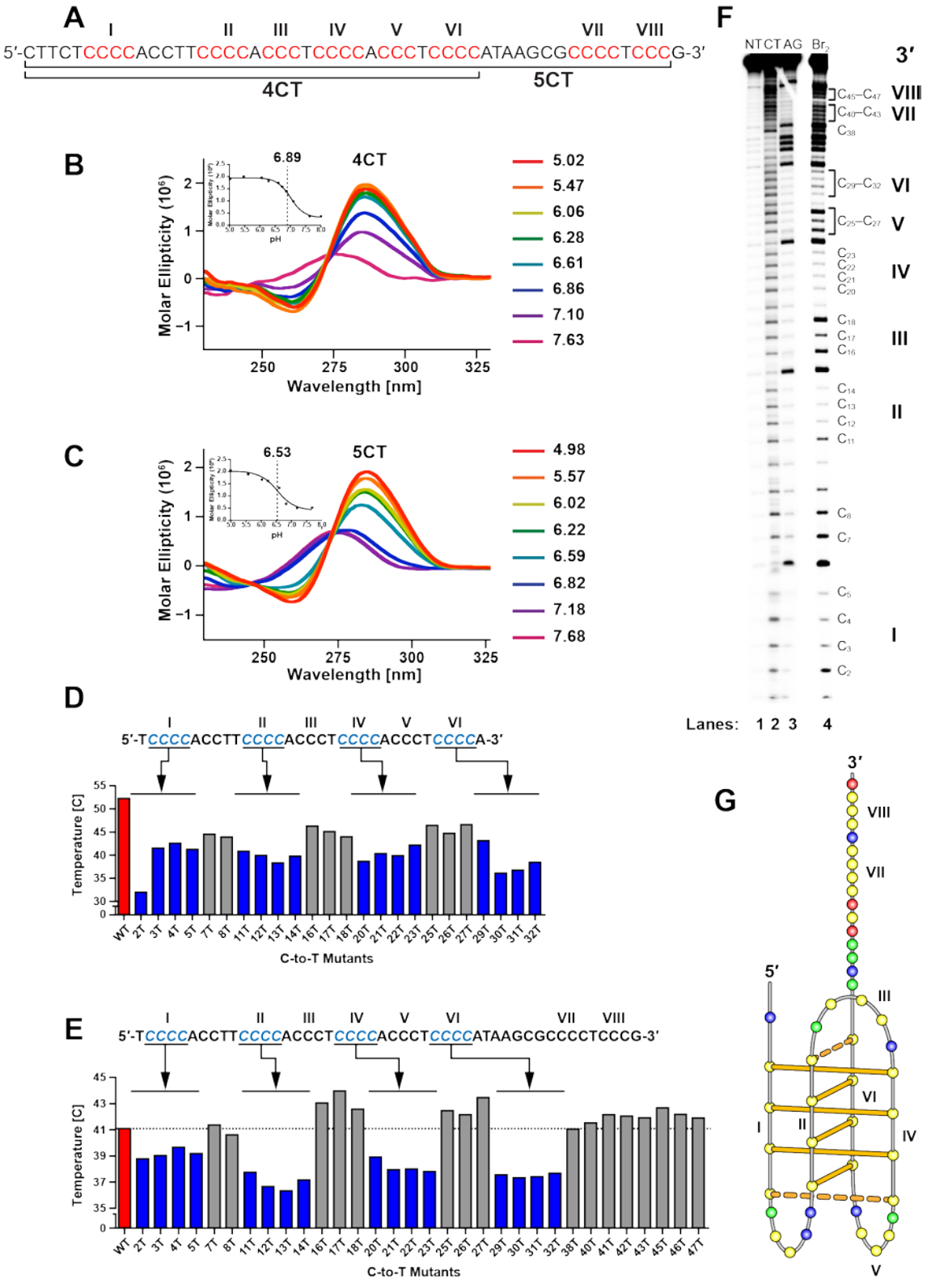


Figure 2. The i-motif in the *MYC* promoter is formed exclusively from the 4CT element and has a 5:5:5 loop configuration. (A) The NHE III_1 sequence showing the 5CT and truncated 4CT sequence, with the cytosine runs labeled I–VIII. Comparison of CD spectra of the 4CT (B) and

5CT (C) elements at different pHs. Inserts show the transitional pHs of the 4CT and the 5CT sequences determined by plotting the CD spectra molar ellipticity at 286 nm versus pH. Histogram representation of the CD melting temperatures of each 4CT cytosine-to-thymine mutant (D) and each 5CT cytosine-to-thymine mutant (E) oligomer. Above each histogram are shown the 4CT and 5CT sequences, and those bases involved in C^+–C base pairing are shown in blue. (F) Bromine footprinting pattern of the WT 5CT oligomer at pH 6.5. Lanes 1–3 represent controls; no treatment (NT) represents the cleavage reaction in the absence of bromine, CT represents pyrimidine, and AG represents purine sequencing. Lane 4 is the Br_2 footprinting pattern. To the right of the gel, the roman numerals refer to each cytosine run (I–VIII), and numbers refer to individual cytosines (C_2–C_{47}). (G) Illustration of the proposed folding pattern of the *MYC* i-motif based on the mutational and bromine footprinting data in E and F. Cytosine runs I–VIII are indicated.

(T_m) were found to be 39.6 °C and 36.4 °C for the 4CT and 5CT elements respectively (data not shown).

To determine the cytosines involved in C^+–C base pairing, single-stranded oligonucleotides containing single C-to-T mutations were each evaluated by measuring the T_m. The results show that the largest decreases in T_m were observed in mutants with C-to-T base substitutions in runs I, II, IV, and VI for both the 4CT and 5CT sequences, whereas little effect was seen with mutants in runs III, V, VII, and VIII (Figure 2, D and E). Bromine footprinting can be used in a confirmatory way to identify the cytosines involved in C^+–C base pairing. Bromination of cytosines that lack steric or electrostatic hindrance occurs at the C5 position, and subsequent reactivity with piperidine results in a cleavage product that can be visualized with a PAGE gel.[48,49] In agreement with the T_m measurements, cytosines in runs I, II, IV, and VI showed protection from bromination and subsequent piperidine cleavage, whereas cytosines in runs III, V, VII, and VIII were not protected by base pairing, and cleavage products were seen (Figure 2F). Thus while it appears that the 4CT sequence forms a more thermodynamically stable i-motif than the 5CT sequence, the results show that the same bases are used for i-motif formation, resulting in a 5:5:5 loop configuration (Figure 2G). This loop folding pattern is different to that reported previously from a supercoiled plasmid experiment[27] and is probably due to the poorer resolution of the footprinting gel in the earlier experiment.

Although hnRNP K Preferentially Binds to the Longer 5CT Sequence, the Initial Recognition Is through the i-Motif in the 4CT Element. Previous work reported hnRNP K binds to the *MYC* promoter and activates transcription through recognition of C-rich ssDNA.[50] By analogy with the *BCL2* i-motif–forming sequence in the promoter element,[16] this recognition by hnRNP K may be through preferential interactions with the constrained form of the sequence in the i-motif rather than ssDNA. Indeed, the consensus sequences (5′-C-C-C/T-T-3′) for the KH domains of hnRNP K[36] are located in all three loops of the *MYC* i-motif. Competition electrophoretic mobility shift assays (EMSAs) demonstrated the enhanced affinity of hnRNP K

for the wild-type (WT) sequence i-motif compared to the mutant, which cannot form the i-motif (Figure 3A). This is in accord with our suspicion that the i-motif scaffold presents these consensus sequences in a preferred, conformationally more constrained structure that is absent in the single-stranded form. Additionally, less binding affinity for hnRNP K was also observed with the *BCL2* i-motif, which contains the GCTCCCG, TTCCT, and GCGCCCG loop sequences (Figure 3A).[48] This finding, alongside the analogous result with the *BCL2* i-motif and hnRNP LL,[16] provides a basis for understanding why the i-motif is an important recognition feature in transcription control of gene expression.

Having established that the i-motif form is preferred for hnRNP K recognition, we next investigated the cellular role of the additional cytosine runs present in the 5CT sequence that are required to promote maximal *MYC* transcription.[23] The presence of these additional runs of cytosines in the fifth CT element suggested that they might also be important for stable binding of hnRNP K. As shown by competition EMSA, unlabeled cold single-stranded 4CT DNA required at least a 10-fold increase in concentration over the 5CT element to compete for hnRNP K binding (Figure 3B). The preference of hnRNP K for the 5CT sequence can be rationalized if we assume the additional CT element present in the 5CT sequence is used as a binding site for the third KH domain in hnRNP K.

To further characterize the binding of the 5CT oligomer with hnRNP K, we compared the bromine footprinting of this element in the presence and absence of hnRNP K. Upon addition of hnRNP K, cytosine runs III, V, and VII, containing the CCCT recognition sequences, are protected from cleavage (Figure 3C). Interestingly, each of the 3′ flanking sequences (IV, VI, and VIII), located next to recognition sequences III, V, and VII, are also protected, showing a consistent pattern. Taken together, these results suggest that the protein is in proximity to an unfolded i-motif covering cytosine runs II through VIII, with the three KH domains binding to the central loop (III), the 3′ lateral loop (V), and the flanking fifth CT (VII/VIII) element. Thus we suggest that the 4CT sequence, when folded into an i-motif, is recognized initially by two of the

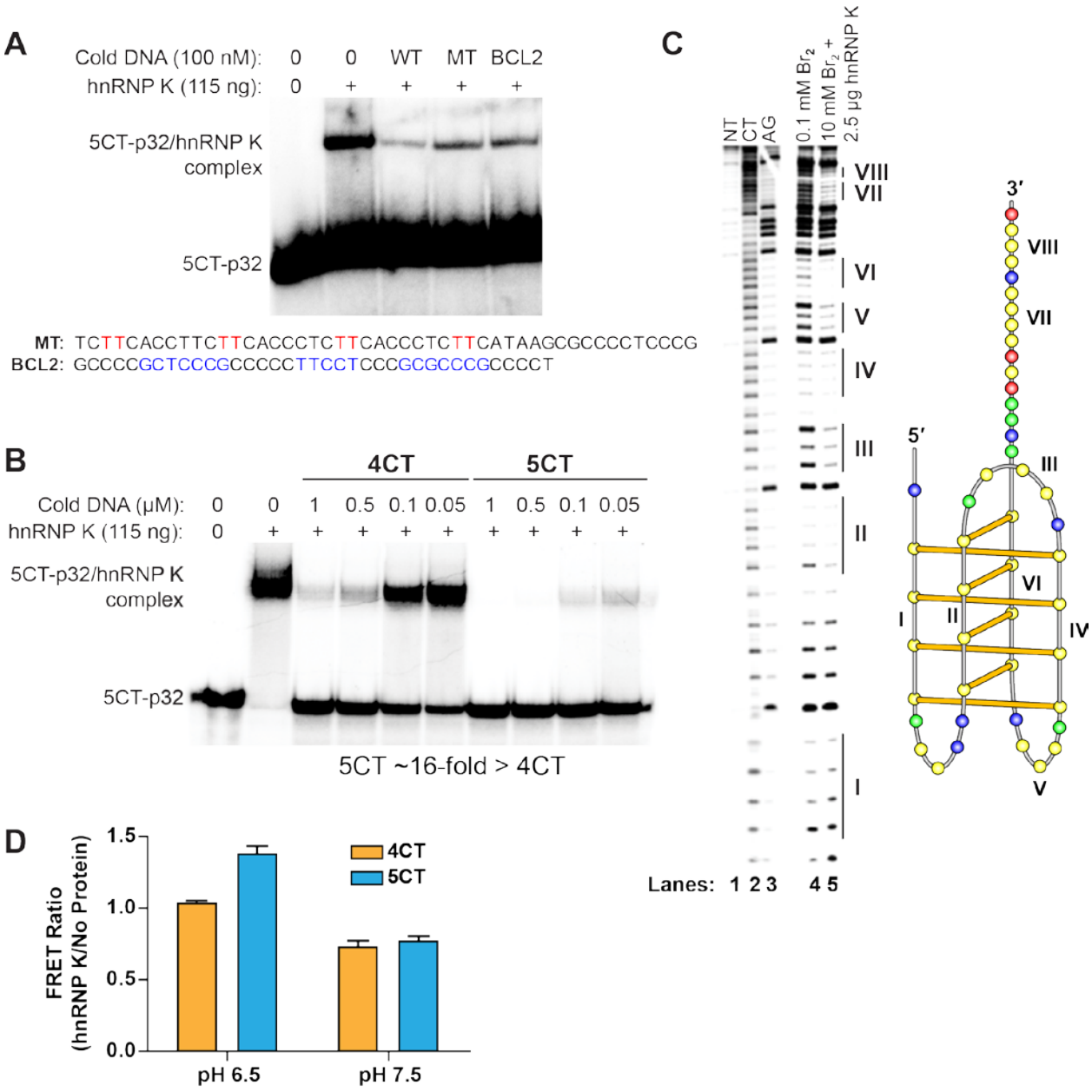


Figure 3. hnRNP K preferentially binds to the 5CT sequence and unfolds the i-motif to form the thermodynamically stable complex. (A) Competition EMSAs between the WT 5CT sequence and the *BCL2* i-motif–forming sequence (loops bases in blue) and also with the 5CT mutant (MT, mutants bases in red), which cannot form an i-motif. The mutant 5CT and *BCL2* sequences are shown below the gel. (B) Competition EMSAs between the 4CT or WT 5CT sequence and full-length hnRNP K at four different 4CT and 5CT cold competitor concentrations. (C) Bromine footprinting pattern of the 5CT sequence with and without hnRNP K at pH 6.5. Lanes 1–3 represent controls, no treatment (NT) represents the cleavage reaction in the absence of bromine, and CT and AG represent pyrimidine and purine sequencing respectively. Lanes 4 and 5 are treatment with Br_2 without and with hnRNP K. The Br_2 footprinting reveals that hnRNP K utilizes cytosine runs III, V, VII, and VIII, as shown by enhanced protection to Br_2 cleavage. The cartoon to the right of the gel shows the proposed *MYC* i-motif folding pattern labeled with cytosine runs I–VIII. (D) FRET results showing hnRNP

K differential unfolding activity of the 5CT i-motif compared to the 4CT i-motif at pH 6.5 and pH 7.5.

KH domains of hnRNP K in a kinetically favored manner over the ssDNA substrate. However, if the fifth CT element becomes available, then all three KH domains are utilized in forming a thermodynamically stable complex between the 5CT element and hnRNP K. Indeed, on the basis of the results of competition experiments with either cold 4CT or 5CT, we found, as expected, that with a mutant hnRNP K protein containing only one KH domain, the preference for the 5CT element was reduced to 2.5-fold, compared to 16-fold for the WT hnRNP K (Supplementary Figure 1). To confirm whether binding of hnRNP K unfolds the i-motif, comparative FRET experiments using dual-labeled 3′ TAMRA/3′ FAM at the 3′ and 5′ ends were carried out for both the 4CT and 5CT sequences at pHs of 6.5 and 7.5 respectively. The results in Figure 3D show that only the 5CT element at a pH of 6.5 had a significant increase (**P <0.01) in the FRET ratio upon addition of hnRNP K, which is consistent with the premise that unfolding requires both the fifth CT element and an i-motif, which displays the central and lateral loops CCCT in a constrained manner. However, at pH 7.5 both exist primarily in the unconstrained form and the FRET ratio is decreased to a similar extent. This implies that a different, nonspecific binding event occurs that is common to both the unfolded 4CT and 5CT at pH 7.5.

hnRNP K Recognizes and Binds to the CCCT Sequence Located in the 3′ Lateral and Central Loops of the i-Motif and to the Single-Stranded Fifth CT Element. To understand how the i-motif and the fifth CT element contribute to transcription activation by hnRNP K, it is important to identify which binding sequences are required for initial recognition and then final binding by hnRNP K's three KH domains. For this purpose we used cold unlabeled mutant oligonucleotides with single or combination mutations in the i-motif loops in a competition EMSA with the labeled WT 5CT sequence. In the competition EMSA, single-loop mutations of loops 1 and 2 (L1 and L2 in Figure 4A) competed the best in comparison to the other mutated sequences, followed by the loop 3 mutant (L3), revealing loop 3 to be the preferred loop for recognition by hnRNP K (Figure 4A). The three loops were then mutated in pairs of two loops at

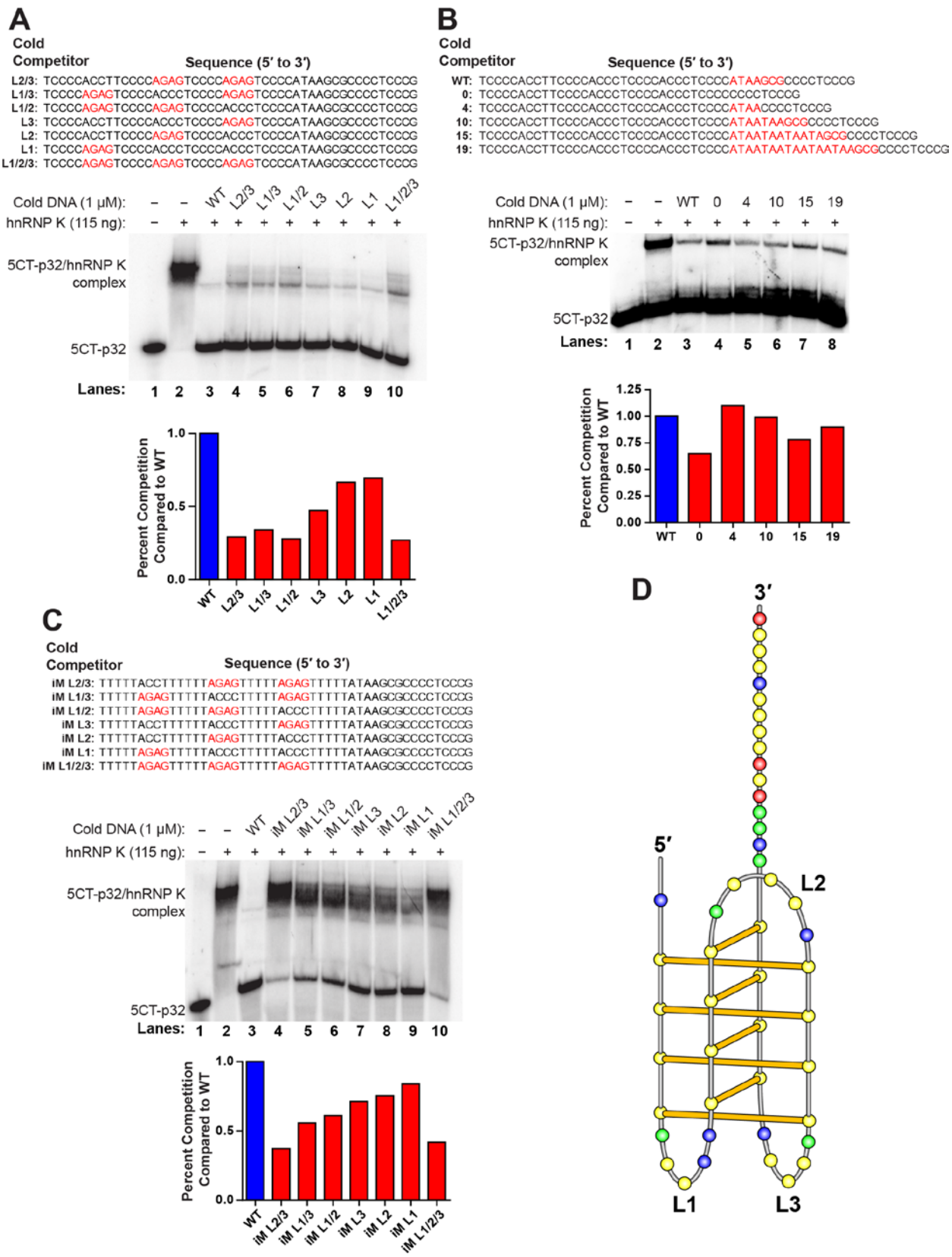

A
Cold Competitor
Sequence (5′ to 3′)
L2/3: TCCCCACCTTCCCCAGAGTCCCCAGAGTCCCCATAAGCGCCCCTCCCG
L1/3: TCCCCAGAGTCCCCACCCTCCCCAGAGTCCCCATAAGCGCCCCTCCCG
L1/2: TCCCCAGAGTCCCCAGAGTCCCCACCCTCCCCATAAGCGCCCCTCCCG
L3: TCCCCACCTTCCCCACCCTCCCCAGAGTCCCCATAAGCGCCCCTCCCG
L2: TCCCCACCTTCCCCAGAGTCCCCACCCTCCCCATAAGCGCCCCTCCCG
L1: TCCCCAGAGTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCCTCCCG
L1/2/3: TCCCCAGAGTCCCCAGAGTCCCCAGAGTCCCCATAAGCGCCCCTCCCG
Cold DNA (1 μM): – – WT L2/3 L1/3 L1/2 L3 L2 L1 L1/2/3
hnRNP K (115 ng): – + + + + + + + + +
5CT-p32/hnRNP K complex
5CT-p32
Lanes: 1 2 3 4 5 6 7 8 9 10
Percent Competition Compared to WT
1.0
0.5
0.0
WT L2/3 L1/3 L1/2 L3 L2 L1 L1/2/3
B
Cold Competitor
Sequence (5′ to 3′)
WT: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCCTCCCG
0: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCCCCCTCCCG
4: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAACCCCTCCCG
10: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAATAAGCGCCCCTCCCG
15: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAATAATAATAGCGCCCCTCCCG
19: TCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAATAATAATAATAAGCGCCCCTCCCG
Cold DNA (1 μM): – – WT 0 4 10 15 19
hnRNP K (115 ng): – + + + + + + +
5CT-p32/hnRNP K complex
5CT-p32
Lanes: 1 2 3 4 5 6 7 8
Percent Competition Compared to WT
1.25
1.00
0.75
0.50
0.25
0.00
WT 0 4 10 15 19
C
Cold Competitor
Sequence (5′ to 3′)
iM L2/3: TTTTTACCTTTTTTAGAGTTTTTAGAGTTTTTATAAGCGCCCCTCCCG
iM L1/3: TTTTTAGAGTTTTTACCCTTTTTAGAGTTTTTATAAGCGCCCCTCCCG
iM L1/2: TTTTTAGAGTTTTTAGAGTTTTTACCCTTTTTATAAGCGCCCCTCCCG
iM L3: TTTTTACCTTTTTTACCCTTTTTAGAGTTTTTATAAGCGCCCCTCCCG
iM L2: TTTTTACCTTTTTTAGAGTTTTTACCCTTTTTATAAGCGCCCCTCCCG
iM L1: TTTTTAGAGTTTTTACCCTTTTTACCCTTTTTATAAGCGCCCCTCCCG
iM L1/2/3: TTTTTAGAGTTTTTAGAGTTTTTAGAGTTTTTATAAGCGCCCCTCCCG
Cold DNA (1 μM): – – WT iM L2/3 iM L1/3 iM L1/2 iM L3 iM L2 iM L1 iM L1/2/3
hnRNP K (115 ng): – + + + + + + + + +
5CT-p32/hnRNP K complex
5CT-p32
Lanes: 1 2 3 4 5 6 7 8 9 10
Percent Competition Compared to WT
1.0
0.5
0.0
WT iM L2/3 iM L1/3 iM L1/2 iM L3 iM L2 iM L1 iM L1/2/3
D
3′
5′
L2
L1
L3

Figure 4. Competition EMSAs demonstrate that the preferred hnRNP K binding substrate requires the presence of two loop sequences in the i-motif and an additional recognition sequence at approximately seven bases downstream of the i-motif. (A) Sequences of single-stranded oligonucleotides contain either single or combination mutants in loops 1–3 of the 5CT i-motif sequence (upper). Competition EMSA with mutant oligomers with single (lanes 7–9) and multiple loop mutants (lanes 4–6 and 10) used as cold competitors against the labeled 5CT sequence (middle). Densitometry scans of the competition EMSA showing band intensities relative to the cold WT 5CT sequence (lane 3 in EMSA) (lower). (B) Sequences of single-stranded oligonucleotides with increasing nucleotide length (from 0 to 19), in which nucleotides where removed or added between cytosine runs VI and VII in the mutant oligomers (upper). Competition EMSA with cold competitors (lanes 3–8) competing against the labeled 5CT sequence for hnRNP K binding (middle). Densitometry scan of the competition EMSA comparing all mutants to WT 5CT cold competitor (lane 3 in EMSA) (lower). (C) Sequences of single-stranded oligonucleotides containing C-to-T mutations in the cytosine runs involved in i-motif formation (I, II, IV, VI) also having additional single or combination mutations in loops 1–3 (upper). Oligomers in which all cytosines involved in C^+–C base pairing were mutated to thymine, in addition to the loop sequences that were mutated in either one loop (lanes 7–9) or two loops (lanes 4–7) (middle). EMSA (middle) and densitometry scan (lower) of the competition EMSA comparing all mutants to WT 5CT cold competitor (lane 3 in EMSA). (D) Proposed folding pattern on *MYC* i-motif with loops labeled.

a time. These mutants were significantly less competitive for hnRNP K binding than the single mutants, and no significant differences were observed between any pairs of the loop mutants (L2/3, L1/3, and L1/2) in Figure 4A. As expected, the mutation of L1, L2, and L3 resulted in the least competitive substrates. Taken together, these results suggest that the i-motif loops provide the recognition motifs for two of the three KH domains, while the third recognizes the fifth CT element. Of the three loop sequences, L3 is preferred for binding to hnRNP K, with L2 marginally preferred over L1. Not unexpectedly, L1/2/3 was the least competitive of all the sequences tested.

While hnRNP K has three KH domains for recognition of loop sequences in the *MYC* i-motif, hnRNP LL has four RNA recognition domains that have been shown to recognize the protein consensus sequences in the lateral loops of the *BCL2* i-motif. Significantly, it was found that there is an optimal 13-nucleotide spacer between hnRNP LL binding sequences, which is the exact sequence length between the lateral loops of the *BCL2* i-motif.[16] On the basis of the results of this study, we propose that the two closely spaced KH1 and KH2 domains bind to two of the i-motif loops, whereas the less closely spaced KH3 binds to the fifth CT element. Using a similar approach to experiments carried out by Kang et al., mutants were designed with different length linkers between the i-motif and the fifth CT element.[16] The results show that the mutants preferred for hnRNP K recognition were those with nucleotide spacers closest in length to the wild-type distance of seven bases (Figure 4B). Therefore, for optimal hnRNP K recognition of the *MYC* promoter's C-rich strand, the consensus sequences must be conformationally restricted by the i-motif and have access to an additional single-stranded CT element seven nucleotides from the 3′ end of the i-motif.

To determine which loop sequences hnRNP K would utilize in the absence of the i-motif, we mutated those cytosines found in the i-motif C^+–C base pairing. Additionally, each loop was mutated individually or in combination and then tested by competition EMSA (Figure 4C). Mutations in C-tracts involved in i-motif formation required a three-fold increase in cold DNA

compared to the previous experiments in Figure 4, A and B, in order to see a competitive effect. The sequence with loop 1 mutated (iM L1) was the most competitive of all the mutants. This loop sequence is the farthest from the fifth CT element and therefore is probably not utilized by any of the three KH domains in the final complex. As anticipated, sequences with mutations in loops 2 or 3 (iM L2 or iM L3) are comparatively less competitive, indicating that they are more likely to be utilized for hnRNP K binding in the final complex. Finally, as expected, the i-motif combination mutants were significantly less competitive compared to the single-loop mutants, implying that hnRNP K does not favor sequences with fewer than two consensus binding sequences. The results of these competition EMSA studies reveal that hnRNP K preferentially utilizes two of the i-motif loops, favoring loops 2 and 3, and recognizes and binds to a third consensus sequence in the fifth CT sequence seven bases downstream of the i-motif.

In Cells, the 5CT Sequence Is Required for Maximal Transcription Activation by hnRNP K. To activate transcription of *MYC* by hnRNP K in living cells, the region in which hnRNP K binds must be denatured.[36] Examination of the *MYC* promoter using a DNase 1 hypersensitivity assay found that this region containing the i-motif can be single-stranded,[51] which is induced by the negative superhelicity generated during transcription activation.[27] To confirm our ex vivo findings in vitro, we developed a two-plasmid reporter assay. The expression plasmid produces the fusion protein VP16/hnRNP K (Figure 5A), a fusion of hnRNP K and the transactivation domain of VP16.[28] The reporter plasmid contains the i-motif DNA sequence upstream of the E1bTATA box. When co-transfected, the fusion protein activates luciferase expression through interaction with the reporter's promoter. This assay allows us to test the ability of hnRNP K to activate luciferase expression using different recognition sequences. Four reporters were designed and cloned with variations of the *MYC* CT elements. We first investigated the ability of hnRNP K to preferentially activate transcription from the 5CT sequence compared to the 4CT sequence (Figure 5B). After 24 h, expression of VP16/hnRNP K significantly increased the luciferase activity 3.1-fold when co-transfected with 5CT-pGL3,

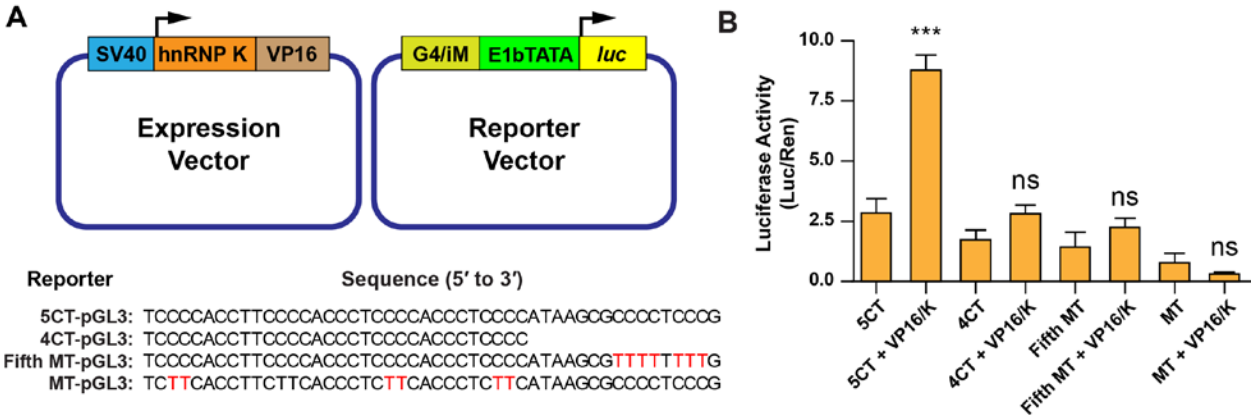


Figure 5. In vitro luciferase assay demonstrates the requirement for the WT 5CT sequence for maximal transcription activation by hnRNP K. (A) Cartoon illustrating the construction of in vitro binding assay plasmids. The expression vector (left) transcribes the fusion protein with hnRNP K as the DNA binding domain and the transactivation domain of VP16. The luciferase reporter vector (right) contains the G-quadruplex/i-motif (G4/iM) DNA binding sequence upstream of the E1bTATA box (transactivation domain motif) and either the 5CT-pGL3, 4CT-pGL3, fifth MT-pGL3, or MT-pGL3 sequence inserted into the G4/iM site. Sequences of different reporter constructs are shown below. (B) Luciferase activity produced by each reporter and empty expression vector compared to the luciferase activity produced by each reporter sequence and VP16/hnRNP K (VP16/K) expression vector, normalized to *Renilla* luciferase.

compared to the 5CT-pGL3 reporter alone (**** P <0.0001). Under the same conditions this activation was not seen with the 4CT-pGL3 reporter. The MT-pGL3 reporter preserves the nucleotide spacer length between the fourth and fifth CT elements but mutates the fifth CT element. This reporter was designed to verify that the fifth CT element is required in addition to the correct nucleotide spacer sequence. As expected, there was no significant difference in luciferase activity between the 4CT-pGL3 and fifth MT-pGL3 reporters in combination with VP16/hnRNP K expression. The final reporter variant, MT-pGL3, inhibited i-motif formation but preserved the fifth CT element and nucleotide spacer length, and as expected, no significant induction of luciferase activity occurred from VP16/hnRNP K expression. Thus we were able to verify the EMSA results that hnRNP K binds optimally to the WT 5CT sequence compared to the truncated 4CT sequence in a cell-based system. Collectively, these results show that the i-motif and the fifth CT element are both required for hnRNP K activation of *MYC* transcription.

The G-Quadruplex and the i-Motif Are Mutually Exclusive in the *MYC* Promoter, and Transcription Factors That Bind to Them Determine the Silencing and Activation of *MYC* Transcription through the NHE III$_1$. Having now established why all 5CT elements are required for transcription activation, alongside our previous understanding of the role of just four of the guanine runs on the opposite strand, we next set out to elucidate how the two secondary DNA structures work in concert to either silence or activate *MYC* transcription. Our first question centered on whether the G-quadruplex and the i-motif could exist at the same time or whether they were mutually exclusive. The contrasting roles of the G-quadruplex and the i-motif suggested to us that they might act together as a molecular switch analogous to the situation in RNA where two different conformational forms are recognized and stabilized by different proteins.[52] Whereas in RNA the single-stranded nature of the template requires mutual exclusivity, in the context of a duplex template there is no immediate requirement that the two conformational forms should be mutually exclusive. A suggestion that this might be the case is from previous studies on the human hINS and hTERT promoters, where mutual exclusivity

between the two structures on opposite strands was demonstrated.[18,53] The mutual exclusivity between the G-quadruplex and the i-motif in the two complementary strands of the *MYC* 4CT and 5CT sequences can be analyzed by a comparison in the formation probabilities of tetraplex structures.[54] Using dual-beam optical tweezers, we tethered the duplex MYC DNA between two particles trapped at two laser foci. By controlling one laser focus using a steerable mirror, we manipulated the position of one particle with respect to another, which changed the tension inside the DNA tether (see Materials and Methods and Supplementary Figure 2). We recorded the tension of the DNA and the distance of the two beads in force-extension (F-X) curves. These F-X curves contained mechanical information, such as the unfolding force and the change-in-contour length,[55] of the unfolding events for DNA G-quadruplexes or i-motifs. With this mechanical information, we then performed a population-based analysis to determine the mutual exclusivity between the G-quadruplex and i-motif in double-stranded DNA (dsDNA).[18,55]

First, using conditions that only allow the formation of a G-quadruplex or i-motif,[53] we estimated, from rupture force histograms, that about 20.6% of G-quadruplexes and 25% of i-motifs form under these two conditions in the 4CT. For the 5CT sequence, these two numbers are 29.0% and 33.5% respectively (Figure 6, A and B, left panels; see Supplementary Figure 3 for F-X curves). In contrast, using 50 mM MES buffer (pH 5.5, 100 mM KCl) that allows the formation of both G-quadruplexes and i-motifs, 25.0% tetraplex (G-quadruplex and i-motif) formation was observed in the 4CT sequence and 33.5% in the 5CT sequence (Figure 6, A and B, right panels). If the formation of one secondary DNA structure does not interfere with the folding of the other, the tetraplex population is expected to be more than the sum of each individual tetraplex (20.6% + 25.0% = 45.6% for the 4CT and 29.0% + 33.5% = 62.5% for the 5CT). Since this value is much higher than observed populations (25% and 31.6% for the 4CT and 5CT sequences respectively), this suggests a negative interaction (or mutual exclusivity) between the G-quadruplex and i-motif in the *MYC* 4CT and 5CT sequences.

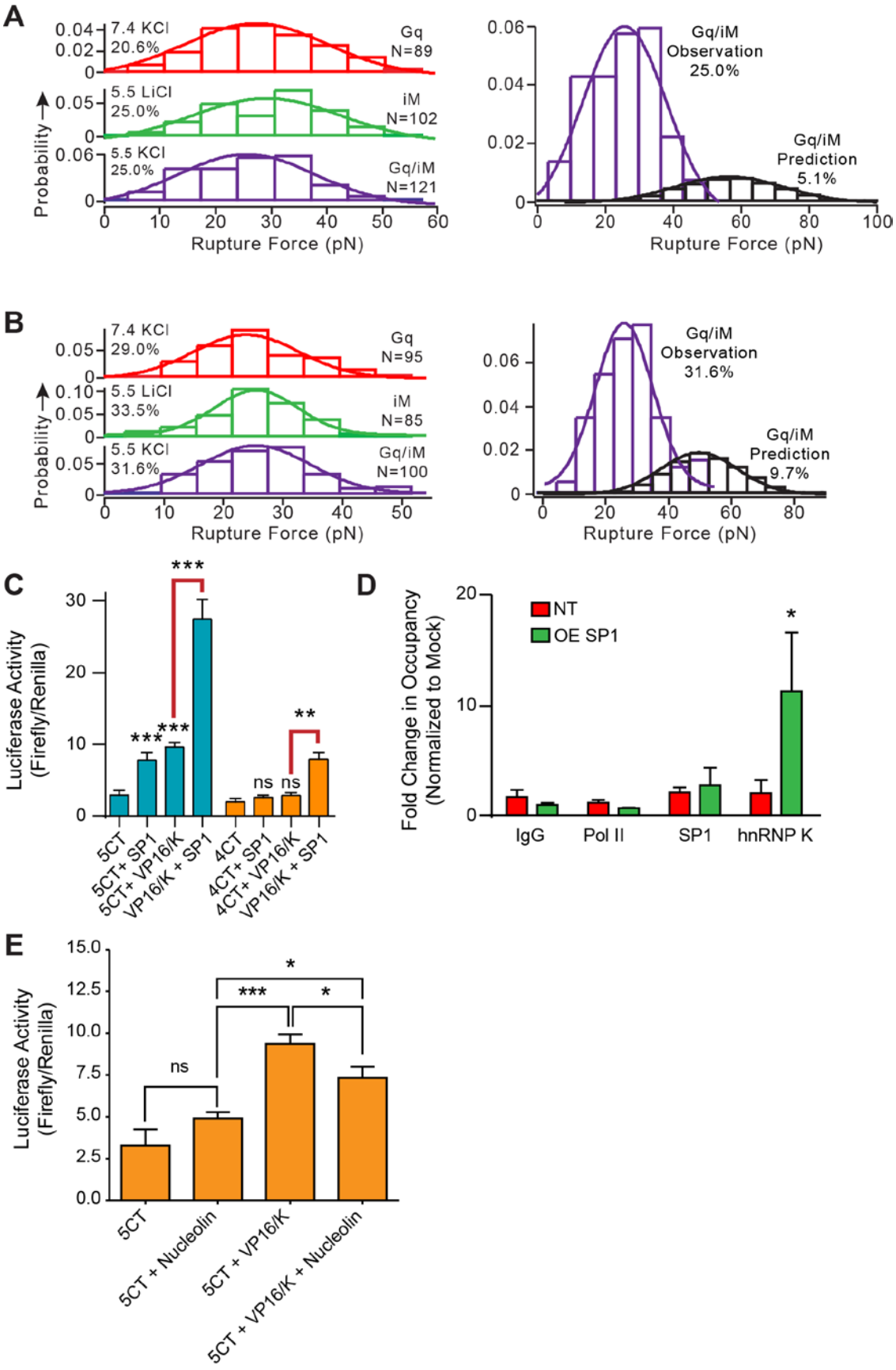

A
7.4 KCl
20.6%
Gq
N=89
5.5 LiCl
25.0%
iM
N=102
5.5 KCl
25.0%
Gq/iM
N=121
Probability
Rupture Force (pN)
Gq/iM
Observation
25.0%
Gq/iM
Prediction
5.1%
Rupture Force (pN)
B
7.4 KCl
29.0%
Gq
N=95
5.5 LiCl
33.5%
iM
N=85
5.5 KCl
31.6%
Gq/iM
N=100
Probability
Rupture Force (pN)
Gq/iM
Observation
31.6%
Gq/iM
Prediction
9.7%
Rupture Force (pN)
C
Luciferase Activity
(Firefly/Renilla)
5CT
5CT+ SP1
5CT+ VP16/K
VP16/K + SP1
4CT
4CT+ SP1
4CT+ VP16/K
VP16/K + SP1
ns
D
Fold Change in Occupancy
(Normalized to Mock)
NT
OE SP1
IgG
Pol II
SP1
hnRNP K
E
Luciferase Activity
(Firefly/Renilla)
ns
5CT
5CT + Nucleolin
5CT + VP16/K
5CT + VP16/K + Nucleolin

Figure 6. The G-quadruplex and i-motif are mutually exclusive in the NHE III_1 sequence, and nucleolin and hnRNP K binding to these structures results in opposite effects on *MYC* expression. Overexpression of SP1-induced negative superhelicity is also required for maximal *MYC* expression. (A) Rupture force histograms of the *MYC* 4CT in 10 mM Tris buffer with 100 mM KCl at pH 7.4 (only G-quadruplex can form, red), in 50 mM MES buffer with 100 mM LiCl at pH 5.5 (only i-motif can form, green), and in 50 mM MES buffer with 100 mM KCl at pH 5.5 (G-quadruplex and i-motif can both form, purple) (left panel). Comparison of the observed (purple) and predicted (black) tetraplex formation rupture force histograms for simultaneous unfolding of the two tetraplexes in 50 mM MES buffer with 100 mM KCl at pH 5.5 (right panel). (B) *MYC* 5CT population analyses of G-quadruplex in 10 mM Tris buffer with 100 mM KCl at pH 7.4 (red), i-motif in 50 mM MES buffer with 100 mM LiCl at pH 5.5 (green), and G-quadruplex and/or i-motif in 50 mM MES buffer with 100 mM KCl at pH 5.5 (purple) (left panel). Predicted probabilities of unfolding both G-quadruplex and i-motif in 50 mM MES buffer supplemented with 100 mM KCl at pH 5.5 (black) (right panel). For comparison, the experimentally observed histogram in this buffer is shown in purple (right panel). (C) Luciferase activity comparing HeLa cells containing the pGL3 reporter containing the 5CT or 4CT promoter with overexpression of SP1, VP16/hnRNP K, or both SP1 and VP16/hnRNP K. (D) ChIP analysis of overexpression of SP1 effects on IgG, pol II, SP1, and hnRNP K occupancy at the *MYC* promoter normalized to the mock transfection. NT = no transfection control; OE SP1 = overexpression of SP1. P-value is <0.05 based on two-way ANOVA. (E). Luciferase activity in HeLa cells showing that nucleolin and hnRNP K compete for the mutually exclusive G-quadruplex and i-motif in the 5CT of the MYC promoter. Luciferase activity produced from the 5CT-pGL5 alone or with just VP16/hnRNP K or nucleolin overexpression was compared to co-overexpression with both VP16/hnRNP K and nucleolin.

It is possible that observed unfolding features represent simultaneous unfolding of the two tetraplexes in the two complementary strands. To evaluate this scenario, we calculated a predicted rupture force histogram (Figure 6, A and B, right panels, black traces) according to a population-based algorithm previously reported.[56] Close inspection of the experimental observation (Figure 6, A and B, right panels, purple traces) showed little population for the simultaneous unfolding (Figure 6, A and B, right panels, black traces), therefore further ruling out the formation of two tetraplexes in the *MYC* sequence. Finally, the absence of two-step unfolding events confirmed the mutual exclusivity of G-quadruplex and i-motif structures in the *MYC* 4CT and 5CT (<0.4%) duplex DNA regions.

We next addressed the question of the roles of SP1, nucleolin, and hnRNP K in controlling *MYC* expression. A number of proteins have been identified to interact with the GC-rich DNA located in the *MYC* promoter, including SP1,[23] nucleolin,[32] hnRNP K,[57] cellular nucleic acid binding protein (CNBP),[58] and a Wilms' suppressor gene product (WT1).[59] In TATA-less promoters, SP1 has been identified as an initiator of basal transcription by recognition of double-stranded GC-rich DNA[60] that can induce negative superhelicity, giving rise to ssDNA that can form a G-quadruplex or i-motif.[61] Nucleolin, a G-quadruplex-binding protein, suppresses transcription,[32] whereas CNBP binds to the single-stranded guanine-rich strand, preventing the formation of the G-quadruplex[62] and enabling i-motif formation. Studies reported here show that hnRNP K recognizes the i-motif and can then activate gene expression if all 5CT elements are accessible. Similarly to CNBP, WT1 can bind to the C-rich DNA, blocking i-motif formation. Proteins such as NM23-H2 unfold the DNA secondary structures and revert the equilibrium back to a double-stranded state.[31] In contrast to previous studies, our luciferase system allowed us to investigate the effects of other proteins on the binding of hnRNP K to the i-motif and the resulting activation.

We then investigated whether the increased negative supercoiling induced by SP1 would affect the transcription activation by hnRNP K. We addressed this by utilizing our in vitro luciferase system described in Figure 5A and co-expressing both SP1 and hnRNP K vectors. Single expression of SP1 (8-fold, *** P <0.001) and VP16/hnRNP K (10-fold, *** P <0.001) resulted in similar transcription activation when using the 5CT promoter sequence. Significantly, co-overexpression of SP1 and VP16/hnRNP K resulted in an additional 2.9-fold (*** P <0.001) increase in luciferase activity compared to VP16/hnRNP K overexpression alone (Figure 6C). For the 4CT promoter sequence a significant increase in luciferase was not observed with single expression of either SP1 or VP16/hnRNP K. A small increase was observed in the 4CT reporter with co-overexpression of SP1 and VP16/hnRNP K (2-fold, ** P <0.01), but the effect was significantly attenuated when compared to the 5CT promoter. To confirm our observation that SP1 and hnRNP K bind synergistically to the *MYC* promoter, we investigated how SP1 overexpression affects hnRNP K binding to the *MYC* promoter using chromatin immunoprecipitation (ChIP). SP1 overexpression did not cause any significant changes in IgG, pol II, or SP1 occupancy at the *MYC* promoter; however, SP1 did significantly (* P <0.05) increase hnRNP K occupancy by 11-fold when compared to the mock treatment (Figure 6D). No significant changes were observed when comparing the mock treatment to the non-treated (NT) control in the ChIP assay (Figure 6D). On the basis of these results, we believe that overexpression of SP1 continuously primes the promoter to initiate transcription, inducing more negative superhelicity, which consequently enhances melting of upstream elements. The increased size of the melted bubble in duplex DNA in the NHE III_1 facilitates access to the fifth CT element, which then results in formation of a thermodynamically stable complex with hnRNP K.

Last, we determined the effect of co-overexpression of nucleolin with the VP16/hnRNP K fusion protein on luciferase activity using the 5CT element. In comparison with the 5CT alone, VP16/K increases luciferase by 3.1-fold (*** P <0.001), which is then reduced by about 25%

when nucleolin is added simultaneously (* P <0.05) (Figure 6E). This shows that nucleolin is antagonistic to VP16/K, which would be expected if the targets are mutually exclusive and have opposite effects when bound by the two different proteins. The underlying mechanism is a change in population dynamics, through each protein competing for either the silencer element (nucleolin–G-quadruplex) or activation element (hnRNP K–i-motif). In a previous study on VEGF, hnRNP K was also found to activate transcription.[63] However, the parallel observation from this study that nucleolin increases gene expression was not replicated in our study. On the basis of the mutual exclusivity and the co-overexpression of hnRNP K with the other proteins involved in recognition of duplex and the noncanonical DNA, we can propose a model for transcription activation and silencing of *MYC* in the NHE III_1 (see Discussion)

DISCUSSION

The role of secondary DNA structures in the control of gene expression and targeting by small molecules has been almost exclusively confined to studies on the involvement of G-quadruplexes, starting with *MYC* and TMPyP4 more than a decade ago.[34] It was only recently that targeting of the i-motif in the *BCL2* promoter to either inhibit or stimulate gene expression was demonstrated, alongside identification of hnRNP LL, as the protein that binds to i-motif.[16,17] In the case of the *MYC* proximal promoter element, there are at least two *cis*-regulatory regions that can form noncanonical DNA sequences upstream of the P1 and P2 promoters (Figure 1). Both require transcriptionally induced torsional stress to form these structures.[27,64] In addition, proteins that bind to these upstream elements to control gene expression have been identified,[22,25] as well as drugs that target the FUSE element[65] and the G-quadruplex.[42] What was still missing was whether an i-motif structure was an important player on the C-rich strand to modulate *MYC* transcription. It has been shown that hnRNP K binds to the CT elements to activate transcription[28] and that for full transcription to occur, a fifth CT element, located external to those sequences required to form either the G-quadruplex or the i-motif, is necessary.[23] Thus

the initial goals of this research were to determine if hnRNP K had an analogous role to hnRNP LL in recognizing and unfolding the i-motif in the *MYC* promoter to activate transcription and to identify the role of the fifth CT element in transcription activation.

A critical property of the i-motif versus the unstructured form found in a single-stranded form is to provide a scaffold for display of the consequently constrained CCCT recognition elements in the central loop and the 3′ lateral loop for recognition and binding by the KH domains of hnRNP K. This provides a kinetic advantage for the initial binding event. Following initial recognition through these two constrained binding sequences, if the fifth CT element is available, then the dynamic property of the i-motif permits an unfolded form of the i-motif to engage with the third, presumably less constrained CCCT sequence in the fifth CT element to form the thermodynamically stable form of the binary complex between hnRNP K and the 5CT element. Significantly, Br_2 footprinting of the binary complex between the i-motif and hnRNP K shows protection of each of these unique CCCT sequences together with their 3′ flanking sequences (III and IV, V and VI, and VII and VIII) (Figure 3C), providing additional evidence for the modular nature of the sequential binding mechanism.

In this contribution we have demonstrated that the i-motif in the *MYC* promoter is formed exclusively from the 4CT element and is initially recognized most likely through the two lateral loops by hnRNP K. However, this complex with the protein utilizes just two of the three KH domains, and in order to form a stable complex, the fifth CT element has to be accessed by additional torsional stress provided by SP1. Under these conditions, unfolding of the i-motif takes place and the third KH domain is engaged in binding to the 5CT element. In this form the three KH domains of hnRNP K most likely bind to the two most 5′ CCCT sequences located in the central loop and the 5′ lateral loop alongside a similar sequence from the fifth CT element. In this form full transcription activation is achieved, presumably because of the thermodynamically favored binding complex involving all three KH domains. This is directly analogous to the unfolded form of the *BCL2* i-motif, where three RNA recognition domains of hnRNP LL are

engaged. The mutual exclusivity of the G-quadruplex and the i-motif permits us to propose that two different dynamic conformational states exist that depend on the degree of negative superhelicity induced by incremental binding of SP1, which in turn determines the transcription firing rate. At a basal level of activation of *MYC* transcription, where there is sufficient torsional stress to create a dynamic equilibrium between the G-quadruplex and the i-motif formed from the 4CT element, the population dynamics are determined by the relative binding affinities of nucleolin and hnRNP K. Since hnRNP K is only weakly bound to the folded i-motif through two KH domains, the superior binding of nucleolin to the G-quadruplex favors the transcriptionally silenced form (A in Figure 7). However, once sufficient torsional stress is exerted through SP1-induced negative superhelicity to further denature the duplex DNA in the NHE III_1, the fifth CT element is melted and hnRNP K is able to access all three KH binding sites. A stable complex is formed that out-competes the binding affinity of nucleolin, and transcription is fully activated (B in Figure 7). Thus we can now provide a firm mechanistic rationale, based on both mutual exclusivity and the population dynamics of the two different transcriptional states (A and B in Figure 7), to explain the conclusions made by DesJardins and Hay over 20 years ago,[23] that there are two different levels of *MYC* transcription activation determined by different levels of SP1 occupancy (C in Figure 7).

Importantly, the differential effects of SP1 occupancy provide the link to the involvement of the i-motif in the NHE III_1 through its recognition by hnRNP K in control of MYC gene expression. This identifies SP1 as a major factor in the control of *MYC* transcription alongside topoisomerase 1, which is implicated in the control of levels of negative superhelicity.[66] Since precise levels of *MYC* expression are important in the control of cellular growth and proliferation, the FUSE/FBP/FIR system works in concert with the NHE III_1/hnRNP K/nucleolin system.[50] Reassuringly, while admittedly a more artificial system, our results can explain precisely the mechanism by which different levels of SP1 control *MYC* transcription activity in the more

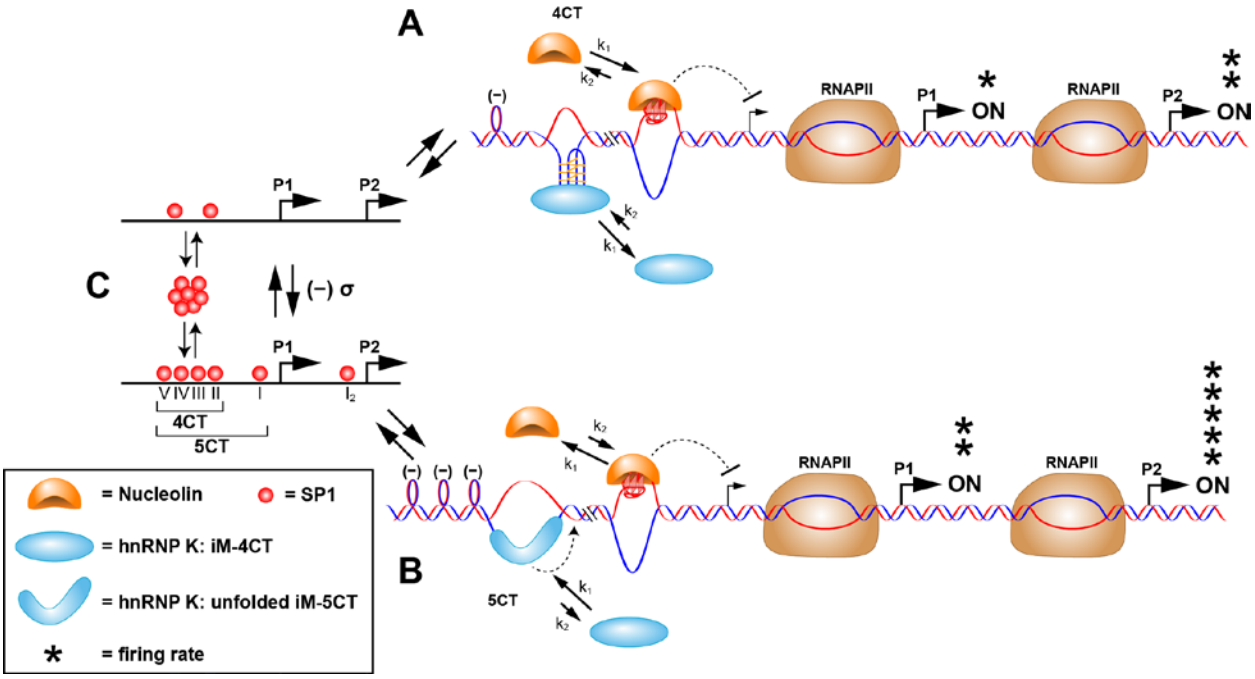


Figure 7. Proposed scheme for the molecular mechanosensor mechanism for differential control of *MYC* expression through the NHE III_1. (A) At low SP1 occupancy of duplex promoter binding sites (as shown in (C), upper), which results in low negative supercoiling, only the 4CT element is accessed and nucleolin binding predominates over hnRNP K binding to the equilibrating G-quadruplex and i-motif, which are mutually exclusive. This results in basal levels of transcription from the P1 and P2 promoters, because hnRNP K only has access to the central and lateral loops of the i-motif that bind to two of the KH domains. (B) At higher occupancy levels of SP1 to the duplex promoter binding sites (as shown in (C), lower), which results in enhanced negative supercoiling, the 5CT element is now fully melted, and hnRNP K forms a thermodynamically stable complex binding through the addition of the CT element. The binding affinity of hnRNP K to the unfolded i-motif and the additional CT element involving three KH domains now exceeds that of nucleolin to the G-quadruplex, and *MYC* expression from P1 and P2 is at peak levels.

biologically relevant system in HeLa cells studied by Desjardins and Hay,[23] thus adding considerable weight to the validity of our conclusions.

The importance of torsional stress in controlling the rate of transcription firing mediated through the NHE III_1 is not without precedent in the *MYC* proximal promoter system. In pioneering work the Levens lab demonstrated that FBP and FIR compete for binding to the noncoding strand in the FUSE element, which is a largely AT-rich element 1.7 kb upstream of the P2 promoter.[50] This element melts into a single-stranded form under torsional stress induced by ongoing transcription and interacts initially through TBP with the p98 subunit of TFIIH to activate transcription. FIR only competes for TBP binding to the p98 subunit of TFIIH when torsional stress is increased to repress transcription. This highly dynamic system was further quantified by Braddock and coworkers.[67] Additional torque results in FIR displacing the TBP from binding to the p98 subunit and canceling the activation effect. Thus this system operates in a real-time way to fine-tune *MYC* transcription. The two-state system we describe for the 4CT and 5CT elements with different secondary DNA structures that compete in a mutually exclusive way for binding to nucleolin and hnRNP K mimics in many ways the FUSE/FBP/FIR system, especially with regard to the fundamental role of torsional stress.

An important lesson from this study, which for the first time takes into account the consequences of a G-quadruplex repressor and i-motif activator elements in the same transcriptional regulatory element, is that what were considered straightforward mutational studies to define the role of either the G-quadruplex or the i-motif cannot be correctly interpreted without considering the consequences on both strands. In addition, our results reported here further emphasize the underlying complexity of these transcriptional regulatory systems involving noncanonical DNA structures, which mimic the intrinsic dynamic chemical behavior more often associated with RNA secondary structures.[68] This behavior is well illustrated by the presence of a tertiary DNA structure in the main *hTERT* G-quadruplex, where the sequential cooperative folding of the G-quadruplex is dependent upon the integrity of a 15-nucleotide

hairpin loop.[55] Likewise, the dynamic conformational flexibility especially prevalent in the C-rich strand, illustrated by equilibrating i-motif and hairpin species in the *BCL2* system, is also reminiscent of the dynamic behavior of RNA.[17]

Finally, a greater appreciation of the complexity of the dual-stranded systems provides new therapeutic opportunities for designing rational combinations of agents that target either secondary DNA structure. For example, because the extent of negative supercoiling determines whether the transcriptional system assumes state A or B shown in Figure 7, topoisomerase inhibitors might well provide useful combination drugs. Indeed, such a precedent exists already, where it was shown that camptothecin, a topoisomerase 1 poison, was synergistic when used in a schedule-dependent way with RHPS4, a G-quadruplex-interactive drug[69] that represses *MYC* expression.[70] While the *MYC* G-quadruplex lacks CpG islands, other G-quadruplexes, such as in the *BCL2* promoter, contain such sites. Methylation of these sites results in stabilization of the G-quadruplexes[71] and therefore might be expected to affect the population dynamics of the G-quadruplex and i-motif and thus the relative binding of *trans*-activating factors.

MATERIALS AND METHODS

Oligonucleotide and Plasmid DNA. All oligonucleotides were obtained from Eurofins MWG Operon and then PAGE purified (8% PAGE with 7 M urea). The firefly (pGL3) and *Renilla* luciferase plasmids (pRL-TK) were purchased from Promega. The GFP, nucleolin, and SP1 overexpression plasmids were provided by Dr. Daekyu Sun (University of Arizona),[25] and the VP16/hnRNP K fusion overexpression plasmid was provided by Dr. David Levens (NIH).[31]

Circular Dichroism. CD analyses were conducted on a Jasco 810 spectropolarimeter using a quartz cell of 1 mm optical path length. Spectra analyses were obtained at an instrument scanning speed of 100 nm/min, with a response time of 1 s, over a wavelength range of 230–350 nm, and recorded three times, averaged, smoothed, and baseline-corrected for signal contributions from buffers. Molar ellipticities for melting curves of the i-motif were

recorded at 286 nm (the λ of the maximum molar ellipticity) over a temperature range of 15–70 °C and then plotted against temperatures for T_m determination. Mutational analysis of the 4CT sequence was done at pH 6.17 and pH 6.47 for the 5CT sequence. The putative i-motif–forming oligonucleotides were prepared at a 5 μM strand concentration in 50 mM Na cacodylate buffer adjusted to the proper pH.

Preparation and End-Labeling of DNA Oligonucleotides. The DNA oligonucleotides were 5′ end-labeled with [γ-^{32}P] ATP upon incubation with T4 polynucleotide kinase for 1 h at 37 °C, then T4 kinase was heat-inactivated for 5 min at 95 °C. Removal of unincorporated radionuclides from the labeled oligonucleotides was achieved using a Bio-Spin 6 chromatography column via centrifugation at 4000 rpm for 5 min. The oligonucleotides were mixed with denaturing loading dye (95% formamide, 20 mM Tris, pH 7.4, and 0.1% bromophenol blue) and further purified by denaturing gel electrophoresis (12% denaturing gel). The bands were excised and eluted in water overnight at room temperature.

Bromine Footprinting. The 5CT oligonucleotide was 5′ end-labeled with [γ-^{32}P] ATP as previously described. The purified 5′ end-labeled 5CT oligomer was incubated in either the presence or absence of recombinant hnRNP K (2 μg) in pH 6.5 binding buffer (40 mM HEPES, 2 mM DTT, 4 mM $MgCl_2$, 2 mM EDTA, 20% glycerol, 200 mM KCl, 0.2 μg/μL BSA, 0.2% Tween-20, and 5 μg/mL poly[d(I-C)]) with molecular bromine formed in situ by mixing an equal molar concentration (0.1 mM and 10 mM) of KBr with $KHSO_5$ for 20 min and terminated by the addition of 60 μL of a 0.6 M sodium acetate and calf thymus DNA (10 mg/mL) solution. Any unreacted bromine was removed in subsequent ethanol precipitation steps. After ethanol precipitation, the DNA pellet was dried and resuspended in 50 μL of a 10% piperidine solution. Samples were heated at 90 °C for 15 min to induce bromination-specific strand cleavage, dried, and resuspended with alkaline sequencing gel loading dye. The bromination-specific strand cleavage was visualized on a sequencing gel (16% PAGE with 7 M urea). A purine sequencing reaction was performed using formic acid and hydrazine for the pyrimidine-specific reaction.

Electrophoretic Mobility Shift Assay. The 4CT and 5CT oligonucleotides were 5′ end-labeled with [γ-^{32}P] ATP as noted above. These oligonucleotides were prepared at 5000 cpm in a binding buffer containing 40 mM HEPES (pH 6.5), 2 mM DTT, 4 mM $MgCl_2$, 2 mM EDTA, 20% glycerol, 200 mM KCl, 0.2 μg/μL BSA, 0.2% Tween-20, and 5 μg/mL poly[d(I-C)] then incubated on ice for 30 min with or without 115 ng hnRNP K. For the cold competition assay, 1–3 μM of unlabeled competitor was added to binding buffer. Protein–DNA complexes were resolved by electrophoresis on a 4% acrylamide gel.

FRET Protein Binding Assay. Dual-labeled 3′ TAMRA/5′ FAM 4CT and 5CT oligonucleotides were synthesized and purchased from Eurofins MWG Operon. The i-motif was induced by heating 500 nM FRET probes in 50 mM Na cacodylate buffer (pH 6.5 or 7.5) to 95 °C for 5 min and slow cooling to room temperature. Either the 4CT or 5CT FRET probe (50 nM) was incubated for 5 min with 50 nM recombinant hnRNP K or no protein in 40 mM HEPES (pH 6.5 or 7.5), 2 mM DTT, 4 mM $MgCl_2$, 2 mM EDTA, 20% glycerol, 200 mM KCl, 0.2 μg/μL BSA, 0.2% Tween-20, and 5 μg/mL poly[d(I-C)], and fluorescent intensity was measured using BioTek Synergy HT at 495 nm (Ex.)/528 (Em.). The data were corrected with background subtraction of corresponding binding buffer with or without protein and then normalized to the no-protein control. Three experimental replicates were done in triplicate, and statistical significance was calculated with T-tests in GraphPad Prism.

Single-Molecule Laser Experiments. *Materials.* All DNA oligonucleotides used in the single-molecule laser experiments were purchased from Integrated DNA Technologies. All chemicals with >99% purity were purchased from VWR. Enzymes used for molecular biology experiments were purchased from New England Biolabs, and surface functionalized beads for the laser tweezers experiments were obtained from Spherotech.

Preparation of DNA constructs. In the optical tweezers experiments, the target DNA fragment was tethered between two dsDNA handles, according to the procedure described previously.[72] Briefly, we first digested the pEGFP vector (Clontech) using two restriction

enzymes, SacI and EagI. After purification using agarose gel, we labeled the SacI end with digoxigenin by terminal deoxynucleotidyl transferase, making a 2690-bp dsDNA handle. The other 2028-bp dsDNA handle was obtained from the PBR322 plasmid (NEB) by PCR and digested with XbaI enzyme. Additionally, one end of this 2028 handle was labeled by a biotinylated PCR primer. Finally, we sandwiched a dsDNA oligo containing the target sequence, which was obtained from annealing of two complementary ssDNA oligos (95–25 °C in 5.5 h), between the 2028-bp and 2690-bp DNA handles through a one-pot ligation reaction using T4 DNA ligase.

Single-molecule force-ramp assay. We first mixed 0.1 ng (3.5×10^{-17} mol) of the DNA prepared above with 1 μL of digoxigenin antibody-coated polystyrene beads (1.87 μM in diameter, 0.5% w/v) in 5–10 μL of three different buffers: 10 mM Tris buffer supplemented with 100 mM KCl at pH 7.4, 50 mM MES buffer supplemented with 100 mM KCl at pH 5.5, and 50 mM MES buffer supplemented with 100 mM LiCl at pH 5.5. The DNA construct and the bead may link together through the digoxigenin-antibody and digoxigenin complex after 30 min incubation, followed by dilution to 800 μL with the same buffer. The diluted mixture was then injected into a 3-channel chamber, ready for further experiments.

The home-built dual-trap 1064-nm laser tweezers to carry out the force-ramp assay at 23 °C has been reported previously.[73,74] Briefly, a mobile laser focus trapped the bead attached with the DNA sample, while the other static laser focus grabbed the other streptavidin-coated bead (2.10 μM diameter; Spherotech). A motorized mirror was used to operate the mobile trap and bring the two beads together to tether the DNA construct between them. The single molecule tether was confirmed in a previously established force ramp assay.[74]

Populations of G-quadruplex or i-motif structures were estimated by their unfolding transitions in the F-X curves up to 60 pN collected during force ramping experiments with a loading rate set at 5.5 pN/s (Supplementary Figure 2). Previously the Mao group reported that a force of up to 60 pN in such a force-ramp assay will completely unfold the tetraplex structures in

dsDNA.[53] An incubation time of 60 s was used to ensure the folding of either G-quadruplex or i-motif species after the unfolding events.[56]

Reporter Construction. E1b-pGL3 was created by the insertion of 5′-GATCTTAGAGGGTATATAATGGATCA (template strand) between the *Bgl*II and the *Hind*III sites of the pGL3 reporter (Promega). 4CT, 5CT, fifth MT, and MT/E1b-pGL3 were created by replacing the *Kpn*1 and *Bgl*II sites with 5′-CTGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCACA; 5′-CTGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCCTCCCA; 5′-CTGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGTTTTTTTTA; 5′-CTGAGTCTCCTCTTCACCTTCTTCACCCTCTTCACCCTCTTCATAAGCGCCCCTCCCA (template strand) respectively. Cloning was verified using the University of Arizona Genetics Core low-volume sequencing facility.

Cell Culture. The HeLa cervix adenocarcinoma cell line was obtained from the American Type Culture Collection. HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Cellgro) supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin. All cell lines were cultured at 37 °C in a humidified atmosphere of 5% CO_2. Cells were counted using a hemocytometer and assessed for viability using trypan blue exclusion prior to use for experimental purposes. HeLa cells were transfected following the TurboFect (Thermo Fisher Scientific) transfection reagent guidelines.

Luciferase Assay. In a 24-well plate, 5×10^4 HeLa cell were plated 24 h prior to transfection (TurboFect, Thermo Fisher Scientific) with 1 μg of each reporter or expression and 10 ng pRL-TK vector (Promega). Samples were then incubated for 24 h and cells washed twice with PBS. Following the Dual-Glo® Luciferase Assay System (Promega) protocol, firefly and *Renilla* luciferase activity was read with a luminometer. Luciferase activity was calculated by obtaining a ratio of firefly luminescence and *Renilla* luminescence.

Chromatin Immunoprecipitation. ChIP analyses were carried out using a modified EZ-ChIP kit (Millipore) from 5×10^6 cells/IP/treatment. At each time point queried, cells were collected, cross linked, and sonicated for 40 cycles of 30 s on/30 s off at a maintained temperature of 4 °C until DNA was 300−500 base pairs. ChIP-quality antibodies for each protein of interest were purchased from AbCam or Millipore. DNA was purified using a QIAquick PCR Purification Kit (Qiagen) before 2× dilution and quantization with qRT-PCR, as described above. Custom TaqMan primers to the NHE III_1 region of the *MYC* promoter and the 3′-UTR of the *MYC* promoter were designed and purchased from ABI; amplification was detected with the Rotor-Gene Q PCR detection system (Qiagen).

Recombinant hnRNP K Protein Production. The cDNA of hnRNP K was purchased from Open Biosystems and subsequently cloned into the Novagen (Merck KGaA) pET28a(+) protein expression vector. Correct sequencing of pET28a-hnRNP K was confirmed by the University of Arizona Functional Genomics Core facility. After sequencing analysis of the pET28a-hnRNP K, the expression construct was transformed into Rosetta-gami™ B (DE3) cells (Novagen), clonal colonies were selected, and transformants were confirmed using PCR with the T7 promoter and T7 terminator primers. pET28-hnRNP K–positive transformants were seeded in Miller's LB broth and grown overnight, then 1 L Miller's LB broth was inoculated with the overnight culture at 1/20 ratio and incubated at room temperature for 2 h. hnRNP K expression was induced by 1 mM IPTG for 6 h at room temperature. Cells were harvested by centrifugation at 3,700 x g for 30 min. Harvested cells were resuspended in a lysis buffer (50 mM NaH_2PO_4 [pH 8.0], 300 mM NaCl, 1% Triton X-100, 1 mg/mL lysozyme [Sigma-Aldrich], and 1× protease inhibitor cocktail [cOmplete™ Protease Inhibitor Cocktail (Roche)]) and underwent 10 cycles of the following: incubation on ice for 15 min, vortexing, and sonication. Cell debris was removed by centrifugation at 14,000 rpm for 30 min at 4 °C, and the supernatant was removed and incubated with HisPur Cobalt Resin (Thermo Fisher Scientific) while rotating for 1 h at 4 °C to allow for the selective binding of histidine-tagged hnRNP K. The resin was washed with a wash

buffer (50 mM NaH_2PO_4 [pH 8.0], 300 mM NaCl, 20 mM imidazole). An elution buffer (50 mM NaH_2PO_4 [pH 8.0], 300 mM NaCl, 250 mM imidazole) was used to separate hnRNP K from the resin. Purified hnRNP K was subjected to buffer exchange into a protein stock buffer with 20 mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 20% glycerol, and 0.05% Tween-20 using a Centricon centrifugal filter (EDM Millipore). Purity of hnRNP K was confirmed by coomassie blue staining while a Bradford assay was performed to determine the concentration of protein. The Arizona Proteomics Consortium at the University of Arizona confirmed the identity of the recombinant protein. The GST-hnRNP K and GST-NKH3 expression constructs, a gift from the Levens lab, were expressed and purified as previously reported.[28]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXX. Three figures, the first showing the importance of the KH domains for binding to the MYC i-motif, the second showing a schematic drawing of the mechanical unfolding experiments, and the third showing F-X curves of different MYC constructs obtained in optical tweezers.

AUTHOR INFORMATION

Corresponding Author

*hurley@pharmacy.arizona.edu

ACKNOWLEDGMENTS

We are grateful to Dr. David Bishop for preparing, proofreading, and editing the final version of the manuscript and figures. This research was supported by the National Science Foundation (CHE-1026532 and CHE-1415883 for HM), the National Institutes of Health (5R01CA153821

and 1R01GM085585 for LHH), and the National Foundation for Cancer Research (VONHOFF-15-01 to for LHH).

REFERENCES

(1) Lutz, W.; Leon, J.; Eilers, M. *Biochim. Biophys. Acta* **2002**, *1602*, 61.

(2) Dang, C. V.; Resar, L. M. S.; Emison, E.; Kim, S.; Li, Q.; Prescott, J. E.; Wonsey, D.; Zeller, K. *Exp. Cell Res.* **1999**, *253*, 63.

(3) Bissonnette, R. P.; Echeverri, F.; Mahboubi, A.; Green, D. R. *Nature* **1992**, *359*, 552.

(4) Amati, B.; Littlewood, T. D.; Evan, G. I.; Land, H. *EMBO J.* **1993**, *12*, 5083.

(5) Frye, M.; Gardner, C.; Li, E. R.; Arnold, I.; Watt, F. M. *Development* **2003**, *130*, 2793.

(6) Dang, C. V.; Le, A.; Gao, P. *Clin. Cancer. Res.* **2009**, *15*, 6479.

(7) Lachman, H. M.; Skoultchi, A. I. *Nature* **1984**, *310*, 592.

(8) Weinstein, I. B. *Science* **2002**, *297*, 63.

(9) Felsher, D. W. *Genes Cancer* **2010**, *1*, 597.

(10) Bernstein, P. L.; Herrick, D. J.; Prokipcak, R. D.; Ross, J. *Genes Dev.* **1992**, *6*, 642.

(11) Delmore, J. E.; Issa, G. C.; Lemieux, M. E.; Rahl, P. B.; Shi, J.; Jacobs, H. M.; Kastritis, E.; Gilpatrick, T.; Paranal, R. M.; Qi, J.; Chesi, M.; Schinzel, A. C.; McKeown, M. R.; Heffernan, T. P.; Vakoc, C. R.; Bergsagel, P. L.; Ghobrial, I. M.; Richardson, P. G.; Young, R. A.; Hahn, W. C.; Anderson, K. C.; Kung, A. L.; Bradner, J. E.; Mitsiades, C. S. *Cell* **2011**, *146*, 904.

(12) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discov.* **2011**, *10*, 261.

(13) Cogoi, S.; Paramasivam, M.; Spolaore, B.; Xodo, L. E. *Nucleic Acids Res.* **2008**, *36*, 3765.

(14) Hsu, S. T.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2009**, *131*, 13399.

(15) Qin, Y.; Hurley, L. H. *Biochimie* **2008**, *90*, 1149.

(16) Kang, H. J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4172.

(17) Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4161.

(18) Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. *Biochemistry* **2016**, *55*, 2291.

(19) Arango, D.; Mariadason, J. M.; Wilson, A. J.; Yang, W.; Corner, G. A.; Nicholas, C.; Aranes, M. J.; Augenlicht, L. H. *Br. J. Cancer* **2003**, *89*, 1757.

(20) Hoffman, B.; Liebermann, D. A. *Oncogene* **2008**, *27*, 6462.

(21) Bazar, L.; Meighen, D.; Harris, V.; Duncan, R.; Levens, D.; Avigan, M. *J. Biol. Chem.* **1995**, *270*, 8241.

(22) Liu, J.; Kouzine, F.; Nie, Z.; Chung, H.-J.; Elisha-Feil, Z.; Weber, A.; Zhao, K.; Levens, D. *EMBO J.* **2006**, *25*, 2119.

(23) DesJardins, E.; Hay, N. *Mol. Cell. Biol.* **1993**, *13*, 5710.

(24) Bentley, D. L.; Groudine, M. *Mol. Cell. Biol.* **1986**, *6*, 3481.

(25) González, V.; Guo, K.; Hurley, L.; Sun, D. *J. Biol. Chem.* **2009**, *284*, 23622.

(26) Simonsson, T.; Pribylova, M.; Vorlickova, M. *Biochem. Biophys. Res. Commun.* **2000**, *278*, 158.

(27) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863.

(28) Tomonaga, T.; Levens, D. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5830.

(29) Dayn, A.; Malkhosyan, S.; Mirkin, S. M. *Nucleic Acids Res.* **1992**, *20*, 5991.

(30) Kouzine, F.; Levens, D. *Front. Biosci.* **2007**, *12*, 4409.

(31) Dexheimer, T. S.; Carey, S. S.; Zuohe, S.; Gokhale, V. M.; Hu, X.; Murata, L. B.; Maes, E. M.; Weichsel, A.; Sun, D.; Meuillet, E. J.; Montfort, W. R.; Hurley, L. H. *Mol. Cancer Ther.* **2009**, *8*, 1363.

(32) González, V.; Hurley, L. H. *Biochemistry* **2010**, *49*, 9706.

(33) Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. *Mol. Cancer Ther.* **2002**, *1*, 565.

(34) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593.

(35) Tomonaga, T.; Levens, D. *J. Biol. Chem.* **1995**, *270*, 4875.

(36) Braddock, D. T.; Baber, J. L.; Levens, D.; Clore, G. M. *EMBO J.* **2002**, *21*, 3476.

(37) Reilly, S. M.; Lyons, D. F.; Wingate, S. E.; Wright, R. T.; Correia, J. J.; Jameson, D. M.; Wadkins, R. M. *Biophys. J.* **2014**, *107*, 1703.

(38) Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 10878.

(39) Sun, D.; Liu, W. J.; Guo, K.; Rusche, J. J.; Ebbinghaus, S.; Gokhale, V.; Hurley, L. H. *Mol. Cancer Ther.* **2008**, *7*, 880.

(40) Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H. *Nucleic Acids Res.* **2007**, *35*, 7698.

(41) Dexheimer, T. S.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2006**, *128*, 5404.

(42) Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. *J. Biol. Chem.* **2011**, *286*, 41018.

(43) Mathur, V.; Verma, A.; Maiti, S.; Chowdhury, S. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1220.

(44) Yang, D.; Hurley, L. H. *Nucleosides Nucleotides Nucl. Acids* **2006**, *25*, 951.

(45) Jin, K. S.; Shin, S. R.; Ahn, B.; Rho, Y.; Kim, S. J.; Ree, M. *J. Phys. Chem. B* **2009**, *113*, 1852.

(46) Vorlíčková, M.; Kejnovská, I.; Bednářová, K.; Renčiuk, D.; Kypr, J. *Chirality* **2012**, *24*, 691.

(47) Phan, A. T.; Mergny, J.-L. *Nucleic Acids Res.* **2002**, *30*, 4618.

(48) Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 17667.

(49) Ross, S. A.; Burrows, C. J. *Nucleic Acids Res.* **1996**, *24*, 5062.

(50) Michelotti, G. A.; Michelotti, E. F.; Pullner, A.; Duncan, R. C.; Eick, D.; Levens, D. *Mol. Cell. Biol.* **1996**, *16*, 2656.

(51) Siebenlist, U.; Hennighausen, L.; Battey, J.; Leder, P. *Cell* **1984**, *37*, 381.

(52) Mercer, T. R.; Mattick, J. S. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300.

(53) Dhakal, S.; Yu, Z.; Konik, R.; Cui, Y.; Koirala, D.; Mao, H. *Biophys. J.* **2012**, *102*, 2575.

(54) Cui, Y.; Koirala, D.; Kang, H.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H. *Nucleic Acids Res.* **2014**, *42*, 5755.

(55) Yu, Z.; Gaerig, V.; Cui, Y.; Kang, H.; Gokhale, V.; Zhao, Y.; Hurley, L. H.; Mao, H. *J. Am. Chem. Soc.* **2012**, *134*, 5157.

(56) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; Balasubramanian, S.; Mao, H. *Nat. Chem.* **2011**, *3*, 782.

(57) Baber, J. L.; Libutti, D.; Levens, D.; Tjandra, N. *J. Mol. Biol.* **1999**, *289*, 949.

(58) Chen, S.; Su, L.; Qiu, J.; Xiao, N.; Lin, J.; Tan, J.-H.; Ou, T.-M.; Gu, L.-Q.; Huang, Z.-S.; Li, D. *Biochim. Biophys. Acta* **2013**, *1830*, 4769.

(59) Han, Y.; San-Marina, S.; Liu, J.; Minden, M. D. *Oncogene* **2004**, *23*, 6933.

(60) Pugh, B. F.; Tjian, R. *Cell* **1990**, *61*, 1187.

(61) Selvam, S.; Koirala, D.; Yu, Z.; Mao, H. *J. Am. Chem. Soc.* **2014**, *136*, 13967.

(62) Qiu, J.; Chen, S.; Su, L.; Liu, J.; Xiao, N.; Ou, T.-M.; Tan, J.-H.; Gu, L.-Q.; Huang, Z.-S.; Li, D. *Biochim. Biophys. Acta* **2014**, *1840*, 2244.

(63) Uribe, D. J.; Guo, K.; Shin, Y. J.; Sun, D. *Biochemistry* **2011**, *50*, 3796.

(64) Kouzine, F.; Liu, J.; Sanford, S.; Chung, H. J.; Levens, D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1092.

(65) Huth, J. R.; Yu, L.; Collins, I.; Mack, J.; Mendoza, R.; Isaac, B.; Braddock, D. T.; Muchmore, S. W.; Comess, K. M.; Fesik, S. W.; Clore, G. M.; Levens, D.; Hajduk, P. J. *J. Med. Chem.* **2004**, *47*, 4851.

(66) Collins, I.; Weber, A.; Levens, D. *Mol. Cell. Biol.* **2001**, *21*, 8437.

(67) Hsiao, H.-H.; Nath, A.; Lin, C.-Y.; Folta-Stogniew, E. J.; Rhoades, E.; Braddock, D. T. *Biochemistry* **2010**, *49*, 4620.

(68) Brown, R. V.; Hurley, L. H. *Biochem. Soc. Trans.* **2011**, *39*, 635.

(69) Leonetti, C.; Scarsella, M.; Riggio, G.; Rizzo, A.; Salvati, E.; D'Incalci, M.; Staszewsky, L.; Frapolli, R.; Stevens, M. F.; Stoppacciaro, A.; Mottolese, M.; Antoniani, B.; Gilson, E.; Zupi, G.; Biroccio, A. *Clin. Cancer. Res.* **2008**, *14*, 7284.

(70) Leonetti, C.; Amodei, S.; D'Angelo, C.; Rizzo, A.; Benassi, B.; Antonelli, A.; Elli, R.; Stevens, M. F.; D'Incalci, M.; Zupi, G.; Biroccio, A. *Mol. Pharmacol.* **2004**, *66*, 1138.

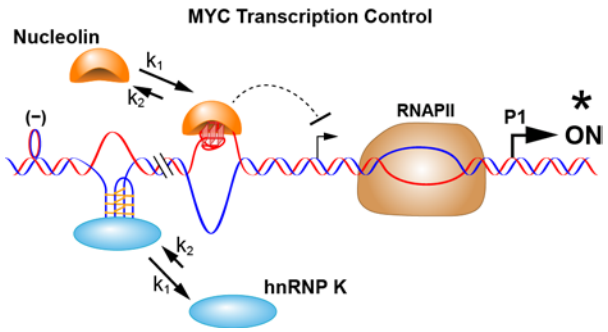
(71) Lin, J.; Hou, J.-Q.; Xiang, H.-D.; Yan, Y.-Y.; Gu, Y.-C.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ou, T.-M.; Huang, Z.-S. *Biochem. Biophys. Res. Commun.* **2013**, *433*, 368.

(72) Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991.

(73) Mao, H.; Luchette, P. *Sensors Actuators B: Chem.* **2008**, *129*, 764.

(74) Yu, Z.; Schonhoft, J. D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2009**, *131*, 1876.

Table of Contents Graphic



Supporting Information

A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the *MYC* Promoter NHE III$_1$

Caleb Sutherland, Yunxi Cui, Hanbin Mao, and Laurence H. Hurley

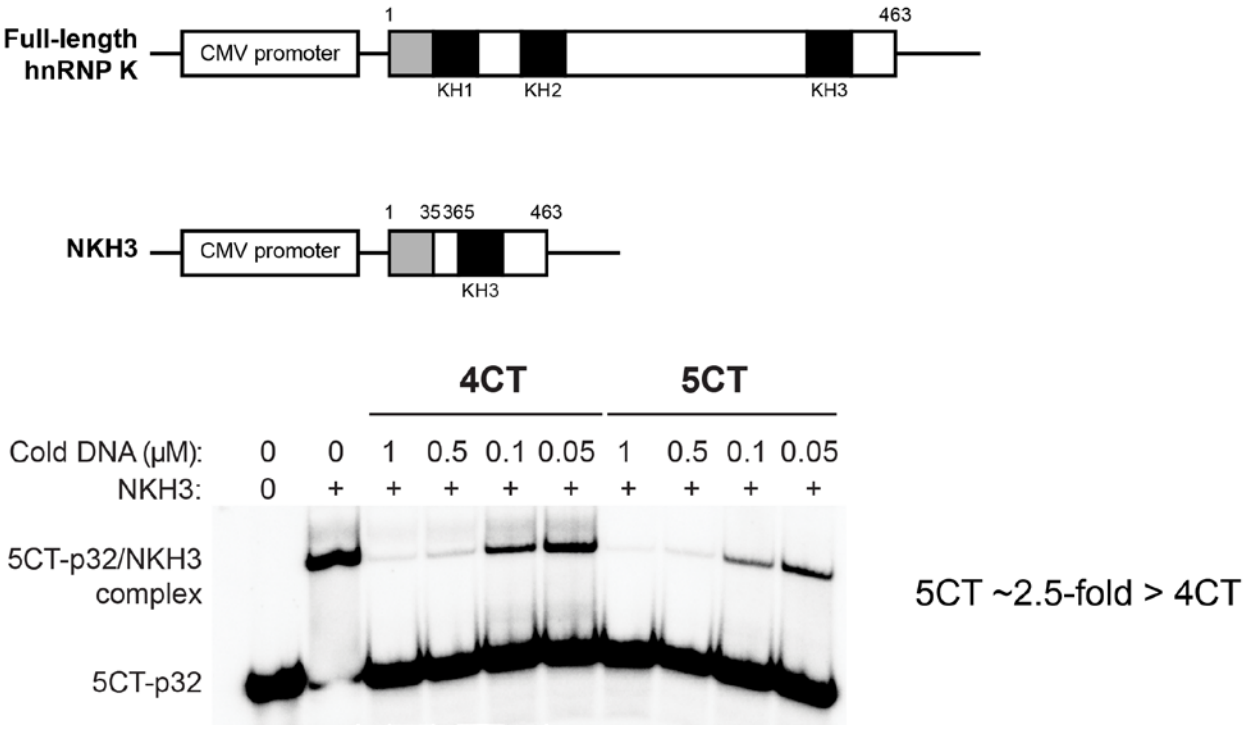


Figure S1. The WT hnRNP K containing three KH domains binds better to the 5CT sequence than the truncated NKH3 protein containing only one KH domain. (Top) Schematic of the two protein expression constructs containing either full-length hnRNP K with three KH domains (top) or NKH3 with only one KH (bottom). (Bottom) Competition EMSA between the 4CT or WT 5CT sequence and NKH3 protein at four different 4CT and 5CT cold-competitor concentrations.

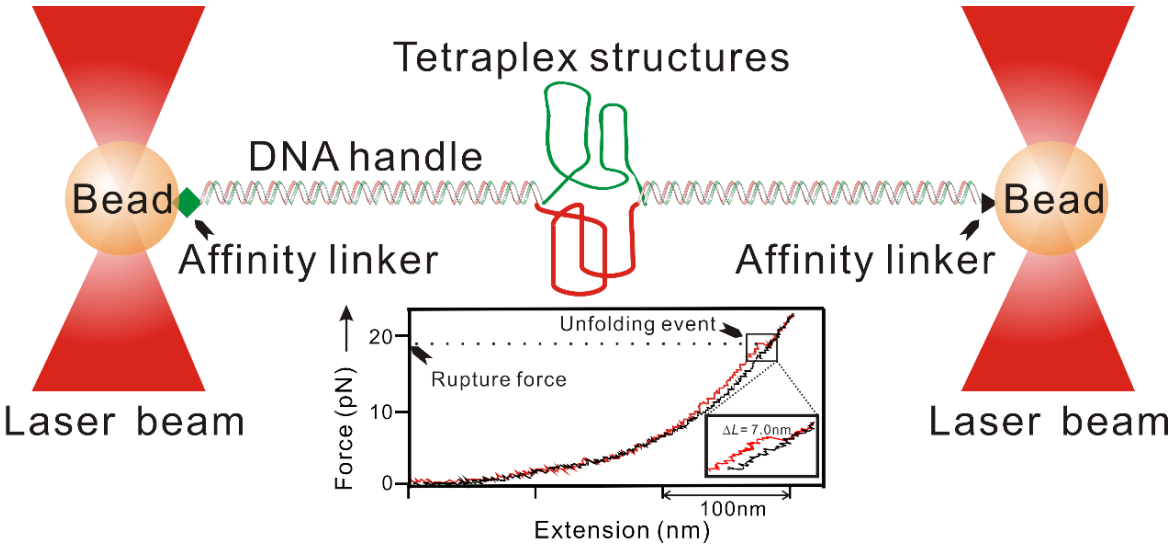


Figure S2. Schematic drawing of the mechanical unfolding experiments. Inset shows a typical F-X curve obtained from the duplex DNA construct *MYC* 4CT (see Materials and Methods for the sequence). Blowup region shows the feature that contains the mechanical information (rupture force and change-in-contour length (Δ*L*) of the unfolding.

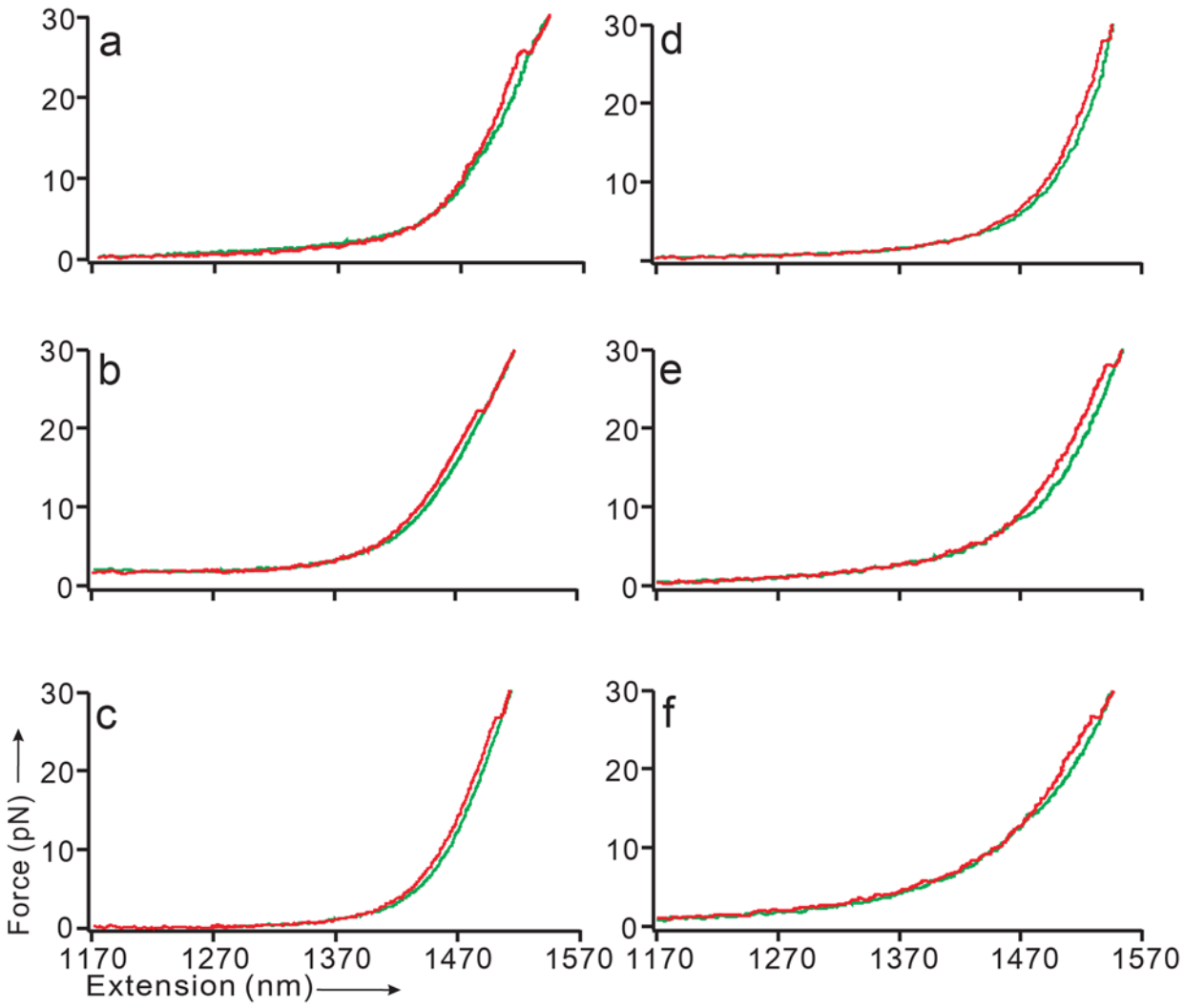


Figure S3. Typical force-extension (F-X) curves of different MYC constructs obtained in optical tweezers. Red and green traces represent stretching and returning curves, respectively. (a) 4CT at pH 5.5 in 50 mM MES buffer with 100 mM KCl. (b) 4CT at pH 5.5 in 50 mM MES buffer with 100 mM LiCl. (c) 4CT at pH 7.4 in 10 mM Tris buffer with 100 mM KCl. (d) 5CT at pH 5.5 in 50 mM MES buffer with 100 mM KCl. (e) 5CT at pH 5.5 in 50 mM MES buffer with 100 mM LiCl. (f) 5CT at pH 7.4 in 10 mM Tris buffer with 100 mM KCl.