DNA G-quadruplex and i-motif structure formation is interdependent in human cells

Jessica J. King[1], Kelly L. Irving[1], Cameron W. Evans[1], Rupesh V. Chikhale[2], Rouven Becker[2], Christopher J. Morris[2], Cristian D. Peña Martinez [3], Peter Schofield[3,4], Daniel Christ[3,4], Laurence H. Hurley[5], Zoë A. E. Waller[2], K. Swaminathan Iyer[1] and Nicole M. Smith[1]*

[1]School of Molecular Sciences, University of Western Australia, Crawley, WA 6009, Australia

[2]School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK,

[3] Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, New South Wales, Australia.

[4]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Kensington, Sydney, New South Wales, Australia.

[5]University of Arizona, College of Pharmacy, 1703 East Mabel Street, Tucson, Arizona 85721, United States

[6]UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX, UK

Supporting Information Placeholder

ABSTRACT: Guanine and cytosine-rich nucleic acid sequences have the potential to form secondary structures such as G-quadruplexes and i-motifs, respectively. We show that stabilization of G-quadruplexes using small molecules destabilizes the i-motifs, and vice versa, indicating these gene regulatory controllers are interdependent in human cells. This has important implications as these structures are predominately considered as isolated structural targets for therapy, but their interdependency highlights the interplay of both structures as an important gene regulatory switch.

Dynamic structural transitions between the common B-DNA and alternative DNA conformations provides an additional layer of regulatory control in gene expression.[1–4] G-quadruplex (G4) and i-motif (iM) are two important classes of non-canonical DNA structures that form within certain guanine and cytosine-rich regions of the human genome, respectively. As iM structures are formed via a stack of intercalating hemiprotonated cytosine base pairs (C+:C), it was initially thought that iM formation required slightly acidic pH, however, it has now been established that these structures form at physiological pH within cellular environments.[5,6] G4 structures are formed from pi-stacked planar G-tetrads, where each G-tetrad consists of four guanine bases held together by Hoogsteen hydrogen bonding, and are further stabilized by physiologically relevant cations.[7–10] G4 and iM folding mechanisms have been used to predict the propensity of their formation across the genome and their overrepresentation in regulatory regions.[5,11] Furthermore, their structural features have been widely exploited as scaffolds for the design of small molecule targeting ligands to stabilize G4 and iM structures and thereby regulate gene expression.[12–16] Most G4s identified to date have been classified as transcriptional suppressors, in contrast, there is increasing evidence that iM structures may play the opposite role in gene regulation in activating gene expression.[1,3,13] Given that these G- and C-rich sequences occur in complimentary strands of duplex DNA, understanding G4 and iM dynamics relative to each other is crucial for developing targeted therapies. Previous in vitro studies of duplex DNA containing complementary G- and C-rich strands have reported that G4 and iM formation are mutually exclusive.[12,13,17] Here, we demonstrate for the first time that inducing and stabilizing G4 or iM simultaneously destabilizes the other in human cells, potentially perturbing dynamic structural transitions across the genome. In particular, we show that targeting one of these non-canonical DNA structures using small molecule stabilizing ligands induces destabilization of the other. Given that these non-canonical structures are important genomic regulatory elements, their dynamic interdependence suggests that it is not only active stabilization but also concomitant destabilization of the other that governs gene regulation.

During transcription local unwinding of duplex DNA facilitated by negative supercoiling stress generates transiently exposed single-strand segments enabling G4 and iM structures to form.[18] In order to establish the propensity of formation of G4 and iM secondary structures in human cells, it was pivotal to comprehensively map G4 and iM formation throughout the entirety of the cell cycle.

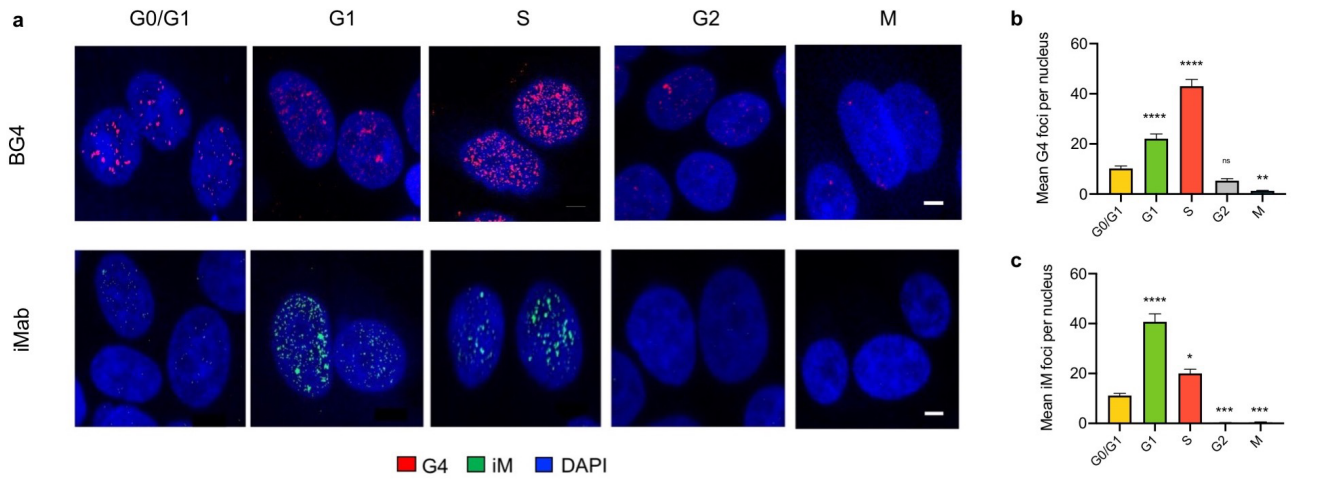

Figure 1. Visualisation and quantification of G4 and iM structures throughout the MCF-7 cell cycle. (a) Representative images of G4 foci (red) and iM foci (green) in MCF-7 nuclei (blue) at each cell cycle stage. Image scale bar, 5 μm. **(b)** Quantification of G4 foci throughout the cell cycle. **(c)** Quantification of iM foci throughout the cell cycle. n = 60 to 107 nuclei, and mean and S E.M. were calculated from two biological replicates. Statistical significance was assessed by one-way ANOVA and Tukey's multiple comparisons test. Significance shown is relative to G0/G1. ***p*<0.01, ****p*<0.001, *****p*<0.0001.

This is essential to map both the prevalence of these secondary structures and the biological roles they play in regulating gene expression. Using recent breakthroughs in the development of structure-specific antibodies scFv BG4[10]and iMab[6], to detect G4 and iM structures, respectively, in human cells, we quantified their formation at the five main cell cycle stages (G_0/G_1, G_1, S, G2, and M) in the MCF-7 cell line (cell cycle arrest was confirmed by flow cytometry, Supplementary Fig. 2). The scFv BG4 antibody[10] was utilized in our study as it does not bind to iM structures, unlike the IgG BG4.[6] We confirmed the propensity for G4 formation was maximal at S phase, during DNA replication, and iM formation was highest at G_1 during high levels of transcription (Fig. 1).[6,10] Not surprisingly, G4 and iM formation were lowest during G_2 and M phase (Fig. 1), when cellular processes such as transcription and replication are restricted, and histone modifications lead to chromatin condensation.[19] This suggests that the number of these secondary structures formed within the genome coincides with chromatin accessibility.[11] Furthermore we validated that mimosine, RO-3306, and colcemid, which were used to arrest the cell cycle, did not affect the stabilisation of G4 or iM structures, confirming that the changes in the number of detectable G4 and iM were indeed related to cell cycle stage (Supplementary Fig. 3).

In order to establish whether treatment of cells with small molecules designed to stabilize G4 and iM would induce a concomitant increase in the number of detectable structures, we treated unsynchronized MCF-7 cells with ellipticine analogues, either a G4 stabilizing ligand (GQC-05, 10 μM) or iM stabilizing ligand (71795, 10 μM) for 1 h (Fig. 2). The efficacy of GQC-05 to

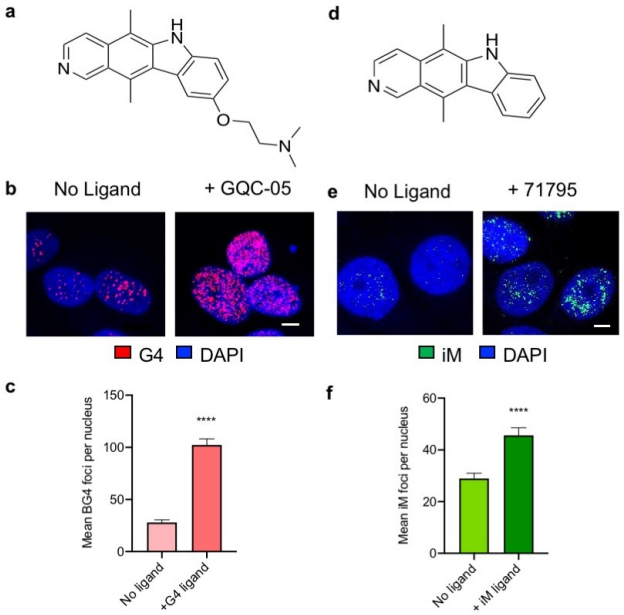

Figure 2. Stabilization of G4 and iM structures within MCF-7 nuclei. (a) Chemical structure of GQC-05 (NSC338258). **(b)** Representative images of G4 foci in MCF-7 nuclei after treatment with GQC-05 (10 μM). **(c)** Stabilization of G4s by GQC-05 increases G4 foci count. **(d)** Chemical structure of 71795 (NSC71795). **(e)** Representative images of iM foci in MCF-7 nuclei after treatment with 71795 (10 μM). **(f)** Stabilization of iMs by 71795 increases iM foci count. Image scale bar, 5 μm. **(c,f)** Quantification of G4 and iM foci in MCF-7 nuclei, n= 138-213. Mean and S.E.M. calculated from two biological replicates. Statistical significance was assessed by t-test, *****p*>0.0001.

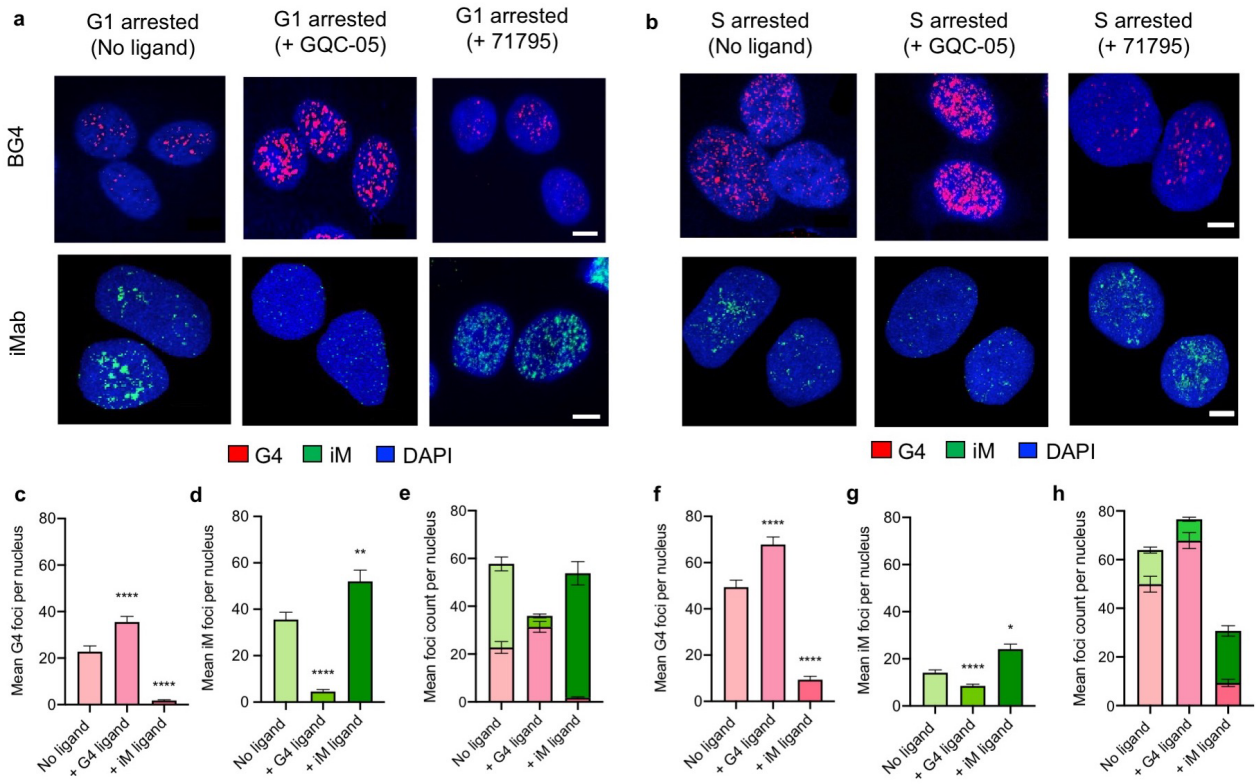

Figure 3. G4 and iM formation is interdependent in human cell nuclei. (a) Immunofluorescence of G4 (red) and iM (green) foci in MCF-7 nuclei (blue) at G1 in the presence of GQC-05 (10 μM) and 71795 (10 μM). **(b)** Immunofluorescence of G4 and iM in S phase arrested MCF-7 nuclei, after treatment with GQC-05 (10 μM) and 71795 (10 μM). **(c-h)** Quantification of G4 and iM foci in G1 **(c-e)** and S **(f-h)** arrested MCF-7 cells. Sample sizes ranged from n= 168 to 204 nuclei from two biological replicates. Statistical significance is shown relative to the no ligand control and was assessed by one-way ANOVA and Tukey's multiple comparisons test. *p<0.1, **p<0.01, ***p<0.001, ****p<0.0001. **(e,h)** Stack plots showing the relative abundance of G4 and iM foci in G1 **(e)** and S **(h)** arrested MCF-7 nuclei.

selectively stabilize G4 DNA has been previously reported.[20] In the case of iM ligand (71795) we performed biophysical analysis to demonstrate a high preference for stabilizing iM DNA structures over G4 and duplex DNA (Supplementary Tables 2-5, Supplementary Figures 5-13). Using FRET melting experiments the ligand-induced change in melting temperature (ΔT_m) for i-motif was found to be much higher (e.g. 29°C for the i-motif from the human telomere) compared to double stranded (2°C) and G-quadruplex DNA (2°C, Supplementary Table 2, Supplementary Figure 5). Similar differences in stabilization properties between i-motif and G-quadruplex and double helical DNA were also observed using circular dichroism spectroscopy (Supplementary Table 3, Supplementary Figures 6-13). Moreover, ligand 71795 was observed to induce folding of i-motif at physiological pH (Supplementary Table 4, Supplementary Figures 5 and 12).

Treatment of MCF-7 cells with GQC-05 and 71795 at 10 μM resulted in statistically significant ($p < 0.0001$) increases in the number of nuclear G4 and iM foci respectively (Fig. 2) confirming ligand-induced stabilization within the cellular environment. Given that both these structural transitions are dynamic, small molecule binding stabilizes the structures, locking them into position.

We next examined the consequences of stabilizing one structure on the prevalence of the other. Given that G4 formation is maximal at early S phase, and iM formation is maximal at G_1, we synchronized MCF-7 cells at these two stages, followed by treatment with GQC-05 (10 μM, 1 h) and 71795 (10 μM, 1 h). In the case of cells arrested in G_1 stage, treatment with GQC-05 resulted in a significant increase in G4 and a decrease in iM formation (Fig. 3c,d,g). Treatment with 71795 resulted in an increase in iM and significant decrease in G4 formation (Fig. 3c,d,g). Similar results were obtained for cells arrested in S phase (Fig. 3e,f,h). This result suggests that during G_1 phase, when there is high propensity towards the formation of iM, the stabilization of G4 structures occurs at the expense of the number of iM structures. Similarly, during S phase, when there is high propensity towards the formation of G4, treatment with the iM-stabilizing ligand triggers the unfolding of G4 structures. This indicates that G4 and iMs are interdependent and act as structural controllers of each other's formation.[12,13] We also treated cells with both GQC-05 and 71795 sequentially after

G4/iM stabilization, as well as adding both ligands simultaneously, and found that these ligands act competitively (Fig. 4).

To verify that these effects were due to the stabilizing ability of the ligands and not compound class, G4-stabilizing ligands that belong to different compound classes, pyridostatin ((*N*,*N*′-bis(quinolinyl)pyridine-2,6-dicarboxamide scaffold)[10] and GSA0932 (Quindoline analogue)[21] were tested and also resulted in reduced iM formation compared the no ligand control (Supplementary Figure 15). To corroborate this relationship further, circular dichroism was conducted of dsDNA containing G4 and iM complementary sequences under conditions that favor both G4 and iM formation, which demonstrated that ligand-induced stabilization of one structure prevents formation of the other (Supplementary Figure 16).

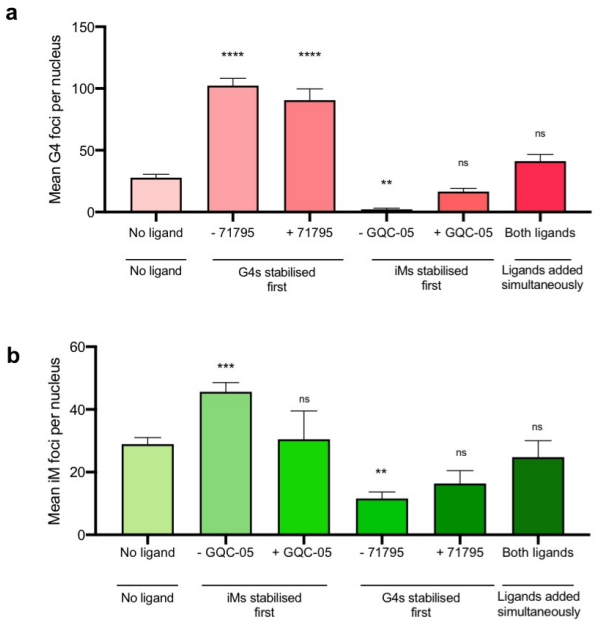

Figure 4. BG4 and iMab quantification in MCF-7 cell nuclei with G4 and iM stabilization. GQC-05 and 71795 ligands were used to stabilize G4 and iM structures, respectively. MCF-7 cell nuclei were treated with GQC-05 (10 μM) after iM stabilization and as well as treatment with 71795 (10 μM) after G4 stabilization. GQC-05 (10 μM) and 71795 (10 μM) were also added simultaneously. **(a)** BG4 quantification in MCF-7 nuclei. **(b)** iMab quantification in MCF-7 nuclei. Mean and S.E.M. are displayed. Sample sizes ranged from n= 47 to 95 nuclei from three technical triplicates. Statistical significance relative to the no ligand control is displayed. ns = nonsignificant, *p<0.1, **p<0.01, ***p<0.001, ****p<0.0001.

In conclusion our findings confirm that while G4 and iM formation in human cells is dependent on chromatin accessibility and cell cycle progression, they are also highly dependent on the formation of each other. In particular we show that stabilization of G4 structures at G1 phase (when iMs are most prevalent) causes a reduction in the number of iM structures, while stabilizing iMs at S phase (when G4s are most abundant) results in a decrease in the number of observed G4 structures. Our results provide insights into the nature of G4 and iM formation and offer a basis for future biological studies and therapeutic targeting for diseases.

ASSOCIATED CONTENT

This information includes experimental procedures, flow cytometry graphs, statistical significance results, immunofluorescence experiments, FRET melting and circular dichroism spectra. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

nicole.smith@uwa.edu.au

Funding Sources

This work was funded by the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC) of Australia, the Cancer Council of Western Australia and the National Breast Cancer Foundation. RVC was funded by Diabetes UK (18/0005820).

ACKNOWLEDGMENT

The authors thank S. Balasubramanian for his gift of pSANG10-3F-BG4. The authors acknowledge the facilities and technical assistance of the Australian Microscopy & Microanalysis Research Facility at the Centre for Microscopy, Characterization & Analysis. The University of Western Australia, a facility funded by the University, State and Commonwealth Governments.

REFERENCES

(1) Kang, H. J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. The Transcriptional Complex between the BCL2 I-Motif and HnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *J Am Chem Soc* **2014**, *136* (11), 4172–4185. https://doi.org/10.1021/ja4109352.

(2) Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. The Dynamic Character of the *BCL2* Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **2014**, *136* (11), 4161–4171. https://doi.org/10.1021/ja410934b.

(3) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proceedings of the National Academy of Sciences* **2002**, *99* (18), 11593–11598. https://doi.org/10.1073/pnas.182256799.

(4) Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and Its Effect on Transcription. *Nucleic Acids Research* **2006**, *34* (9), 2536–2549. https://doi.org/10.1093/nar/gkl286.

(5) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif

Structures at Neutral PH. *Nucleic Acids Res* **2017**, *45* (6), 2951–2959. https://doi.org/10.1093/nar/gkx090.
(6) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nature Chemistry* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.
(7) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Research* **2006**, *34* (19), 5402–5415. https://doi.org/10.1093/nar/gkl655.
(8) Parkinson, G. N. Fundamentals of Quadruplex Structures. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; Royal Society of Chemistry: Cambridge, 2006; pp 1–30. https://doi.org/10.1039/9781847555298-00001.
(9) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57* (1), 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003.
(10) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nature Chemistry* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.
(11) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; Tannahill, D.; Balasubramanian, S. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat Genet* **2016**, *48* (10), 1267–1272. https://doi.org/10.1038/ng.3662.
(12) Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016**, *55* (15), 2291–2299. https://doi.org/10.1021/acs.biochem.6b00016.
(13) Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J Am Chem Soc* **2016**, *138* (42), 14138–14151. https://doi.org/10.1021/jacs.6b09196.
(14) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nature Reviews Drug Discovery* **2011**, *10* (4), 261–275. https://doi.org/10.1038/nrd3428.
(15) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res* **2007**, *35* (2), 406–413. https://doi.org/10.1093/nar/gkl1057.
(16) Kaiser, C. E.; Van Ert, N. A.; Agrawal, P.; Chawla, R.; Yang, D.; Hurley, L. H. Insight into the Complexity of the I-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J Am Chem Soc* **2017**, *139* (25), 8522–8536. https://doi.org/10.1021/jacs.7b02046.
(17) Dhakal, S.; Yu, Z.; Konik, R.; Cui, Y.; Koirala, D.; Mao, H. G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophysical Journal* **2012**, *102* (11), 2575–2584. https://doi.org/10.1016/j.bpj.2012.04.024.
(18) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52* (9), 2863–2874. https://doi.org/10.1021/jm900055s.
(19) Martin, R. M.; Cardoso, M. C. Chromatin Condensation Modulates Access and Binding of Nuclear Proteins. *The FASEB Journal* **2010**, *24* (4), 1066–1072. https://doi.org/10.1096/fj.08-128959.
(20) Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. Demonstration That Drug-Targeted Down-Regulation of MYC in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-Quadruplex. *J. Biol. Chem.* **2011**, *286* (47), 41018–41027. https://doi.org/10.1074/jbc.M111.274720.
(21) Miranti, C. K.; Moore, S.; Kim, Y.; Chappeta, V. R.; Wu, K.; De, B.; Gokhale, V.; Hurley, L. H.; Reyes, E. M. Nucleolin Represses Transcription of the Androgen Receptor Gene through a G-Quadruplex. *Oncotarget* **2020**, *11* (19), 1758–1776.

Table of Contents Image: