

**PROTEIN AND LIGAND INTERACTIONS OF MYC PROMOTER G-  
QUADRUPLEX**

by

**Guanhui Wu**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**



Department of Medicinal Chemistry and Molecular Pharmacology

West Lafayette, Indiana

May 2020

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**  
**STATEMENT OF COMMITTEE APPROVAL**

**Dr. Danzhou Yang, Chair**

Department of Medicinal Chemistry and Molecular Pharmacology

**Dr. Elizabeth J. Tran**

Department of Biochemistry

**Dr. Daniel P. Flaherty**

Department of Medicinal Chemistry and Molecular Pharmacology

**Dr. Tony R. Hazbun**

Department of Medicinal Chemistry and Molecular Pharmacology

**Approved by:**

Dr. Andy Hudmon

*The universe inside our cells*

## ACKNOWLEDGMENTS

This dissertation would not have been possible without the support of many people. First, I would like to express my sincere gratitude to Dr. Danzhou Yang. I appreciate her constant mentoring and guidance. Her passion and enthusiasm for the research were contagious and motivational for me throughout the last several years.

I would also like to express my thanks to the members of my thesis committee, Dr. Elizabeth Tran, Dr. Tony Hazbun, and Dr. Daniel Flaherty, for their invaluable guidance in my scientific pursuit. I am also thankful for the past and present members of the Yang lab who helped me through the journey. I am indebted to Dr. Megan Carver, Dr. Clement Lin, Dr. Buket Onel, Dr. Kaibo Wang, Dr. Jonathan Dickerhoff, Dr. Wenting Liu, and Luying Chen for their invaluable help and companionship.

Finally, I would like to thank my family and girlfriend for their unconditional love, support, and encouragement throughout these years. I am so proud to have a family who always supports me and pushes me towards my dreams.

## TABLE OF CONTENTS

LIST OF FIGURES .....	8
ABSTRACT.....	10
CHAPTER 1. INTRODUCTION .....	11
1.1 G-quadruplex nucleic acids .....	11
1.2 Functions of G-quadruplexes in telomeres.....	13
1.3 Biological functions and structures of G-quadruplex DNA in gene promoters .....	13
1.3.1 Functions of G-quadruplexes in gene promoters .....	13
1.3.2 Negative superhelicity in promoter G-quadruplexes formation.....	16
1.3.3 Biophysical properties and conformations of promoter G-quadruplexes .....	17
1.4 <i>MYC</i> promoter G-quadruplex is a novel anti-cancer target.....	19
1.4.1 Myc protein and Cancer .....	19
1.4.2 Current progress in targeting Myc protein and <i>MYC</i> gene.....	21
1.4.3 Transcriptional regulation of the <i>MYC</i> gene through the promoter G-quadruplex .	24
1.4.4 Structure of the <i>MYC</i> promoter G-quadruplex .....	25
1.4.5 Targeting the <i>MYC</i> promoter G-quadruplex with small molecules .....	27
1.5 Detection of DNA G-quadruplex <i>in vivo</i> .....	27
1.6 Footnotes .....	29
CHAPTER 2. MATERIALS AND METHODS .....	30
2.1 Cell culture, protein preparation, and oligonucleotides.....	30
2.2 Cell viability assay .....	30
2.3 Chromatin immunoprecipitation (ChIP) coupled with quantitative PCR .....	30
2.4 Circular Dichroism (CD) Spectroscopy .....	31
2.5 DMS footprinting .....	31
2.6 Electrophoretic mobility shift assay (EMSA) .....	32
2.7 Enzyme-linked immunosorbent assay (ELISA).....	32
2.8 Fluorescence anisotropy experiments.....	33
2.9 G-quadruplex unfolding assay.....	33
2.10 Luciferase assay.....	34
2.11 Quantitative reverse transcription PCR (qRT-PCR) .....	36

2.12	Western Blot.....	36
2.13	Footnotes .....	36
CHAPTER 3. DDX5 HELICASE RESOLVES G-QUADRUPLEX AND IS INVOLVED IN MYC GENE TRANSCRIPTIONAL ACTIVATION .....		38
3.1	Introduction .....	38
3.2	Results .....	41
3.2.1	Characterization of DDX5 as a proficient G-quadruplex resolvase.....	41
3.2.2	G-quadruplexes are preferred substrates of DDX5 <i>in vitro</i> and <i>in vivo</i> .....	43
3.2.3	ATP hydrolysis is not directly coupled to G-quadruplex-unfolding by DDX5 .....	49
3.2.4	Conformational change of MycG4 DNA induced by DDX5.....	50
3.2.5	DDX5 directly interacts with the MYC promoter G-quadruplex-forming region and activates MYC expression in tumorigenic cells .....	54
3.2.6	DDX5-induced MycG4-unfolding and MYC gene activation are inhibited by G-quadruplex-interactive compounds.....	59
3.2.7	DDX5 inhibition and MycG4 stabilization lead to synergistic lethality in cancer cells .....	63
3.2.8	ChIP analysis of DDX5 and other transcription factors at the MYC promoter .....	65
3.3	Discussion.....	65
3.4	Footnotes .....	67
CHAPTER 4. INDENOISOQUINOLINE TOPOISOMERASE INHIBITORS STRONGLY BIND AND STABILIZE THE MYC PROMOTER G-QUADRUPLEX AND DOWNREGULATE MYC.....		68
4.1	Introduction .....	68
4.2	Results .....	71
4.2.1	Indenoisoquinolines can induce and stabilize MycG4.....	71
4.2.2	Some indenoisoquinolines significantly lower MYC levels in cancer cells .....	73
4.2.3	MYC-inhibiting indenoisoquinolines are strong MycG4 binding ligands.....	76
4.2.4	Molecular docking study of the binding of indenoisoquinoline 5 to MycG4 .....	78
4.2.5	Binding selectivity of MycG4-interactive indenoisoquinolines and 7-azaindenoisoquinolines.....	79
4.2.6	Structure-activity relationship of MycG4 binding by indenoisoquinolines .....	82

4.3	Conclusion.....	84
4.4	Footnotes .....	84
CHAPTER 5. REVIEW OF THE MOLECULAR RECOGNITION OF THE HYBRID-TYPE G-QUADRUPLEXES IN HUMAN TELOMERES.....		86
5.1	Introduction .....	86
5.2	Structural conformation of telomeric G-quadruplexes.....	86
5.3	Molecular recognition of human telomeric hybrid-2 G-quadruplex by epiberberine ....	87
5.4	Molecular recognition of human telomeric hybrid-1 G-quadruplex by Pt-tripod in monomeric and dimeric complexes .....	93
5.5	Insights obtained from the complex structures.....	97
5.6	Footnotes .....	100
CHAPTER 6. A DNA POLYMERASE STOP ASSAY FOR CHARACTERIZATION OF G-QUADRUPLEX FORMATION AND IDENTIFICATION OF G-QUADRUPLEX-INTERACTIVE COMPOUNDS.....		101
6.1	Introduction .....	101
6.2	Materials.....	104
6.3	Methods .....	105
6.4	Notes.....	110
6.5	Footnotes .....	111
CHAPTER 7. CONCLUSION.....		112
REFERENCES .....		115
PUBLICATIONS.....		130
PAPER ABSTRACTS.....		132

## LIST OF FIGURES

Figure 1.1. Schematic illustrations of G-tetrad, and intramolecular and intermolecular G-quadruplexes .....	12
Figure 1.2. Biological implications of targeting telomeric G-quadruplexes using G-quadruplex-interactive ligands .....	14
Figure 1.3. Promoter G-quadruplexes are negative supercoiling-dependent structures .....	18
Figure 1.4. The domain structure and functions of Myc protein in transcription.....	22
Figure 1.5. Schematic illustrations of the <i>MYC</i> gene and the <i>MYC</i> promoter G-quadruplexes ...	23
Figure 1.6. The G-quadruplex in the <i>MYC</i> NHE III <sub>1</sub> region is a transcriptional silencer.....	26
Figure 1.7. Structures of four MycG4-interactive ligands.....	28
Figure 2.1. Schematic diagram of pGL4.10-MYC-WT Pu43 and pGL4.10-MYC-KO Pu43 luciferase reporter constructs .....	35
Figure 3.1. DDX5 actively unfolds MycG4 shown by a FRET-based helicase unfolding assay .	39
Figure 3.2. DDX5 unfolds MycG4 without flanking tails .....	42
Figure 3.3. DDX5 preferentially recognizes and unfolds G-quadruplex structures in an ATP-independent manner .....	44
Figure 3.4. Further characterization of DDX5 as a G-quadruplex binder .....	46
Figure 3.5. The positive correlation between the expressions of <i>MYC</i> and DDX5 in multiple immortal cell lines.....	47
Figure 3.6. DDX5 chromatin binding and <i>MYC</i> promoter interaction .....	48
Figure 3.7. MycG4 unfolding by DDX5 occurs in the absence of ATP.....	52
Figure 3.8. DMS footprinting and CD spectroscopy showing the conformation change of MycG4 induced by DDX5 .....	53
Figure 3.9. DDX5 regulates <i>MYC</i> transcriptional activity in a G-quadruplex-dependent manner	55
Figure 3.10. Normalized data from Figure 3.9E.....	57
Figure 3.11. The effects of DDX5 depletion using two different shRNAs in 22Rv1 prostate carcinoma cells.....	58
Figure 3.12. G-quadruplex-interactive compounds inhibit G-quadruplex-binding and -unfolding of DDX5.....	60
Figure 3.13. G-quadruplex-interactive molecules, such as TMPyP4, can disrupt transcriptional activation of DDX5 on <i>MYC</i> and lead to cancer cell death .....	61

Figure 3.14. G4-interactive molecule TMPyP4 can disrupt transcriptional activation of DDX5 on <i>MYC</i> .....	62
Figure 3.15. <i>MYC</i> transcriptional regulation via MycG4 in cancer cells.....	64
Figure 4.1. Chemical structures of indenoisoquinoline topoisomerase I inhibitors in Phase I clinical trials and quindoline, as well as the <i>MYC</i> promoter and <i>MYC</i> promoter G-quadruplex .....	69
Figure 4.2. Indenoisoquinolines can induce and stabilize MycG4 .....	72
Figure 4.3. <i>MYC</i> inhibitory activities of indenoisoquinoline analogs .....	74
Figure 4.4. SAR of selected indenoisoquinolines. N.D., not determined.....	75
Figure 4.5. 1D <sup>1</sup> H NMR titrations of MycPu22 DNA with indenoisoquinolines and 7-azaindenoisoquinolines .....	77
Figure 4.6. A model of the 2:1 complex of 7-azaindenoisoquinoline 5:MycG4 suggested by Glide docking in different views .....	80
Figure 4.7. Binding selectivity of MycG4-interactive indenoisoquinolines.....	81
Figure 4.8 Dual inhibition of <i>MYC</i> and topoisomerase I by indenoisoquinolines.....	83
Figure 5.1. The chemical structures of epiberberine and Pt-tripod.....	88
Figure 5.2. The folding pattern of hybrid-2 and hybrid-1 telomeric G-quadruplexes .....	89
Figure 5.3. The specific recognition of human hybrid-2 telomeric G-quadruplex by EPI (PDB ID: 6CCW) .....	90
Figure 5.4. Conversion of human telomeric G-quadruplex structures to the hybrid-2 form induced by EPI.....	92
Figure 5.5. Solution structure of 1:1 and 4:2 Pt-tripod/hybrid-1 G-quadruplex complexes (PDB ID: 5Z80 and 5Z8F) .....	94
Figure 5.6. Arm-groove interactions of Pt-tripod at 5' and 3' sites in the 4:2 Pt-tripod/hybrid-1 complex.....	95
Figure 5.7. Comparison of the 5'-binding modes between EPI and Pt-tripod with human telomeric hybrid-type G-quadruplexes .....	99
Figure 6.1. Overview of DNA polymerase stop assay.....	103

## ABSTRACT

G-quadruplexes (G4s) are non-canonical secondary structures formed in single-stranded guanine-rich nucleic acid sequences, such as those found in oncogene promoters and telomeres. *MYC*, one of the most critical oncogenes, has a DNA G4 (MycG4) in its proximal promoter region that functions as a transcriptional silencer. MycG4 is very stable and the pathological activation of *MYC* requires its active unfolding. However, it remains unclear what drives MycG4 unfolding in cancer cells. We have studied the interactions of DDX5 with the MycG4 at both molecular and cellular levels and discovered that DDX5 actively unfolds the MycG4 and is involved in the *MYC* gene transcriptional regulation, which is described in the first part of this dissertation. DDX5 is extremely proficient at unfolding the MycG4 and ATP hydrolysis is not directly coupled to the G4-unfolding of DDX5. In cancer cells, DDX5 is enriched at the *MYC* promoter and activates *MYC* transcription. G4-interactive small molecules inhibit the DDX5 interaction with the *MYC* promoter and DDX5-mediated *MYC* activation. The second part of this dissertation describes the study of interactions of indenoisoquinoline anticancer drugs with MycG4. The MycG4 transcriptional silencer is a very attractive therapeutic target. Compounds that bind and stabilize the MycG4 have been shown to repress *MYC* gene transcription and are antitumorigenic. Indenoisoquinolines are human topoisomerase I inhibitors in clinical testing. However, some indenoisoquinolines with potent anticancer activity do not exhibit strong topoisomerase I inhibition, suggesting a separate mechanism of action. Our studies show that indenoisoquinolines strongly bind and stabilize MycG4 and lower *MYC* levels in cancer cells. Moreover, the analysis of indenoisoquinoline analogues for their *MYC* inhibitory activity, topoisomerase I inhibitory activity, and anticancer activity reveals a synergistic effect of *MYC* inhibition and topoisomerase I inhibition on anticancer activity. Besides the MycG4, human telomeric G4s are also attractive targets for anticancer drugs due to their ability to inhibit telomere extension in cancer cells. The last part of this dissertation reviews two recent solution structural studies on small molecule complexes with the hybrid-2 telomeric G4 and the hybrid-1 telomeric G4. Structural information of those complexes can advance the design of telomeric G4-interactive small molecules in the cancer therapeutic areas.

## CHAPTER 1. INTRODUCTION

### 1.1 G-quadruplex nucleic acids

G-quadruplexes (G4s) are four-stranded non-canonical secondary structures formed in guanine-rich DNA and RNA sequences. In the early 1900s, Bang first reported the self-assembly phenomenon of guanine, in which high concentrations of guanylic acid/guanosine monophosphate (GMP) form gels in aqueous solution<sup>1</sup>. Almost fifty years later, Gellert and coworkers proposed a cyclic coplanar arrangement of guanine bases after examination of dried GMP gels using X-ray diffraction technique<sup>2</sup>. This coplanar arrangement became known as the G-tetrad plane, in which four guanines connect each other through Hoogsteen hydrogen bonds (**Figure 1.1A**). Physiologically relevant monovalent cations, especially  $K^+$ , stabilize G4 structures by coordinating with eight guanine carbonyl oxygen atoms of the two adjacent G-tetrads<sup>3-5</sup>. Two or more G-tetrad planes stack on top of each other and form a G4 structure (**Figure 1.1B**). A G4 structure can form within a single G-rich strand, or through dimerization and tetramerization of separate G-rich strands<sup>6</sup> (**Figure 1.1B**). The intramolecular G4 structures are of intense interest in the present work because single-stranded DNA or RNA can readily form those structures within the cell.

The biological relevance of those non-Watson-Crick nucleic acid structures was largely ignored until the late 1980s. In 1987, G4 structures were discovered in the G-rich sequences of telomeric DNA<sup>7, 8</sup> and immunoglobulin switch region<sup>9</sup>. Afterwards, those structures have been shown to be involved in many cellular processes, including gene transcription<sup>10, 11</sup>, DNA replication<sup>12</sup>, genome stability<sup>13, 14</sup>, mRNA translation<sup>15, 16</sup>, and RNA processing<sup>17, 18</sup>. In the human genome, G4 structures are more prevalent in the regulatory regions and specifically enriched in the promoters of cancer-related genes<sup>19, 20</sup>, implying their role in the transcriptional regulation. In the human transcriptome, they are enriched in UTRs<sup>21</sup> and also associated with miRNA target sites<sup>22, 23</sup>, mediating RNA processing<sup>17, 18</sup>, and stability<sup>24</sup>.

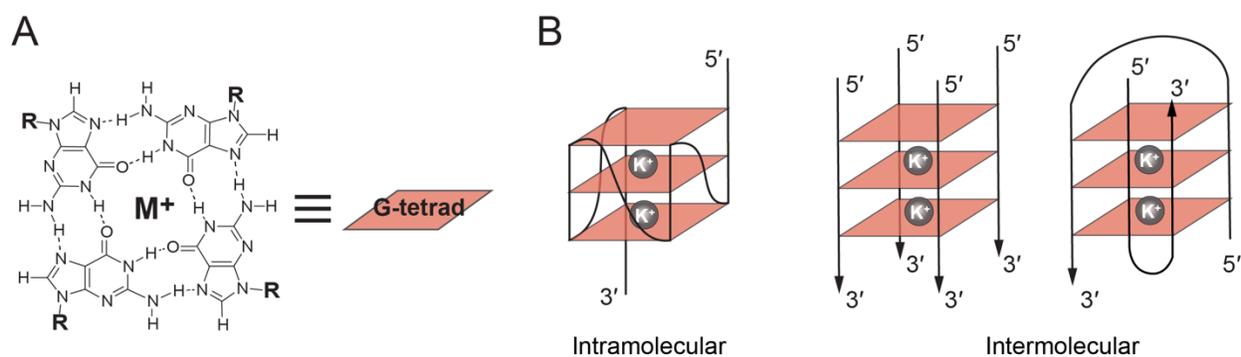


Figure 1.1. Schematic illustrations of G-tetrad, and intramolecular and intermolecular G-quadruplexes

(**A**) Structure of G-tetrad.  $M^+$  stands for monovalent cation. (**B**) Schematics of intramolecular and intermolecular G-quadruplexes consisting of three G-tetrads.  $K^+$  represents potassium cation (Figure is adapted from REF.<sup>25</sup>)

## 1.2 Functions of G-quadruplexes in telomeres

Human telomeres are essential DNA-nucleoprotein complexes capping the termini of chromosomes to protect them from end-to-end fusion and degradation<sup>26, 27</sup>. Telomeres play an important role in genomic stability<sup>28</sup>, aging<sup>29</sup>, and cancers<sup>30</sup>. Human telomeric DNA consists of TTAGGG tandem repeats 5-20 kb in length, terminating in a 30-500 nucleotide single-stranded 3'-overhang<sup>31</sup>. In human somatic cells, telomeric DNA undergoes a progressive shortening with every cell division, known as the end-replication problem<sup>32</sup>. When the telomere length reaches a critical limit, the cell undergoes apoptosis or senescence. However, as one of the hallmarks of cancer, cancer cells counteract the progressive loss of the telomere length to achieve limitless replication potential<sup>30, 33</sup>. One telomere length maintenance mechanism is provided by the reverse transcriptase telomerase (**Figure 1.2**), which is activated in 85-90% of cancer cells<sup>34-36</sup>. Another mechanism is the Alternative Lengthening of Telomeres (ALT) pathway (**Figure 1.2**), which maintains telomere integrity in 10-15% cancer cells that lack detectable telomerase activity<sup>36-40</sup>

G-quadruplexes have been found to form in human telomeres and G-quadruplex formation inhibits the activity of telomerase<sup>41</sup>. The single-stranded G-overhangs of telomeres can spontaneously fold into G-quadruplexes<sup>42</sup>, which can also be dynamically regulated by specific chaperones for their folding<sup>42</sup> and helicases for their unfolding<sup>43</sup>. The therapeutic possibilities of targeting telomeric G-quadruplexes to inhibit telomerase were first reported in 1997<sup>44</sup> and have been actively pursued<sup>45-48</sup>. G-quadruplex-interactive ligands can inhibit telomerase (**Figure 1.2**) and induce apoptosis in cancer cells<sup>49</sup>. In addition, G-quadruplex-interactive ligands were also shown to inhibit the ALT pathway, before which maintains telomere stability in a telomerase-independent manner<sup>50-54</sup>, and can thus avoid drug resistance through inhibiting both telomerase and the ALT pathway<sup>55</sup> (**Figure 1.2**). Therefore, human telomeric G-quadruplexes are considered as attractive cancer-specific drug targets.

## 1.3 Biological functions and structures of G-quadruplex DNA in gene promoters

### 1.3.1 Functions of G-quadruplexes in gene promoters

As a result of research on human telomeric G4s and the cellular consequences of targeting them with small molecules, the interest to G4s has been expanded in the past two decades.

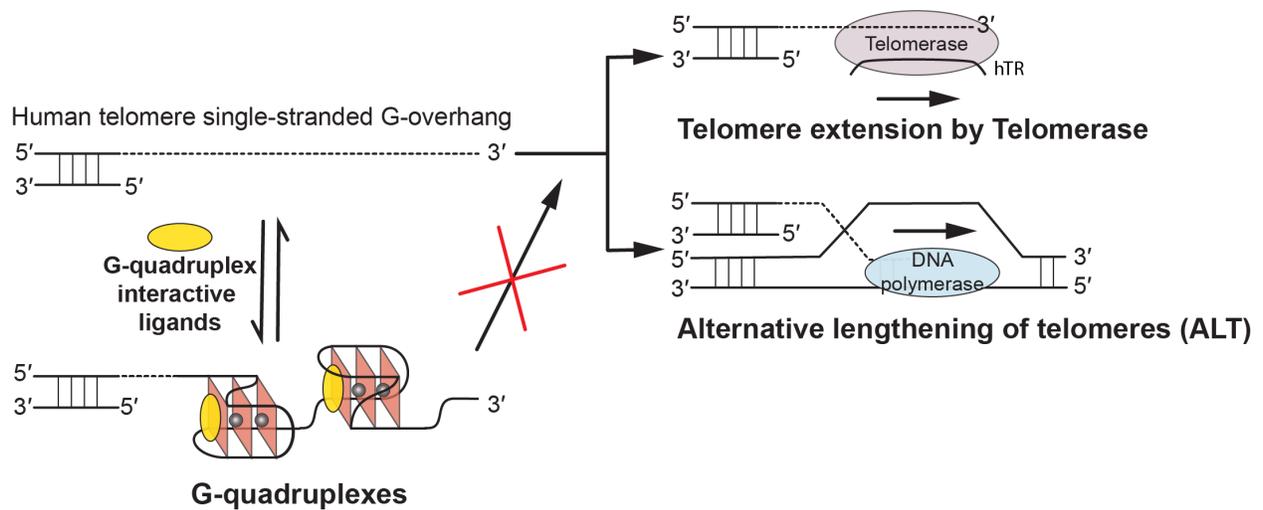


Figure 1.2. Biological implications of targeting telomeric G-quadruplexes using G-quadruplex-interactive ligands

In human tumors, the free single-stranded telomere G-overhang can be extended by telomerase, which reverse transcribes the template region of its RNA subunit (hTR), or by alternative lengthening of telomeres (ALT), which copies telomeric template DNA via homologous recombination. G-quadruplex-interactive ligands can promote the formation and stabilize the telomeric G-quadruplex structures, thus inhibit telomere extension processes which are required for the indefinite growth of human tumors. (Figure is reproduced, with permission, from REF. <sup>56</sup> Copyright © 2019 MDPI)

This was sparked by studies showing telomeric G4-interactive molecules transcriptionally repress *MYC* oncogene through targeting of a G4 structure in the proximal promoter region of the *MYC* gene<sup>10, 57</sup>. The subsequent bioinformatic analysis found about 50% of human genes contain G4s near promoter regions<sup>58</sup>. More convincingly, various G4-sequencing studies using either G4-specific antibody<sup>20</sup> or chemical probes<sup>59</sup> showed that G4 structures are associated with actively transcribed genes and prevalent in regulatory, nucleosome-depleted regions, particularly in the cancer-related genes. These observations suggest chromatin opening allows for transcriptional activation and creates favorable conditions for the G4 formation.

An important question is the role of promoter G4 structures in gene transcriptional regulation. According to the *in vitro* transcription experiments, G4 structures in the template strand of the transcribed region can function as physical obstacles to directly block the movement of transcriptional machinery<sup>60</sup>. Alternatively, G4 motifs present in the non-template strand and can form DNA-RNA hybrid G4s with nascent RNA, leading to transcriptional repression<sup>61</sup>. Another piece of evidence about the functions of G4 motifs comes from the bioinformatic analysis of transcriptional pausing sites in the human genome. G4 motifs near transcription start sites (TSS) positively correlate with proximal-promoter transcriptional pausing sites, implying the role of G4s in the transcriptional pausing<sup>62</sup>. Furthermore, the formation of G4s near TSS has also been found to prevent transcriptional activators binding to gene promoter<sup>63</sup>. For example, a well-established transcription activator, Sp1, typically recognizes a duplex DNA region with the minimal consensus of ‘GGGCGGG’ upstream of TSS<sup>64</sup>. G4 formation in those sequences has been shown to inhibit the binding of Sp1 and to repress gene transcription<sup>63, 65</sup>.

Another piece of evidence suggests the significance of G4 DNA in gene transcription emerging from epigenetics. Although epigenetics is a hotly pursued field, the studies of the relationship between G4s and epigenetic regulation have lagged behind. Recently, new experimental data suggest G4 structures could facilitate chromatin reorganization and epigenetic changes. For example, the ATP-dependent helicase — ATRX is a member of Swi/Snf-like chromatin remodeler, which assists histone H3.3 positioning at telomeric, rDNA and pericentromeric repeats, whereas approximately half of genomic binding sites of ATRX overlap with G4 motifs<sup>66</sup>. ATRX exerts effects on targeted gene expression via the interactions with G-rich repeats<sup>67</sup>, implying the potential linkage between G4s and chromatin remodeling-induced transcriptional regulation. The critical role of CpG methylation in transcription is universally

acknowledged. In 2016, Cree and co-workers found strong interactions between DNA methyltransferase 1 (DNMT1) and DNA G4 structures *in vitro*<sup>68</sup>. A later genome wide-sequencing analysis showed that G4 sites in open chromatin are enriched by DNMT1 and associated with the hypomethylation of CpG islands<sup>69</sup>. Further biochemical analysis revealed those G4 structures inhibit DNMT1 enzymatic activity<sup>69</sup>, suggesting G4 formation may sequester DNMT1 from methylation and inhibit the methylation activity of DNMT1, thus influence gene transcription.

### 1.3.2 Negative superhelicity in promoter G-quadruplexes formation

Telomeric G4 structures are readily available in the single-stranded telomeric DNA overhangs. However, G4 structures in the gene promoter regions are restricted by the duplex nature of genomic DNA. Therefore, an important question is how G4 structures can form in the duplex DNA region and compete with the energetically favorable corresponding Watson-Crick duplex.

The prerequisite of forming G4 structures from duplex DNA is that the regions containing the G4-forming sequences must be nucleosome-free, which allows the separation of the duplex strands. Both bioinformatic<sup>62, 70</sup> and experimental evidence<sup>20, 59</sup> have revealed that these G4-forming regions generally are associated with high nuclease-sensitivity sites and are close to transcription start sites. Furthermore, energy-induced local dissociation/melting of the duplex strand must occur for the G4 formation in those regions. In a relaxed double-helical DNA, two strands twist around the helical axis, while the DNA segment can also be under-twisted (negative supercoiled) during replication or transcription. The negative supercoiling was known to facilitate the formation of alternative DNA structures<sup>71-76</sup>, therefore G4 structures were proposed to be supercoil-dependent structures. However, negative supercoiling alone is not sufficient to drive G4 formation<sup>77</sup>. The G4-formation requires dynamic high-level negative supercoiling torque that is associated with high transcriptional activity (**Figure 1.3**). Those dynamic torques can transmit along linear DNA and be further regulated by DNA-protein interactions<sup>59, 78-80</sup>, resulting in the melting and subsequent G4-formation at susceptible genomic sequences.

The remaining question is why these GC-rich regions are the favored sequences for unwinding rather than other AT-rich sequences, which have a lower energy barrier for the formation of intermediate single-stranded DNA. This can be addressed in two aspects. First, those GC-rich regions can form G4 structures, which provide more metastable intermediates compared

to AT-rich sequence<sup>81, 82</sup>. Second, once those structures formed, they can be trapped out by different proteins that bind and stabilize them<sup>82, 83</sup>, lowering the energy barrier for the formation of intermediate single-strand DNA in GC-rich regions.

### 1.3.3 Biophysical properties and conformations of promoter G-quadruplexes

Unlike the repetitive human telomeric DNA sequence, G4 forming regions in gene promoters are diverse with different numbers of G-tracts and intervening bases<sup>84</sup>. Those regions often contain a continuous stretch of G-rich sequences with more than four G-tracts. Potentially, they can use different combinations of four G-tracts to form intramolecular G4s. However, compared to the telomeric G4s, which contain invariant TAA loops and only have small ground-state energy differences between different hybrid isomers<sup>85</sup>, promoter G-rich sequences often fold into one major and several minor conformations due to the significant energy differences between them as revealed by the distinct thermostabilities.

A few promoter G4 structures have been reported to date, such as *MYC*<sup>86-89</sup>, *VEGFA*<sup>90</sup>, *BCL2*<sup>91, 92</sup>, *KRAS*<sup>93</sup>, *KIT*<sup>94-96</sup>, *RET*<sup>97</sup>, *PDGFRB*<sup>98, 99</sup>, and *PARP1*<sup>100</sup>. In contrast to the hybrid conformations of telomeric G4s, parallel-stranded structures appear to be common for the human promoter G4 structures<sup>6</sup>. In general, they contain three G-tetrads and two 1-nt loops (1st and 3rd), but a variable-length central loop. While they share common parallel folding topology, they can have different capping and loop structures, which are formed by specific flanking sequences and central loops. Most of these promoter G4 structures are highly stable and energetically favorable to form, but their stabilities can be different. For example, In the presence of 90 mM K<sup>+</sup>, the melting temperature of the MycG4 is over 80 °C and cannot be determined because that is very close to the boiling point of water<sup>87</sup>. In comparison, unlike the 2-nt central loop of the MycG4, *VEGFA* primary G4 and the MycG4 (1:6:1) isomer has longer central loops; the loops are stretch over the terminal G-tetrad and form specific capping structures with flanking segments<sup>90</sup>. The melting temperature of the *VEGFA* G4 is shown to be around 75 °C in K<sup>+</sup> solution. The distinct biophysical properties from different

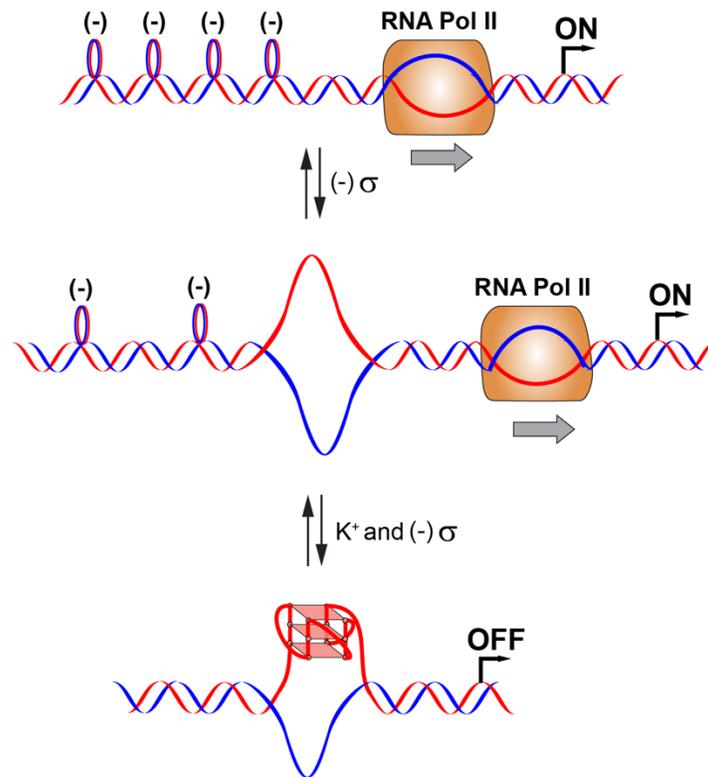


Figure 1.3. Promoter G-quadruplexes are negative supercoiling-dependent structures

When RNA polymerase II (RNA pol II or RNAP II) is moving from left to right (as shown by the grey arrow), the DNA behind the transcription machinery becomes under-twisted (negatively supercoiled;  $(-)\sigma$ ) (top). The supercoiling can diffuse into an upstream region and melt the base pairing in the nucleosome-free region (middle). The melting may subsequently trigger the G-quadruplex formation in the presence of  $K^+$  when the melting passes through the G-quadruple motif (bottom).

G4s indicate the specific flanking sequence and loop size of each parallel G4 together determine the stability of the overall G4 structure<sup>101</sup>. Despite the common parallel folding topology of the promoter G4s, several variant parallel structures have also been found to form in promoter sequences. Instead of using four continuous G-strands, the G-rich sequences in *PDGFRB*<sup>6, 99</sup> and *KIT*<sup>94-96</sup> promoters can adopt unique structures both with a broken strand, to form parallel G4s with three 1-nt propeller loops and an additional lateral loop. More recently, complex G4 systems have been found in the several oncogene promoter regions, such as *BCL2* and *TERT*. *BCL2* promoter contains two G4-forming sequences located in upstream of major P1 promoter region<sup>92</sup>. The most upstream one (a 39-nt G-rich sequence, Pu39) can fold into an equilibrium of a hybrid-type G4<sup>102</sup> and a parallel G4 with a 13-nt central loop<sup>91</sup>. The other 29-nt G-rich sequence just downstream of the Pu39 and immediately upstream of the P1 promoter adopts an equilibrium of two parallel structures<sup>92</sup>; one regular G4 with two 1-nt loops and a 12-nt central loop and another broken-strand G4 with three 1-nt loops and an 11-nt central loop. Both structures contain a unique hairpin (stem-loop duplex) structure in the central long loop. Another complex G4 system was shown in the *TERT* core promoter, which includes a unique end-to-end stacked pair of G4 structure<sup>103-106</sup>. This G4 system consists of a parallel-stranded G4 and another hybrid-type G4 formed with a 26-nt loop, which is likely to adopt a unique hairpin conformation. Importantly, two oncogenic mutations have been found in the G4-forming region of human *TERT* promoter<sup>107</sup>. They are highly recurrent in various cancer types and correlate with higher telomerase levels, indicating the importance of the G-rich sequence on the transcriptional control of the *TERT* oncogene.

## **1.4 MYC promoter G-quadruplex is a novel anti-cancer target**

### **1.4.1 Myc protein and Cancer**

Myc protein is a critical DNA-binding transcriptional factor and is universally expressed in proliferating somatic cells, but absent or low expressed in differentiated cells. In normal cells, the half-life of both Myc protein and mRNA is very short and tightly controlled by mitotic signaling<sup>108, 109</sup>. However, Myc is frequently deregulated and overexpressed in cancer cells. About 450,000 Americans each year are diagnosed with Myc-dependent cancer<sup>110</sup>. The upregulation of Myc in patients is often related to poor prognosis and clinical outcome<sup>111</sup>. The excessive Myc level in the tumor cells can be a result of retroviral promoter insertion, chromosomal

amplification/translocation, enhanced transcription owing to upstream signaling abnormalities, or activation of super-enhancers within the *MYC* gene<sup>112</sup>. Studies in transgenic mice have revealed that Myc inactivation leads to tumor cell redifferentiation. Even brief inactivation of Myc is sufficient to induce sustained tumor regression<sup>113-115</sup>. Therefore, Myc has attracted considerable interest as a potential therapeutic target.

Structure wise, Myc is a member of a class of dimeric transcription factors that contains basic-helix-loop-helix (bHLH) and leucine-zipper (LZ) motifs at the C-terminus and three highly conserved amino acid sequences, known as Myc boxes 1-3 at the N-terminus<sup>112</sup> (**Figure 1.4A**). The LZ motif allows Myc to form a heterodimer with another leucine zipper protein Max (**Figure 1.4B**). The bHLH motif enables it to bind DNA with a preference for a palindromic E-box DNA sequence (CACGTG)<sup>116</sup>.

As a master transcriptional factor, the Myc/Max heterodimer can recruit a chromatin-modifying complex and activate transcription through the interaction with the E-boxes located in the transcriptional regulatory regions of targeted genes<sup>112</sup> (**Figure 1.4C**). The understanding of Myc in transcriptional control has been painfully slow although it was discovered more than three decades ago. Early studies suggested Myc regulates up to one-third of transcriptome<sup>117-119</sup>, while it seems Myc always plays a central role in regulating most of the critical genes in biological processes<sup>129</sup>. Later findings suggested the pleiotropic effects of Myc may be caused by the accumulation of Myc in the promoter regions of actively transcribed genes, leading to transcriptional amplification<sup>130, 131</sup>. Therefore, Myc was suggested being a nonlinear amplifier but not a simple on-off initiator of gene expression. However, this theory has been the subject of intense debate in the last years<sup>120</sup>. A recent analysis of nascent RNA transcripts showed only a few hundred genes are associated with decreased messenger RNA output upon rapid degradation of Myc protein and the vast majority of the transcriptome remaining unaffected<sup>121</sup>. Previous widespread effects might be caused by indirect transcription responses of Myc<sup>121</sup>. New studies provide new critical scientific insights into Myc-dependent genes, but the former finding may still be available and applicable to cancer. The increased cellular concentration of Myc can reduce target-specificity and induce non-specific binding to other E-box motifs<sup>122</sup>, finally leading to a more widespread effect in cancer.

## 1.4.2 Current progress in targeting Myc protein and *MYC* gene

Drug research and development that aims to target Myc protein is very challenging<sup>123</sup>, which is revealed by every steps from work-bench to preclinical studies during the process. First, the functions of Myc are often related to the protein-protein interaction with Max *in vivo* (**Figure 1.4B**). In comparison to oncogenic kinases, the Myc/Max complex lacks enzymatic activity, making *in vitro* characterization of Myc-inhibitors on Myc functions difficult<sup>124</sup>. Second, unlike protein kinases, Myc does not have an obvious binding pocket for small molecules. Myc protein contains largely unstructured regions at the N-terminus, which is not druggable. The C-terminus of Myc is more well-structured, but this region only contains a basic-helix-loop-helix motif, which is hard to be specifically targeted by small molecules (**Figure 1.4B**). As a consequence, the compounds with reported activity against Myc are often associated with low potency, selectivity, and durability when moving to *in vivo* test<sup>125, 126</sup>.

Because direct targeting of the Myc protein has not been achieved so far, targeting the factors that are involved in Myc regulation has been explored as an alternative strategy to treat Myc-driven cancers. Strategies mainly include inhibition of transcription, translation, and activation of *MYC* gene. Among them, targeting transcription of the *MYC* gene has shown great promise. For example, bromodomain containing 4 (Brd4) can regulate the transcription of *MYC* gene through the recruitment of positive transcription elongation factor b (P-TEFb) complex, which can induce transcriptional elongation of paused RNA polymerase II (RNAP II) complex in the proximal promoter region of *MYC* gene<sup>127, 128</sup>. Inhibition of Brd4 using Bromodomain and Extra-Terminal motif (BET) inhibitor, JQ1, has shown potent *MYC* inhibition and anti-cancer effects<sup>129, 130</sup>. However, due to its short half-life, JQ1 is not used in human clinical trials. The second generation of BET inhibitor has been generated and is currently in early-phase clinical trials (ClinicalTrials.gov: NCT01943851, NCT03266159). Another attractive strategy to inhibit *MYC* transcription is to target the *MYC* promoter G4 and associated proteins. It was first demonstrated in 2002 that the G4 structure (MycG4) in the *MYC* promoter is a transcriptional silencer element and that stabilization of the MycG4 leads to transcriptional repression of *MYC*<sup>10</sup>. Since then, numerous MycG4-targeting molecules have been developed, and results are highly encouraging<sup>65</sup>.

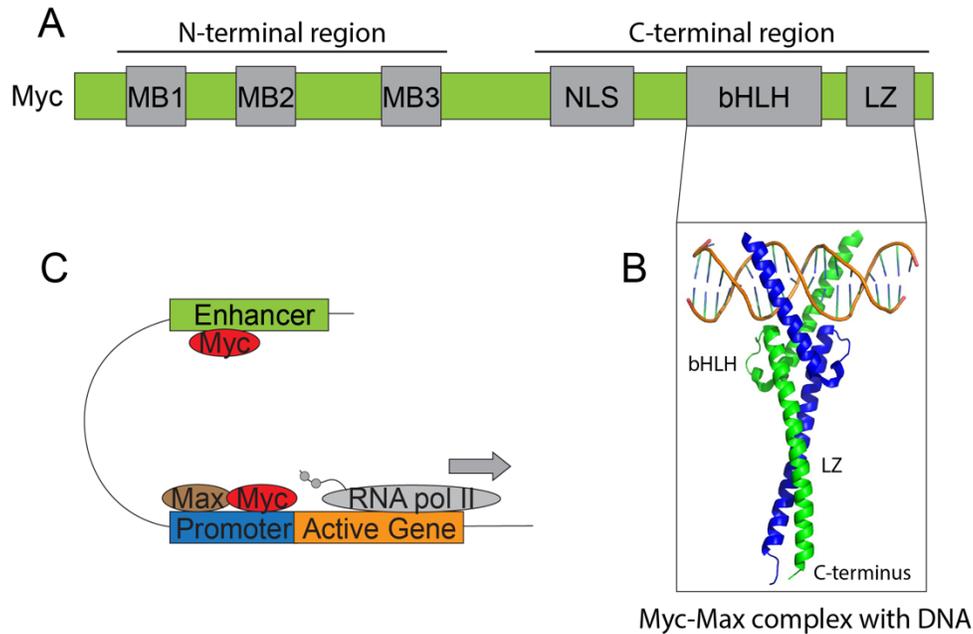


Figure 1.4. The domain structure and functions of Myc protein in transcription

(A) A schematic of the domain structure of Myc protein. N-terminal region serves as a transactivation domain (TAD) and contains three Myc boxes (MB1-3). C-terminal region include nuclear localization signal (NLS), basic-helix-loop-helix (bHLH), and leucine zipper (LZ) motifs. (B) As a transcription factor, Myc partners with Max and binds to the targeted DNA region. (C) Myc has been proposed as a transcriptional amplifier. Myc accumulates in the promoter and enhancer region of the active gene and causes transcriptional signal amplification.

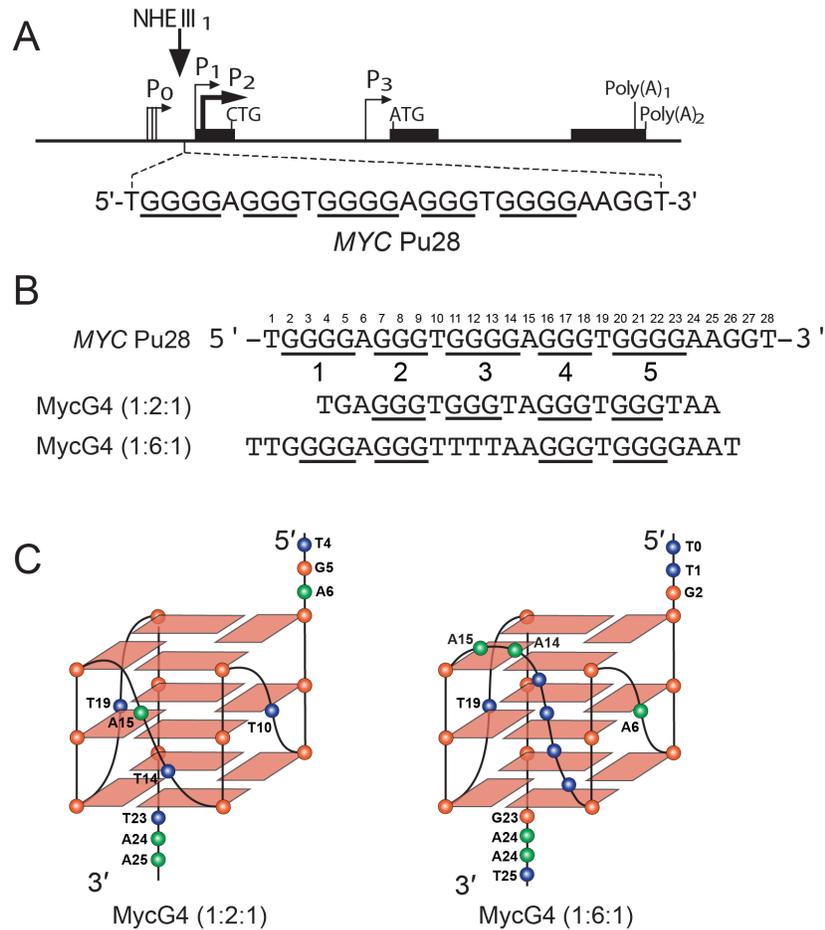


Figure 1.5. Schematic illustrations of the *MYC* gene and the *MYC* promoter G-quadruplexes

(A) Schematic illustration of the promoter of the human *MYC* gene. The full-length G-rich 28-mer G-quadruplex forming sequence (*MYC* Pu28) in the *MYC* promoter NHE III1 region is shown. (B-C) Sequences within the *MYC* Pu28 form a major and a minor G-quadruplex structure in an overlapping region (B). The 1:2:1 loop isomer formed by MycG4 (1:2:1) sequence forms the dominant parallel-stranded *MYC* G-quadruplex structure (C, left; PDB ID: 1XAV). The 1:6:1 loop isomer formed by MycG4 (1:6:1) sequence forms the minor parallel-stranded *MYC* G-quadruplex structure (C, right; PDB ID: 6NEB).

### 1.4.3 Transcriptional regulation of the *MYC* gene through the promoter G-quadruplex

Two tandem principal promoters, termed P1 and P2, govern the transcription of the *MYC* gene<sup>131</sup> (**Figure 1.5A**). They are located 161 bp apart in the core promoter region and integrate many signals, such as transcriptional factors and cis-regulatory elements, to precisely regulate *MYC* expression<sup>132</sup>. In the normal and resting cells, P2 is the pre-dominant promoter producing up to 90% of *MYC* RNA transcripts, while P1 is used to generate about 25% of *MYC* RNA transcripts<sup>131</sup>. However, the transcription of *MYC* can be mis-regulated and predominantly controlled by the P1 promoter in cancer cells<sup>131, 133-136</sup>.

In the *MYC* gene, there are several additional elements that are critical for the transcriptional regulation and can potentially undergo strand separation to form non-duplex structures under dynamic torsional stress<sup>73</sup>. The most characterized elements are the Far Upstream Element (FUSE)<sup>137, 138</sup> and Nuclease hypersensitive element III<sub>1</sub> (NHE III<sub>1</sub>, also known as CT-element)<sup>64, 139, 140</sup>. FUSE is located 1.7 kb upstream of the P2 promoter. Melting of the FUSE element allows the recruitment of FUSE-binding protein (FBP) and FBP-interacting repressor to the FUSE element, establishing a dynamically remodeled loop with transcription factor II H (TFIIH) complex at the P2 promoter to release the torsion stress and regulate transcription<sup>138</sup>. However, the FUSE element itself cannot initiate *MYC* transcription, which requires Sp1 to bind the duplex form of NHE III<sub>1</sub><sup>64</sup> (**Figure 1.6**). NHE III<sub>1</sub> contains five runs of three or more guanines and is located immediately upstream of the P1 promoter, which is important for the *MYC* gene transcription in many cancers<sup>142-146</sup>. The single-strand form of NHE III<sub>1</sub> is again produced as a consequence of dynamic negative supercoiling and is the recognition sites for transcriptional factors, hnRNP k<sup>75, 141, 142</sup> and CNBP<sup>140</sup>, which are known for the transcriptional activation of the *MYC* gene. The G-rich strand of this element is well-known for its ability to form the DNA secondary structure – G-quadruplex<sup>139</sup> (**Figure 1.6**). When G4 is folded in this sequence, it can prevent Sp1 and CNBP binding and consequently silence transcription<sup>10</sup>.

The MycG4 is very stable under physiologically relevant salt conditions with a melting temperature over 85 °C in K<sup>+</sup> solution<sup>143-145</sup>. Complete resolution G4 structures by complementary DNA can take more than three hours *in vitro*<sup>146</sup>. Therefore, the dynamic regulation of *MYC* transcription would require an active unfolding of the MycG4 structure. In this dissertation, I

described that the DDX5 helicase proficiently unfolds the *MYC* promoter G4 and is involved in the *MYC* gene transcriptional activation in human cancer cells.

#### 1.4.4 Structure of the *MYC* promoter G-quadruplex

The NHE III<sub>1</sub> in the *MYC* promoter contains a purine-rich 28-nt strand that has five consecutive runs of three or more guanines (**Figure 1.5A**). To determine the G4 structures formed in the *MYC* promoter, a polymerase stop assay was carried out using the wildtype *MYC* promoter G-rich strand as the template. A single arrest site in the template indicates the G-rich strand adopts a predominant G4 conformation<sup>10</sup>. The G-tracts 2-5 in the *MYC* Pu28 showed a strong protection pattern in chemical footprinting experiments, indicating that these four runs are used for the predominant G4 formation *in vitro*<sup>10</sup> (**Figure 1.5B**). Further studies using biophysical approaches on the MycG4 revealed that it adopts a parallel-stranded fold<sup>86, 87</sup> (**Figure 1.5C, left**).

The NMR and crystal structures of the *MYC* promoter G4 showed that the MycG4 consists of three G-tetrads with four parallel DNA strands, which are linked by three propeller loops<sup>87, 147</sup>. Both T10 and T19 adopt a single loop conformation pointing out toward solvent. In comparison to the single nucleotide loops, the thymine T14 of the two-nucleotide loop also sticks out to the solvent, but the adenine A15 base points toward the G-tetrad groove. Because of the 1:2:1 loop arrangement, the major MycG4 is very stable. Below the 3'-end tetrad, the TAA flanking adopts a well-defined fold-back structure. T23 and A25 form a base pair using Hoogsteen-type hydrogen bonds stacking on the G-tetrad. The T23 base is stack over on the top of the G22 residue, and A25 is on the top of G9 residue. This capping structure is important for the stability of MycG4, and removal of the capping structures significantly reduces the melting temperature.

Interestingly, gel electrophoresis of the NHE III<sub>1</sub> G-rich strand in the presence of K<sup>+</sup> resulted in a major and a minor band<sup>10</sup>, implying the existence of another minor conformation in the same sequence. The subsequent structure analysis revealed that the minor species adopts a parallel G4 structure with 1:6:1 loop arrangement but uses G-tracts 1, 2, 4, and 5<sup>10</sup> (**Figures 1.5B and 1.5C, right**). The solution NMR structure of the 1:6:1 conformer<sup>89</sup> showed that its

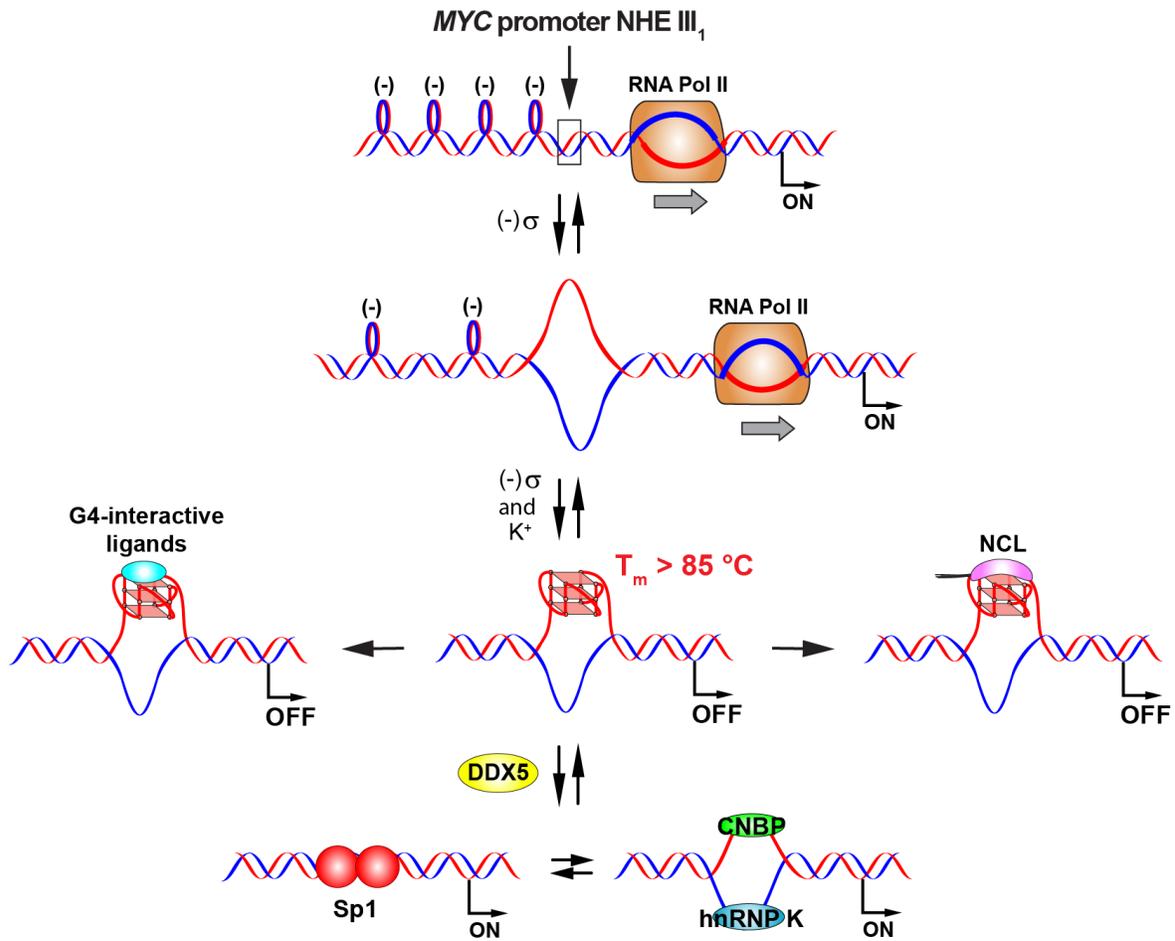


Figure 1.6. The G-quadruplex in the *MYC* NHE III<sub>1</sub> region is a transcriptional silencer. The transcriptional regulation of the *MYC* promoter G-quadruplex involves the resolution of G-quadruplex by DDX5 (this dissertation) and subsequent transcriptional activation by transcription activator Sp1, heterogeneous nuclear ribonucleoprotein (hnRNP K) and cellular nucleic acid binding protein (CNBP). G-quadruplex-interactive ligands or nucleolin (NCL) protein can maintain the *MYC* transcription off.

5'-capping structure is distinctive from the 1:2:1 form, with the 6-nt central loop interacting with the 5'-flanking and forming the capping structure. Additionally, the central solvent-exposed long loop adopts unique conformations. It has been shown many RNA-binding proteins can recognize and target the stem-loop structures. The unique loop conformations and other structural features of MycG4 (1:6:1) from the MycG4 (1:2:1) may provide recognition sites for G4-specific binding proteins<sup>89</sup>.

#### 1.4.5 Targeting the *MYC* promoter G-quadruplex with small molecules

The traditional dogma on optimal MycG4-interactive ligands focused on molecules with symmetric cyclic-fused rings<sup>48</sup>. Targeting MycG4 using a porphyrin derivative TMPyP4 (**Figure 1.7A**) demonstrates tight binding to MycG4 and induction of *MYC* transcriptional repression<sup>10, 57, 144</sup>. Unfortunately, such symmetric fused compounds are nonselective to different G4 structures and also not “drug-like” based on Lipinski’s rule of five<sup>148</sup>. Quindoline (**Figure 1.7B**) and its derivatives, more “drug-like” molecules, were illustrated to stabilize MycG4 and produce *MYC* down-regulation effectively in cells<sup>149, 150</sup>. However, an detailed analysis, which exploits particular translocation features of the *MYC* promoter in the Burkitt’s lymphoma cells, revealed the *MYC* repression caused by quindolines is not directly through the MycG4<sup>151</sup>.

In contrast, an ellipticine derivative (GQC-05, NSC338258; **Figure 1.7C**), which was found by FRET high throughput screening method, showed a directly down-regulates *MYC* transcription through the MycG4 using the Burkitt’s lymphoma cells line<sup>152</sup>. More recently, using a microarray-based screening method resulted in the identification of a selective MycG4 binder (DC-34, **Figure 1.7D**) that directly represses *MYC* through the MycG4 as validated by the Burkitt’s lymphoma cell lines<sup>153, 154</sup>. In this dissertation, we also found indenoisoquinolines topoisomerase inhibitors bind and stabilize the MycG4 and result in *MYC* transcriptional repression.

### 1.5 Detection of DNA G-quadruplex *in vivo*

By using a high-affinity single-chain antibody fragment (scFv), the telomeric DNA of *Stylonychia macronuclei* was shown to adopt G4 structures in 2001<sup>155</sup>. The formation of G4 structures is cell cycle-dependent and controlled by telomere end-binding protein TEBP $\alpha$  and

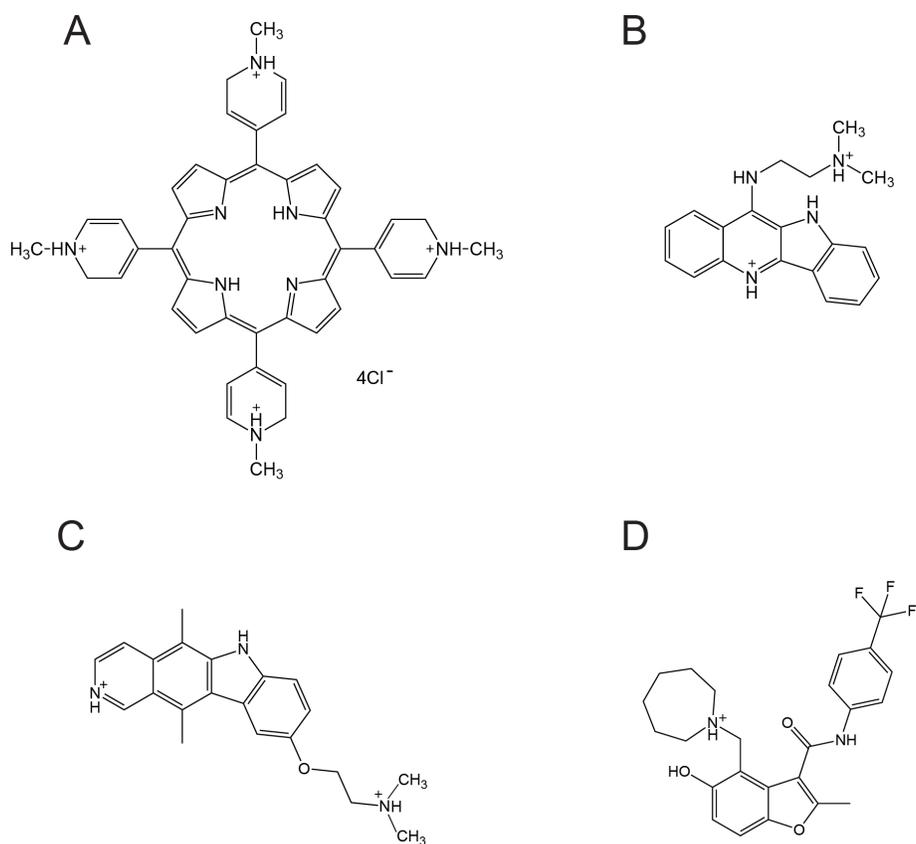


Figure 1.7. Structures of four MycG4-interactive ligands  
**(A)** TMPyP4; **(B)** Quindoline; **(C)** GQC-05; **(D)** DC-34.

TEBP $\beta$ <sup>156, 157</sup>. Subsequently, a few G4-binding proteins such as human ATRX<sup>66</sup> and yeast Pif1 were mapped to G-rich regions in the genome by using ChIP-seq experiments, indicating the G4 formation *in vivo*. More recently, G4-specific antibodies (BG4<sup>158</sup> and 1H6<sup>159</sup>) visualized DNA G4 structures in human cells at both telomeric and non-telomeric regions of human chromosomes. Employing various G4-sequencing methods, ~10,000 G4 structures have been observed in human genome, providing direct evidence and a high resolution map for those structures *in vivo*<sup>20, 59</sup>.

## 1.6 Footnotes

Chapter 1.2 has been previously published as: Wu, G., Chen, L., Liu, W., & Yang, D., Molecular recognition of the hybrid-type G-quadruplexes in human telomeres. *molecules*. 2019; 24(8):1578. (Copyright © 2019 MDPI)

Author contributions: G.W., L.C., W.L., and D.Y. wrote the paper.

## CHAPTER 2. MATERIALS AND METHODS

### 2.1 Cell culture, protein preparation, and oligonucleotides

MCF7/S cells were originally obtained from the Arizona Cancer Center and grown in RMPI 1640 (10-040-CV, Corning) supplemented with 10% fetal bovine serum (35-010-CV, Corning). For DDX5 knockdown in MCF7 cells, a siRNA targeting DDX5 was obtained from Cell Signaling Technology (8627) and a RISC-Free siRNA control was purchased from Dharmacon (D-001600-01). Cells were transfected with a final concentration of 100 nM siRNA by transfection reagent Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. DDX5 transient transfection was performed using Fugene HD (Promega) with pcDNA3.1-DDX5-FLAG<sup>160</sup>. pBV-luc empty vector was used as the control for plasmid transfection. Recombinant MBP-DDX5-GST helicases were expressed and purified as previously described<sup>161</sup>. 5'-biotinylated DNA oligonucleotides, labeled DNA oligonucleotides, and unlabeled RNA oligonucleotides were obtained from Sigma-Aldrich or synthesized using an Expedite 8909 Nucleic Acid Synthesis system (Applied Biosystem, Inc) as previously described<sup>87</sup>.

### 2.2 Cell viability assay

MCF7 cells were seeded at a density of 1,500 cells per well in 96-well plates. 24 hr after seeding, the medium was exchanged for fresh medium containing transfection mixtures (100 nM of DDX5-siRNA or RISC-Free siRNA with Lipofectamine 3000). Next day, the G4-compound solution was directly added to the wells to achieve indicated concentrations. 2 days after treatments, the medium was replaced with fresh medium containing the G4-interactive ligands. Cell viabilities were measured 4 days post treatments using CellTiter 96 Aqueous MTS cell proliferation assay (Promega). The relative cell viability was expressed as a percentage relative to the corresponding transfected cells in the absence of G4-interactive molecules.

### 2.3 Chromatin immunoprecipitation (ChIP) coupled with quantitative PCR

ChIP experiments were carried out using Magna ChIP A/G Kit (17-10085, Millipore) according to the manufacturer's instructions with minor modifications. In brief, untreated or treated  $1 \times 10^7$  MCF7 cells were trypsinized, collected, crosslinked in 10 mL RMPI with 1%

formaldehyde for 9 min at room temperature (rt), and quenched with 0.125 M glycine for 5 min at rt. Cell nuclei were resuspended in 1 mL nuclear lysis buffer and sonicated for total 3 min (1 s 'on', 3 s 'off', 15% amplitude) using a Fisher Scientific Model 500 Sonic Dismembrator with a 1/8" microtip probe. 1/10 of nuclear extract was subject to immunoprecipitation using monoclonal anti-DDX5 antibody (1:100 dilution; 9877, Cell Signaling Technology). Normal rabbit IgG (2729, Cell Signaling Technology) was used as negative control. qPCR was performed on a QuantiStudio 6 Flex Real-Time PCR System (Applied Biosystems) with SYBR Green PCR Master Mix (4309155, Applied Biosystems). Agarose gel electrophoresis was carried out to confirm correct PCR products.

## **2.4 Circular Dichroism (CD) Spectroscopy**

CD spectroscopy was performed on a Jasco J-1100 spectropolarimeter equipped with a thermoelectrically controlled cell holder. Samples were measured in a quartz cell with an optical path length of 1 mm for CD measurements. A blank sample containing only buffer (50 mM Tris-acetate, 2.5 mM MgCl<sub>2</sub>, 100 mM KCl) was used for the baseline correction. Each CD spectroscopy measurement was the averages of three scans collected between 320 and 190 nm at 25 or 30 °C. The scanning speed of the CD instrument was 50 nm/min.

## **2.5 DMS footprinting**

Gel-purified MycG4 Pu28 oligonucleitides (with 5T flanking ends) 5'-end labeled using [ $\gamma$ -<sup>32</sup>P]-ATP in the presence of T4 polynucleotide kinase. The samples were then methylated by treatment with 0.5% (final concentration) DMS, 1  $\mu$ g calf thymus DNA at desired conditions for 7 or 21 mins at room temperature. The reaction was stopped by the addition of  $\beta$ -mercaptoethanol. The methylated DNA was subsequently purified using a phenol/chloroform/isoamyl alcohol mixture to remove proteins. The cleavage at methylated guanines was induced by treatment with 10% piperidine for 18 min at 90 °C. A Speedvac was used to remove the piperidine and two successive water washes. The cleaved products were analyzed on a 16% sequencing (denaturing) PAGE gel.

## 2.6 Electrophoretic mobility shift assay (EMSA)

Gel-purified oligonucleotides were 5'-end labeled using [ $\gamma$ - $^{32}$ P]-ATP in the presence of T4 polynucleotide kinase. 1 nM of labeled oligonucleotides were then incubated with proteins in total 10- $\mu$ L reactions (50 mM Tris-HCl, pH 9, 50 mM KCl, 1 mM MgCl<sub>2</sub>, 0.5 mM DTT, 100  $\mu$ g/mL BSA) on ice for 15 min. Incubation was followed by addition of 5% (v/v) glycerol and resolution by electrophoresis on 4% polyacrylamide gels in cold 0.5X TBE buffer, pH 9. Labeled DNA was visualized by a Typhoon FLA 9500 scanner (GE Healthcare Life Sciences).

## 2.7 Enzyme-linked immunosorbent assay (ELISA)

ELISAs to affinity, specificity of DDX5 to G4 structures were performed using standard capture methods. Briefly, biotinylated oligonucleotides were hybridized to streptavidin-coated plates by incubation overnight at 4 °C. Next day, plates were washed three times with washing buffer (25 mM Tris, pH 7.4, 100 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.05% Tween 20) to remove unbound oligonucleotides and then proteins were captured by incubating the plates with 50 nM of DDX5 at 4 °C for 2 hours. After the incubation, all unbound proteins were removed by washing the plates 3 times with binding buffer (25 mM Tris, pH 7.4, 100 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.1% BSA). Detection of bound DDX5 proteins was achieved with a monoclonal anti-DDX5 antibody (1:400 dilution) and horseradish peroxidase (HRP)-linked anti-rabbit IgG (1:5000 dilution; 7074, Cell Signaling Technology). All unspecific bound antibodies were removed by washing plates four times (total 40 min) with washing buffer right before addition of TMB substrates (PI34028, Pierce). The signal was generated by the conversion of TMB substrates into colored products catalyzed by HRP and measured at 450 nm. For testing the effects of G4-interactive ligands on the DDX5 binding, G4-interactive ligands were added to wells before adding DDX5 and then co-incubated with 50 nM of DDX5. After 2 hr incubation, G4-interactive ligands and unbound proteins were washed away prior to addition of antibodies. For testing the effect of ATP on the DDX5 binding, different concentrations of ATP were added to wells containing the DDX5 complex with immobilized MycG4. After incubation, ATP and transiently released DDX5 were washed away prior to addition of antibodies. Data were analyzed using GraphPad Prism (GraphPad Software, Inc.). G4-forming DNA, duplex DNA, DNA hairpin, and RNA hairpin were annealed in 100 mM

K<sup>+</sup>-containing phosphate buffer with a total 10 μM of oligonucleotides by slowly cooling down from 95 °C to rt.

## 2.8 Fluorescence anisotropy experiments

Fluorescence anisotropy experiments were performed in a reaction buffer containing 3 nM annealed 3'-end 6-TAMRA labeled MycG4 oligonucleotides (Sigma-Aldrich), 50 mM Tris-acetate, 2.5 mM MgCl<sub>2</sub>, 50 mM KCl, and 0.5 mM DTT. DDX5 proteins were gradually titrated into the solution. The anisotropy signal of 6-TAMRA was measured with excitation at 555 nm and emission at 580 nm using Jasco FP-8300 Spectrofluorometer at 30 °C. Normalized anisotropy was analyzed using Graphpad Prism and fitted into the one-site specific binding equation.

## 2.9 G-quadruplex unfolding assay

Molecular-beacon MycG4 FRET probe was synthesized using β-cyanoethylphosphoramidite solid phase chemistry (Applied Biosystem Expedite 8909) as described previously with minor modifications<sup>150</sup>. 3'-(6-FAM) CPG (20-2961-xx) and 5'-BHQ-1 phosphoramidite (10-5931-xx) were obtained from Glen Research Corporation. The probe is labeled with 6-FAM (6-fluorescein) on 3'-end and BHQ-1 (Black Hole-1 quencher) on the other end. The synthesized DNA probe was eluted from the column with a 50%:50% mixture of 40% methylamine:ammonia, incubated overnight at rt and subjected to dialysis with water before lyophilization. The MycG4 FRET probe was annealed in 100 mM K<sup>+</sup>-containing Tris-acetate buffer with a total 10 μM of oligonucleotides by slowly cooling down from 95 °C to rt. Annealed G4 substrates were aliquoted into small tubes and frozen at -20 °C. Unfolding reactions were carried out in a buffer containing 20 nM oligonucleotides, 50 mM Tris-acetate, 2.5 mM MgCl<sub>2</sub>, 0.5 mM DTT with indicated concentrations of KCl and adenosine nucleotides. The time-course unfolding analysis was performed using Jasco FP-8300 Spectrofluorometer at 30 °C by addition of indicated concentrations of helicase under constant stirring (200 rpm). Endpoint unfolding analysis was carried out at 5 min after addition of helicase. Fluorescence changes were monitored with excitation wavelength at 490 nm and emission at 522 nm. Percent of unfolding was calculated as  $100 \times \Delta F / \Delta F_{\max}$ <sup>162</sup>, where ΔF is the fluorescence change after adding helicase at indicated time point, ΔF<sub>max</sub> is the difference of fluorescence between the unannealed probe in a 50 mM Tris-

acetate buffer and the annealed probe in an unfolding buffer (50 mM Tris-acetate, 2.5 mM MgCl<sub>2</sub>, 50 mM KCl). For determination of Michaelis constant of G4 unfolding, data were fit with hyperbolic curves using GraphPad Prism and equation:

$$U = \frac{U_{max} \cdot X}{K_m + X}$$

where  $U$  is the percent of G4 unfolding,  $U_{max}$  is the maximum percent of unfolding,  $X$  is the concentration of helicase, and Michaelis constant  $K_m$  is the helicase concentration at which percent of unfolding is the half of  $U_{max}$ .

## 2.10 Luciferase assay

MCF7 cells were seeded at a density of  $6 \times 10^4$  cells per well in 24-well plates. 24 hr after seeding, the medium was exchanged for fresh medium containing 1X Pen/Strep antibiotics (15070063, Gibco) and the cells were transfected with 100 nM of DDX5-siRNA or RISC-Free siRNA using Lipofectamine 3000. Next day, Fugene HD transfection mixtures that contain 500 ng of reporter constructs (pGL4.10-MYC-WT Pu43 or pGL4.10-MYC-KO Pu43<sup>163</sup>, **Figure 2.1**) were directly added to the wells. These two constructs contain a transcription template that fused a luciferase gene with either wild-type MYC promoter (MYC-WT) or MYC G4-knockout promoter sequence (MYC-KO). Luciferase activities were measured 24 hr after reporter-transfection using Dual-Luciferase Reporter Assay Systems (Promega). The results of the luciferase assays are expressed as a percent of *Firefly* luciferase activities after normalization against *Renilla* luciferase activities<sup>163</sup>.

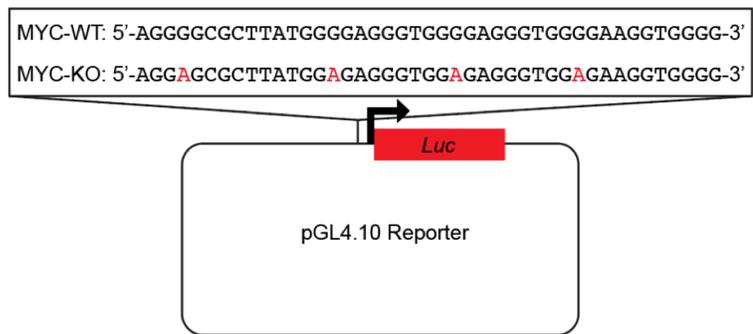


Figure 2.1. Schematic diagram of pGL4.10-MYC-WT Pu43 and pGL4.10-MYC-KO Pu43 luciferase reporter constructs

## 2.11 Quantitative reverse transcription PCR (qRT-PCR)

Total RNA was isolated using TRIzol reagent (Invitrogen). To remove phenol contamination, purified RNA was dissolved in DEPC-treated water and re-precipitated with 75% ethanol. 1 µg of RNA was subjected to cDNA synthesis using the qScript cDNA Synthesis kit (Quanta Biosciences) according to manufacturer's instructions. Real-time PCR was performed in triplicate reactions. For each reaction, a mix of the following reaction components was prepared to the indicated end-concentration: 3 µL water, 1 µL of 1:5 diluted cDNA synthesis products, 0.25 µM of each primer for MYC or GAPDH and 5 µL of SYBR Green PCR Master Mix. Cycling conditions were 95 °C for 5 min, followed by 40 cycles of 95°C for 15 s, 60°C for 15 s and 72°C for 15s. Relative gene expression was calculated by using the  $2^{-\Delta\Delta CT}$ , in which the amount of MYC mRNA was normalized to an endogenous reference (GAPDH). Melting curve analysis or agarose gel electrophoresis was carried out to confirm correct PCR products.

## 2.12 Western Blot

After collecting cells from 6-well plates, the cell pellets were re-suspended in 150 µL of 1X RIPA buffer supplemented with 1X Protease Inhibitor Cocktail (11836153001, Roche) and 1X NuPAGE LDS Sample Buffer (NP0007, Invitrogen) and then proteins were immediately denatured at 80 °C for 10 min. After sonication, 4 µL of each sample was analyzed using 4-15% Mini-PROTEAN TGX Gels (456-1086, Bio-Rad). The gels were cut into strips that contain the proteins of interest and transferred to nitrocellulose membrane (IB23002, Invitrogen) using an iBlot 2 Dry Transfer Device (Invitrogen). Immunoblotting was carried out according to standard procedures using ECL detection method. The membrane was hybridized with following antibodies: monoclonal anti-MYC (1:1000 dilution; 5605, Cell Signaling Technology), monoclonal anti-β-Actin (1: 2000 dilution; 4970, Cell Signaling Technology or A5441, Sigma), monoclonal anti-DYKDDDDK (FLAG) Tag (1:2000; 2368, Cell Signaling Technology), monoclonal anti-DDX5 (1:1000 dilution; 9877, Cell Signaling Technology or 05-850, Millipore).

## 2.13 Footnotes

Parts of this chapter have been previously published as:

Wu, G., Xing, Z., Tran, E.\*, & Yang, D.\*, DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. Proc Natl Acad Sci U S A. 116, 20453-20461 (2019).  
(Copyright © 2019 National Academy of Sciences)

## CHAPTER 3. DDX5 HELICASE RESOLVES G-QUADRUPLEX AND IS INVOLVED IN *MYC* GENE TRANSCRIPTIONAL ACTIVATION

Reprinted with permission from

Guanhui Wu, Zheng Xing, Elizabeth J. Tran\* and Danzhou Yang\*, DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. *Proc Natl Acad Sci U S A*. 116, 20453-20461 (2019). Copyright © 2019 National Academy of Sciences

### 3.1 Introduction

G-quadruplexes (G4s) are four-stranded, non-canonical secondary structures formed in guanine-rich DNA and RNA sequences<sup>3, 8</sup>. G4 structures are built upon stacked square-planar G-tetrads connected with Hoogsteen hydrogen-bonds and stabilized by monovalent cations such as K<sup>+</sup> or Na<sup>+</sup><sup>48</sup>. DNA G4s have recently been found to be involved in a number of critical cellular processes, including gene transcription, replication, and genome instability<sup>10-12, 164</sup>. In particular, *MYC*, one of the most commonly deregulated genes in human cancers, has a G4 DNA motif in its proximal G/C-rich promoter region, known as the Nuclease Hypersensitive Element III<sub>1</sub> (NHE III<sub>1</sub>; also known as the CT element, **Figure 3.1A**), that functions as a transcriptional silencer element<sup>10, 64, 80, 139, 140</sup>. Recently, DNA G-quadruplexes have been visualized in chromosomes in human cells using a G4-specific antibody<sup>158</sup>, revealing enrichment of these structures in regulatory regions of chromatin particularly the *MYC* promoter<sup>20</sup>. Transcriptional regulation of *MYC* expression is complex, with P1 and P2 being the predominant promoters<sup>132, 165</sup>. The NHE III<sub>1</sub> element is upstream of the P1 and P2 promoters and is an important cis-element for transcriptional regulation of *MYC*<sup>139, 166</sup> (**Figure 3.1A**). When the *MYC* gene is not actively transcribed, inherent supercoiling generally does not make MycG4 to form in the double-stranded promoter region<sup>77</sup>. However, in highly transcribed cells, the transcription machinery generates dynamic, negative supercoiling behind the moving machinery<sup>71, 73, 167</sup>. This dynamic, transcription-coupled negative supercoiling can be transmitted to the NHE III<sub>1</sub> region where it promotes melting of genomic DNA to the intermediate single-stranded forms that can spontaneously form the metastable stable G-quadruplex structures<sup>59, 78, 79, 168</sup>. The formation of G4s in the *MYC* promoter inhibits *MYC* transcription by preventing the binding of double-stranded (Sp1) or single-stranded (CNBP and

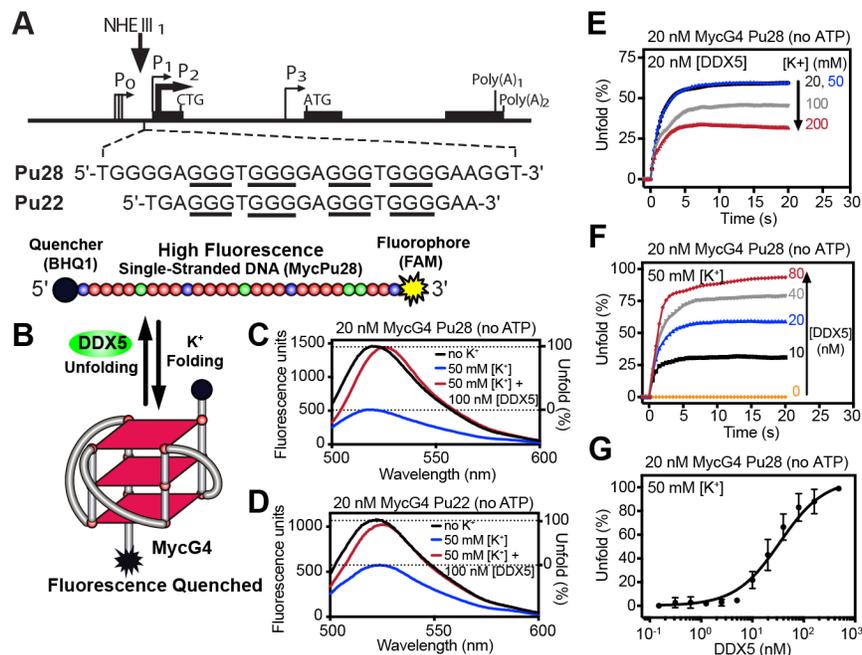


Figure 3.1. DDX5 actively unfolds MycG4 shown by a FRET-based helicase unfolding assay

(A) The *MYC* promoter structure with the G4-forming region NHE III<sub>1</sub> sequence Pu28 is shown. The truncated Pu22 sequence that forms the major MycG4 is also shown. (B) A molecular beacon MycG4 FRET probe was designed for helicase unfolding assays using the MycG4 Pu28 and MycG4 Pu22 sequences. Each oligo was labeled with the fluorophore 6-FAM (6-fluorescein) on 3' end and quencher BHQ-1 (Black Hole-1 quencher) on the 5' end. (C and D) MycG4 unfolding by DDX5 in MycG4 Pu28 (C) and MycG4 Pu22 (D). Percent of MycG4 unfolding was calculated as  $100 \times \Delta F / \Delta F_{\max}$ , where  $\Delta F_{\max} = (\text{black line} - \text{blue line})$ , the difference of fluorescence between the unfolded probe (black line) in 0 mM K<sup>+</sup> and the folded probe (blue line) in 50 mM K<sup>+</sup> solution (50 mM pH 7.4 Tris-acetate, 20 nM MycG4 Pu28 or MycG4 Pu22, 2.5 mM MgCl<sub>2</sub>, no ATP), and  $\Delta F = (\text{red line} - \text{blue line})$ , the fluorescence change after adding DDX5 (red line) from the folded probe (blue line). (E) Time course analysis of 20 nM MycG4 unfolding by 20 nM DDX5 at various concentrations of K<sup>+</sup> (20, 50, 100, and 200 mM) (no ATP). For relative binding of DDX5 to MycG4 at different concentrations of K<sup>+</sup>, see **Figure 3.4**. (F) Time course analysis of 20 nM MycG4 unfolding at various concentrations of DDX5 (0, 10, 20, 40, and 80 nM) in 50 mM K<sup>+</sup> (no ATP) showing DDX5 unfolds MycG4 in a dose-dependent manner. MycG4 Pu28 FRET probe was used as the substrate. (G) The observed Michaelis constant ( $K_m$ ) for DDX5-mediated G4 unfolding was  $32 \pm 8$  nM of DDX5 (50 mM K<sup>+</sup>, no ATP). Unfolding reactions were incubated for 5 min before taking fluorescence measurements.  $n = 3$  independent measurements. Error bars represent mean  $\pm$  s.d..

hnRNP K) transcriptional factors<sup>80, 169</sup>. Compounds that bind and stabilize this G4 have been shown to reduce *MYC* expression and are antitumorigenic<sup>6, 10, 80</sup>. We previously determined the molecular structure of the major G4 formed in the *MYC* promoter NHE III<sub>1</sub> G-rich strand (Pu28, **Figure 3.1A**) in physiologically relevant K<sup>+</sup> solution, revealing a parallel-stranded structure<sup>87</sup> (MycG4, **Figure 3.1B**). The MycG4 structure is very stable under physiologically relevant salt conditions with a melting temperature over 85 °C in 100 mM K<sup>+</sup> solution<sup>88, 143, 144</sup>. Therefore, regulation of *MYC* expression would require active unfolding of the MycG4 G-quadruplex structure. Such a resolvase, however, has yet to be discovered.

DEAD-box (DDX) proteins define the largest family of ds RNA helicases, which contain 12 conserved sequence motifs including the eponymous DEAD motif (Asp-Glu-Ala-Asp, or D-E-A-D)<sup>170</sup>. The DEAD-box helicases are unique among helicase families, in that they are non-directional and non-processive. The human DEAD-box protein 5 (DDX5), also named p68, is one of the founding members of the DEAD-box RNA helicase family and is implicated in a number of critical cellular processes including cell proliferation and organ development<sup>43, 171-173</sup>. While its level is low in normal cells, DDX5 is overexpressed in a number of major human cancers, including colon, lung, breast, and prostate, and is shown to promote tumorigenesis, tumor progression, and cellular transformation<sup>174-179</sup>. DDX5 has been suggested to be a transcriptional regulator<sup>180-182</sup>. The DDX5 yeast ortholog Dbp2 has been shown to directly associate with transcriptionally active chromatin<sup>183, 184</sup>. Intriguingly, DDX5 has been shown to activate *MYC* expression in multiple solid tumors by an unknown mechanism<sup>176, 177, 185</sup>. However, the mechanism of transcriptional regulation by DDX5 is poorly understood, as the RNA helicase activity of DDX5 does not seem to be required for transcriptional regulation in many cases<sup>174, 182</sup>.

Here we report that DDX5 is a highly active new DNA and RNA G4 resolvase that does not require a single-stranded overhang. ATP hydrolysis is shown not to be directly coupled to the DDX5 G4-unfolding. DDX5 proficiently unfolds the *MYC* promoter DNA G4 and activates *MYC* expression in human cancer cells in a G4-dependent manner. Further, the DDX5 unfolding of MycG4 and DDX5-induced *MYC* activation are inhibited by G4-stabilizing small molecules. Thus, the DDX5-MycG4 interaction represents a potential future target for *MYC* downregulation for cancer intervention.

## 3.2 Results

### 3.2.1 Characterization of DDX5 as a proficient G-quadruplex resolvase

We first investigated the G4-unfolding activities of DDX5. We designed a fluorescence resonance energy transfer (FRET) probe using the full-length *MYC* promoter G4 DNA (MycG4-Pu28, **Figure 3.1A**), which is labeled with 6-fluorescein (6-FAM) on the 3'-end and Black Hole-1 quencher (BHQ-1) on the 5'-end (**Figure 3.1B**). The stable formation of G-quadruplexes requires the presence of  $K^+$  or  $Na^+$  cations, with a preference for  $K^+$  (**Figure 3.1B**). In the absence of  $K^+$ , the MycG4-Pu28 existed in the single-stranded form with the two ends far apart, as evidenced by high FAM-fluorescence (**Figure 3.1B top, 3.1C black line**). In the presence of 50 mM  $K^+$ , DNA G4 is formed and FAM-fluorescence was quenched (**Figure 3.1C blue line**). Upon addition of DDX5 to MycG4-Pu28 DNA in the  $K^+$  solution, the FAM-fluorescence was dramatically increased (**Figure 3.1C red line**) to the same level of unfolded Pu28 in the absence of  $K^+$  (**Figure 3.1C black line**), indicating that MycG4 is unfolded to the single-stranded form by DDX5.

A unique feature of DEAD-box helicases is that they separate strands locally rather than in a translocation-based manner, therefore their dsRNA unwinding activity does not require an extended flanking tail<sup>170, 186</sup>. All active G4 helicases/resolvases reported to date require a long single-stranded tail at either the 3'- or 5'- end for loading and subsequent G4-unfolding<sup>43</sup>. To test whether the extended tail is needed for G4-unfolding by DDX5, we first prepared a truncated MycG4 FRET probe, MycG4-Pu22<sup>87</sup>, which contains the four guanine-runs with minimal flanking capping segments that are required to form the well-defined major G4-conformation in the full-length MycG4-Pu28. Upon addition of DDX5, the FAM-fluorescence was dramatically increased (**Figure 3.1D**), indicative of MycG4-Pu22 unfolding by DDX5, as observed for Pu28 (**Figure 3.1C**). We then prepared MycG4-Pu16 that contains only the core G4-forming sequence without any flanking segments and found the same G4-unfolding activity of DDX5 (**Figure 3.2**). These results demonstrate that DDX5 can load on and unfold the MycG4 without extended flanking tails.

We next examined the effect of  $K^+$  because it can stabilize G4 structures<sup>143</sup>. Using equimolar concentrations of DDX5 and DNA (20 nM), the MycG4 unfolding activity of DDX5 was similar in 20 mM and 50 mM  $K^+$  concentrations (50%). Increasing the  $K^+$  concentration

MycG4 Pu16: 5'-GGGTGGGGAGGGTGGG-3'

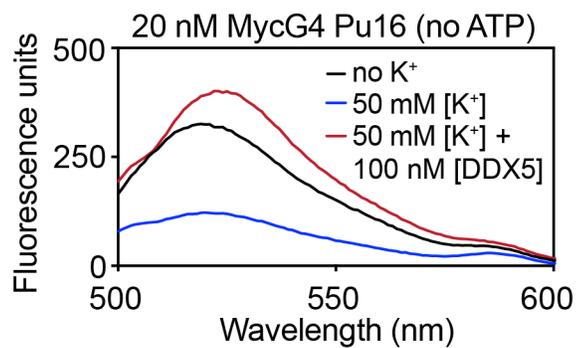


Figure 3.2. DDX5 unfolds MycG4 without flanking tails

The sequence of MycG4 Pu16 (top) is the core MycG4-forming sequence without any flanking tails. MycG4 Pu16 is labeled with the fluorophore 6-FAM (6-fluorescein) on 3' end and quencher BHQ-1 on the 5' end. MycG4 Pu16 is folded in 50 mM K<sup>+</sup> solution (blue line). Upon addition of DDX5, the FAM-fluorescence is dramatically increased (red line) indicating the unfolding of MycG4.

above 50 mM reduced the unfolding activity of DDX5 (**Figure 3.1E**). Because DDX5 showed high unfolding amplitudes up to 50 mM K<sup>+</sup>, this potassium concentration was used in most FRET assays.

To investigate the unfolding activity of DDX5, we carried out time-course analyses of G4-unfolding by DDX5 using the FRET assays. The results revealed that the MycG4 structure was rapidly unfolded by DDX5, with a half-time of a few seconds (**Figures 3.1E and 3.1F**). In addition, DDX5 appeared to unfold the MycG4 Pu28 in a dose-dependent manner (**Figure 3.1F**); DDX5 unfolds about 50% of MycG4 DNA at equimolar concentration of DNA (20 nM). The observed Michaelis constant ( $K_m$ ) of unfolding of MycG4 Pu28 DNA was  $32 \pm 8$  nM of DDX5 (**Figure 3.1G**).

Taken together, our results demonstrate that DDX5 proficiently unfolds G4 DNA without a requirement for ssDNA tail, representing a distinct mechanism among known G4 resolvases to date.

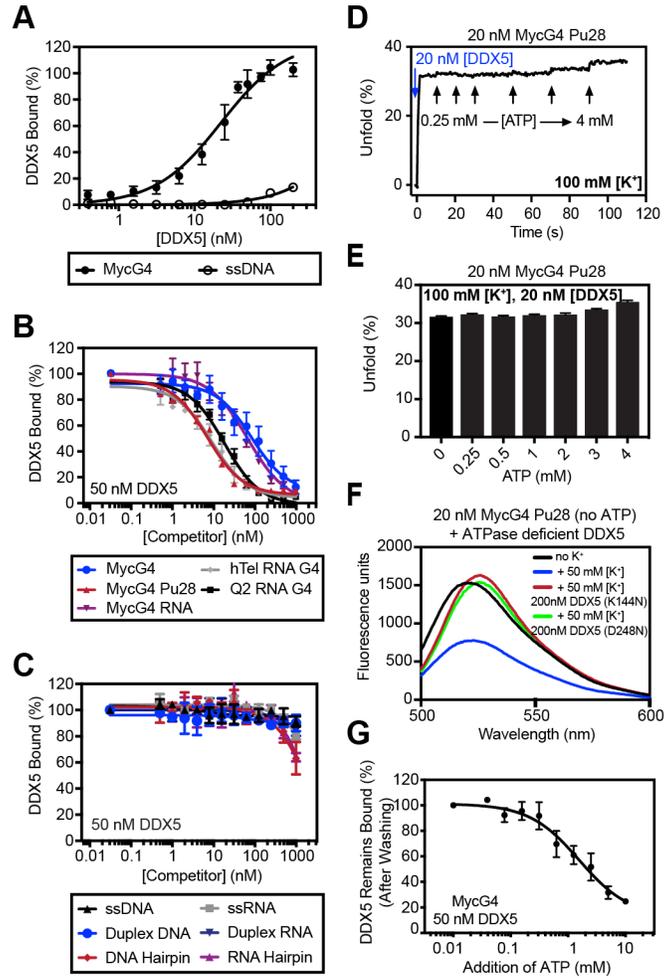
### 3.2.2 G-quadruplexes are preferred substrates of DDX5 *in vitro* and *in vivo*

We next measured the G4 binding activity of DDX5. Utilizing biotinylated DNA immobilized to streptavidin-coated plates in an Enzyme-Linked ImmunoSorbent Assay (ELISA), we measured the direct binding of DDX5 to various DNA substrates in 100 mM K<sup>+</sup> solution (**Figure 3.3A**). DDX5 displayed a high binding affinity to MycG4 DNA, markedly greater than to single-stranded (ss) DNA (**Figure 3.3A**). As K<sup>+</sup> induces<sup>187</sup> and stabilizes<sup>143</sup> the G4-formation, we tested the relative binding of DDX5 to immobilized unannealed MycG4 DNA in ELISA in the presence of increasing concentrations of K<sup>+</sup>. These results showed that K<sup>+</sup> enhances the DDX5 binding of MycG4 DNA, with maximal binding at 50 mM K<sup>+</sup> (**Figure 3.4A**). This observation may explain the similar G4-unfolding activity of DDX5 in 20 mM and 50 mM K<sup>+</sup> concentrations (**Figure 3.1E**). The dissociation constant ( $K_d$ ) of DDX5 binding to MycG4 DNA was determined to be  $13 \pm 1$  nM using fluorescence anisotropy (**Figure 3.4B**). Additionally, DDX5 showed a preference for the G-rich sequence over other non-G-rich sequences in the presence of K<sup>+</sup> (**Figure 3.4C**).

To test whether DDX5 binds to RNA G4 structures, or other ds or ss DNA and RNA conformations, we performed ELISA competition experiments using immobilized MycG4 DNA.

Figure 3.3. DDX5 preferentially recognizes and unfolds G-quadruplex structures in an ATP-independent manner

(A) Binding determined by ELISA in 100 mM K<sup>+</sup> showing that DDX5 has a high affinity for MycG4 structure ( $K_d = 22 \pm 2$  nM).  $K_d$  was calculated using the one-site specific binding equation.  $n = 3-6$  independent measurements. Error bars represent mean  $\pm$  s.d. (B and C) Competition binding ELISA experiments with the immobilized MycG4 DNA in showing DDX5 preferentially recognizes G4 structures. DDX5 was co-incubated with increasing concentrations of either G4-forming oligonucleotides (MycG4 Pu28, MycG4, MycG4 RNA, TERRA RNA G4, or Q2 RNA G4) (B) or non-G4-forming oligonucleotides (ssDNA, ssRNA, DNA hairpin, RNA hairpin, double-stranded (ds) DNA, or dsRNA) (C). Data were fitted with a one-site competition binding equation.  $n = 4-6$  independent measurements. Error bars represent mean  $\pm$  s.d. (D-E) Time-course analysis of 20 nM MycG4 FRET unfolding in 100 mM K<sup>+</sup> showing that the unfolding of MycG4 by DDX5 is slightly enhanced by the addition of ATP (D). Black arrows from left to right correspond to total concentrations of ATP in the solution (0.25, 0.5, 1, 2, 3, and 4 mM). The blue arrows correspond to the concentration of DDX5 in the solution (20 nM). The percentages of 20 nM MycG4 unfolding by 20 nM DDX5 in 100 mM K<sup>+</sup> at different ATP concentrations are as follows: 32%, 32%, 32%, 32%, 32%, 34%, and 35%, from left to right. (E). (F) G4 unfolding FRET assay showing the ATPase-deficient DDX5 mutant proteins K144N and D248N are able to unfold MycG4 structures. The mutant DDX5 protein K144N cannot bind ATP, while D248N can bind but not hydrolyze ATP. (G) ELISA experiments showing the addition of ATP into DDX5/MycG4 complex decreases the population of DDX5 bound to MycG4 after washing steps. Different concentrations of ATP were added to wells containing the DDX5 complex with immobilized MycG4. After incubation, ATP and transiently released DDX5 were washed away prior to the addition of antibodies.  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.e.m..



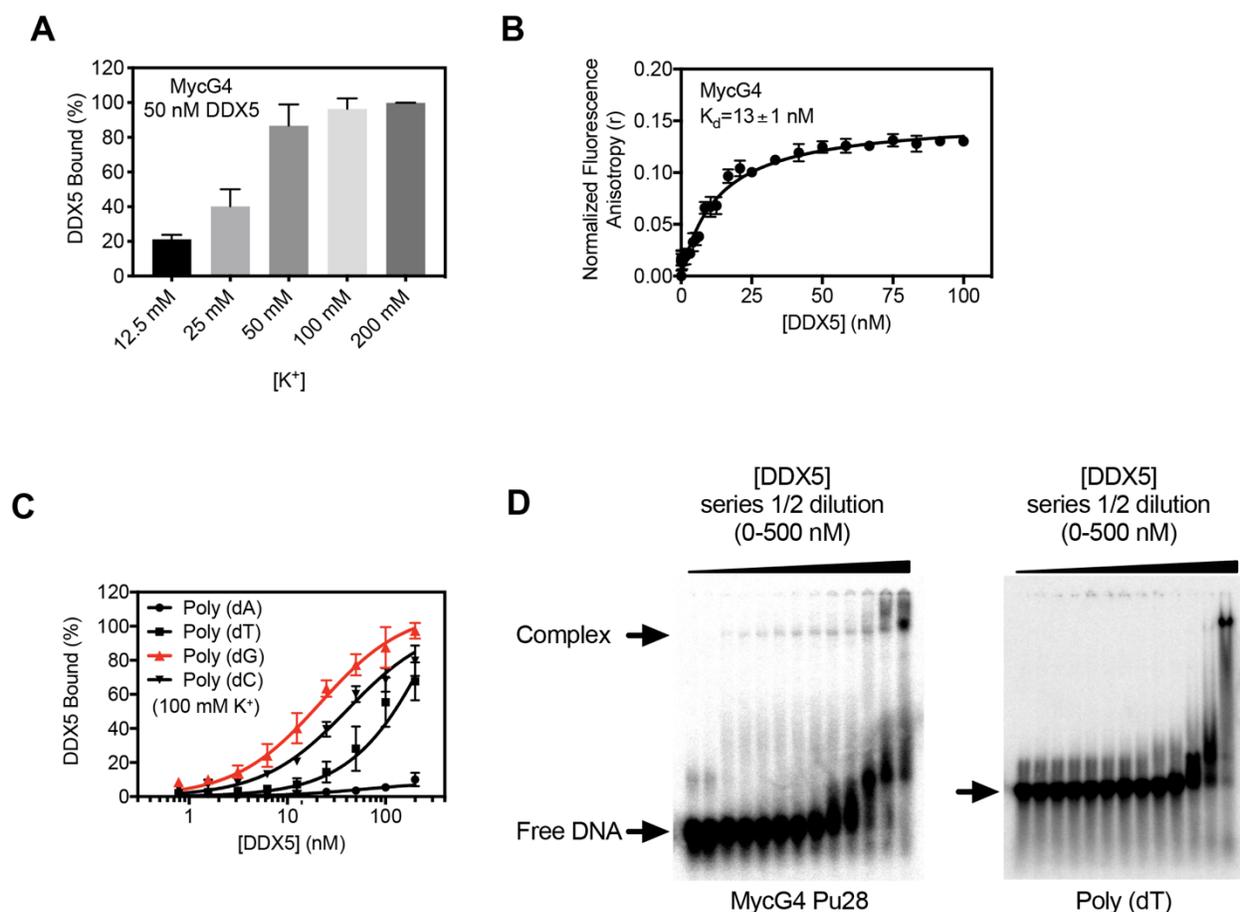


Figure 3.4. Further characterization of DDX5 as a G-quadruplex binder

(A) Relative binding as determined by ELISA showing that increasing concentration of  $K^+$  initially promotes DDX5/G4 interaction, which is saturated at 50 mM of  $K^+$ .  $n = 3$  independent measurements. Error bars represent mean  $\pm$  s.d. (B) The binding curve as determined by fluorescence anisotropy experiments showing that DDX5 has a high affinity for the MycG4. 3'-end 6-TAMRA labeled MycG4 was used as the substrate. The dissociation constant ( $K_d$ ) was calculated using one-site specific binding equation. (C) Relative binding curves as determined by ELISA showing that DDX5 has a preference for G-rich oligonucleotides.  $n = 3$  independent measurements. Error bars represent mean  $\pm$  s.d. (D) DDX5 has a preference for MycG4 Pu28 compared to poly (dT) as demonstrated by EMSA. The concentration of labeled oligonucleotides (MycG4 Pu28 or poly (dT)) is 1 nM. DDX5-DNA complex can be formed in the presence of trace amounts of DDX5 ( $\sim$ 500 pM, the third lane from left) with MycG4 DNA (left), but not with poly-(dT) DNA (right). As previously observed<sup>184</sup>, the DDX5 protein with a calculated pI of 9.06. DDX5 is positively charged at pH 7.4, which makes it difficult to get into the gel during electrophoresis. Addition of DDX5 to the MycG4 DNA caused significant changes on the migration of the unbound probe but could not induce the complete shifting of MycG4. We attribute the shift of the unbound probe to the binding and unfolding by DDX5, as well as gradual dissociation of the unfolded DNA from the DDX5-complex during the gel running process, which has been observed in other protein-DNA systems<sup>188</sup>.

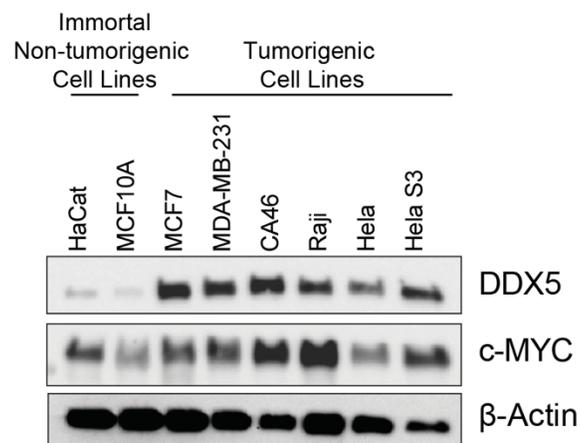


Figure 3.5. The positive correlation between the expressions of *MYC* and *DDX5* in multiple immortal cell lines

The western blot analysis indicates that the positive correlation between the expressions of *MYC* and *DDX5* presents in multiple immortal cell lines and contributes to tumorigenesis.



The binding of DDX5 to immobilized MycG4 DNA was significantly inhibited by both DNA and RNA G4 structures (**Figure 3.3B**), but not by other DNA or RNA conformations, such as dsRNA, RNA hairpin, ssRNA, and dsDNA, DNA hairpin, and ssDNA (**Figure 3.3C**). Importantly, the MycG4 DNA and MycG4 RNA showed the same inhibitory activity of DDX5 binding to immobilized MycG4 DNA (**Figure 3.3B**), suggesting DDX5 recognizes the G4-conformation through a structure-based mechanism. The binding of DDX5 to MycG4 was confirmed by electromobility shift assay (EMSA) (**Figure 3.4D**). Addition of DDX5 to the MycG4 DNA caused significant changes to the migration of the unbound probe. We attribute the shift of the unbound probe to the binding and unwinding of DDX5, as well as gradual dissociation of the unfolded DNA from the DDX5-complex during the gel running process, which has been observed in other protein-DNA systems <sup>188</sup>.

We then investigated the DDX5's preferred chromatin-binding sites reported *in vivo*. We analyzed publicly available DDX5 ChIP-seq data <sup>189</sup>, which was obtained from HeLa S3 cells that have high expression levels of both DDX5 and Myc (**Figure 3.5**). We found that most of the top 15 DDX5-binding motifs identified by MEME (motif-based sequence analysis tools) <sup>190</sup> are G-rich sequences (**Figure 3.6A**), suggesting G4 structures are one of the preferred substrates for DDX5 in human chromatin.

### 3.2.3 ATP hydrolysis is not directly coupled to G-quadruplex-unfolding by DDX5

Unexpectedly, our results showed that the unfolding activities of DDX5 on MycG4 structure occurred in the absence of ATP (**Figure 3.1**). DDX5's binding and unwinding activity of dsRNA is ATP-dependent and saturated at 2 mM ATP <sup>161</sup>. To confirm that our observations were not caused by ATP contamination of purified DDX5, hexokinase and glucose were added to purified protein prior to the assessment of its G4 unfolding activity. Hexokinase phosphorylates glucose using ATP and has been used previously to deplete copurifying ATP from helicase protein preparations <sup>191</sup>. No significant difference in MycG4 unwinding activity was observed with or without pretreatment of hexokinase (**Figure 3.7**), indicating our observations were not because of ATP contamination. To further investigate the effect of ATP, we titrated ATP into the unfolding reaction up to 4 mM. Our results showed that the addition of ATP does not appreciably alter the MycG4 unfolding activity of DDX5 (**Figures 3.3D and 3.3E**).

To confirm the ATP-independence of DDX5 G4-unfolding, we examined MycG4 unfolding using two ATPase-deficient mutant DDX5 helicases, mutant K144N and D248N (**Figure 3.3F**). The mutant DDX5 protein K144N cannot bind ATP, while D248N can bind but cannot hydrolyze ATP <sup>160</sup>. In contrast to dsRNA unwinding, which is ATP dependent, our results showed that both of the DDX5 mutant proteins were able to unfold MycG4. This suggests that DDX5 utilizes a mechanism for unwinding G4 that is distinct from dsRNA unwinding.

We further examined the effects of ATP on the binding of DDX5 to MycG4-DNA. Interestingly, the addition of ATP to the pre-formed complex of DDX5 and immobilized MycG4-DNA in ELISA experiments followed by washing steps significantly reduced the amount of DDX5 protein bound to immobilized MycG4-DNA (**Figure 3.3G**). Since the addition of ATP did not change DDX5 unfolding activity in the FRET assays (**Figures 3.3D and 3.3E**), which didn't contain washing steps, this result suggests that ATP is likely required for release of unfolded G4 from DDX5. Collectively, these data suggest that the ATP hydrolysis is not directly coupled to G4-unfolding by DDX5.

### 3.2.4 Conformational change of MycG4 DNA induced by DDX5

To investigate the conformational change of the MycG4 DNA induced by DDX5, we carried out DMS footprinting experiments in which exposed N7s of guanine nucleotides are methylated by DMS, cleaved by subsequent piperidine treatment, and then visualized on a sequencing gel <sup>192</sup>. G-tetrad (**Figure 3.8A**) formation <sup>10</sup> and protein/nucleic acid interactions <sup>192</sup> have been shown to protect the N7s of guanines against DMS methylation, while protein-induced DNA/RNA conformational changes can alter the DMS methylation-induced cleavage pattern <sup>192-194</sup>. In the presence of K<sup>+</sup>, four consecutive G-runs (R2-R5) of Pu28 displayed a clear protection pattern against DMS methylation-induced cleavage (**Figure 3.8B, lane 2**), indicating the formation of MycG4 in the Pu28 sequence (**Figure 3.1B**) <sup>10</sup>. Addition of DDX5 to the Pu28 DNA enhanced DMS methylation-induced cleavage (**Figure 3.8B, lane 3**) as compared to free Pu28 (**Figure 3.8B, lane 2, and autoradiogram densitometric scans**). In the presence of DDX5, guanines in the G-run R5 showed significantly greater cleavage, while both R3 and R4 G-runs showed higher cleavage, indicating N7s of guanines in G-runs R3, R4, and R5 were more exposed to DMS methylation. This result demonstrates that the intramolecular G4 structure was disrupted by DDX5. Interestingly, R2, R3, and R4 were still partially protected from DMS-methylation-induced

cleavage (compare lane 1 and lane 3), suggesting association of DDX5 with the unfolded MycG4 DNA.

We also performed CD spectroscopy measurements to analyze the conformational change of MycG4 induced by DDX5. The MycG4-Pu28 DNA forms a parallel-type G4 structure in  $K^+$  solution, as shown by a characteristic CD maximum at 260 nm <sup>143</sup> (**Figure 3.8C**). DDX5 protein does not have any signal at 260 nm (**Figure 3.8D**). Upon addition of DDX5 to MycG4 Pu28 DNA, the CD intensity at 260 nm decreased as the concentration of DDX5 increased (**Figure 3.8C**), indicating the disruption of G4 structure by DDX5.

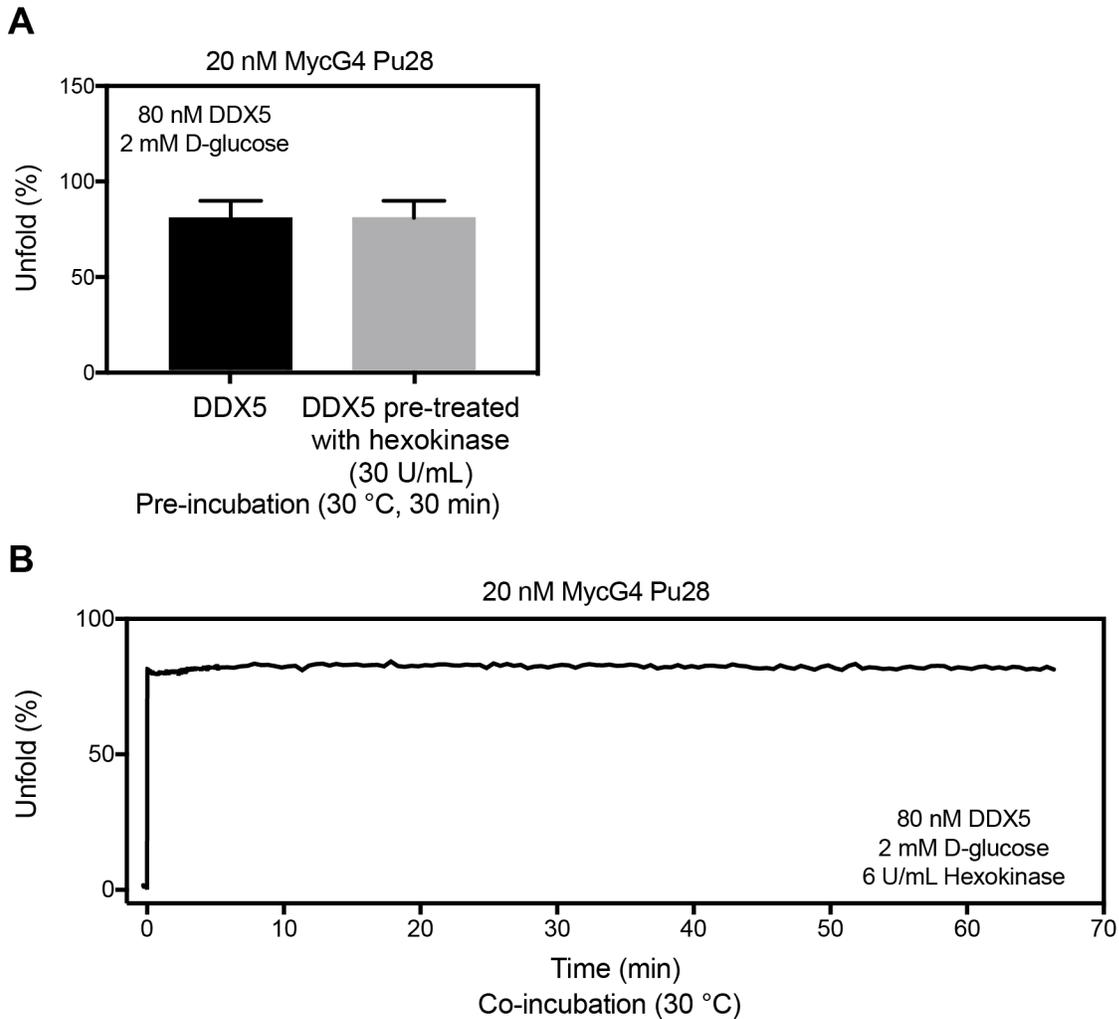


Figure 3.7. MycG4 unfolding by DDX5 occurs in the absence of ATP

(A) Hexokinase is able to catalyze the transfer of a phosphoryl group from ATP to a D-glucose molecule, therefore including hexokinase plus D-glucose in the reaction solution can exclude the effect of ATP contamination. DDX5 was incubated with 30 U/mL of hexokinase and 2 mM D-glucose for 30 min at 30 °C prior to assessment of its G4-unfolding activity. We found pre-incubation with hexokinase plus D-glucose did not compromise G4 unfolding activity of recombinant DDX5. (B) Time course analysis showing the presence of hexokinase plus D-glucose cannot inhibit G4 unfolding activity of DDX5. The unfolding was occurred in a buffer containing 50 mM Tris-acetate (pH 7.4), 2.5 mM MgCl<sub>2</sub>, 50 mM KCl, 0.5 mM of DTT, 6 U/mL of hexokinase and 2 mM D-glucose. The signal was recorded using Jasco FP-8300 Spectrofluorometer at 30 °C under constant stirring (200 rpm). Percent of unfolding was calculated as  $100 \times \Delta F / \Delta F_{\max}$ .

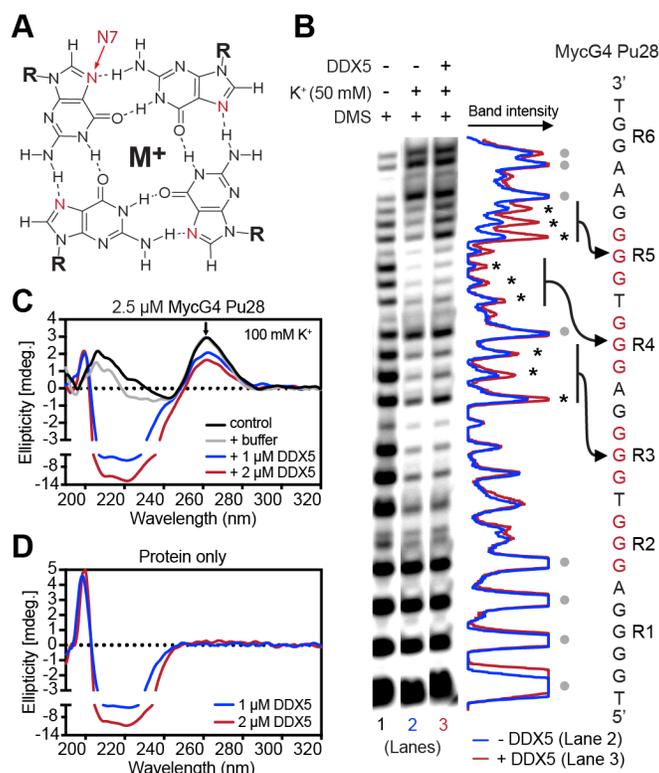


Figure 3.8. DMS footprinting and CD spectroscopy showing the conformation change of MycG4 induced by DDX5

(A) Schematic diagram showing the structure of G4. N7s of guanine nucleotides are labeled in red.  $M^+$  represents monovalent metal cation. (B) DMS methylation assay of the MycG4 Pu28 showing the conformational change of Pu28 induced by DDX5. Pu28 oligonucleotides were incubated with DMS for 21 min at appropriate conditions. Autoradiogram densitometric scans are shown to the right of the gel comparing Pu28 cleavage patterns in the absence (blue; lane 2) and presence (red; lane 3) of DDX5 (500 nM). Guanines with enhanced cleavage in the presence of DDX5 are labeled with black asterisks. Guanines that are not involved in G4 formation (and therefore have very similar cleavage patterns in the absence and presence of DDX5) are labeled with grey dots. (C) CD titration spectra of MycG4 Pu28 in 100 mM  $K^+$  solution with increasing concentrations of DDX5. MycG4 Pu28 forms a parallel G-quadruplex structure in the presence of 100 mM  $K^+$ , as indicated by characteristic maxima at near 260 nm (indicated by the arrow).

Upon addition of DDX5, the CD intensity at 260 nm was significantly reduced. All measurements were using the same MycG4 Pu28 stock solution, with Pu28 oligonucleotides pre-annealed in the presence of 100 mM  $K^+$ . (D) CD spectra of DDX5 protein at different concentrations.

### 3.2.5 DDX5 directly interacts with the *MYC* promoter G-quadruplex-forming region and activates *MYC* expression in tumorigenic cells

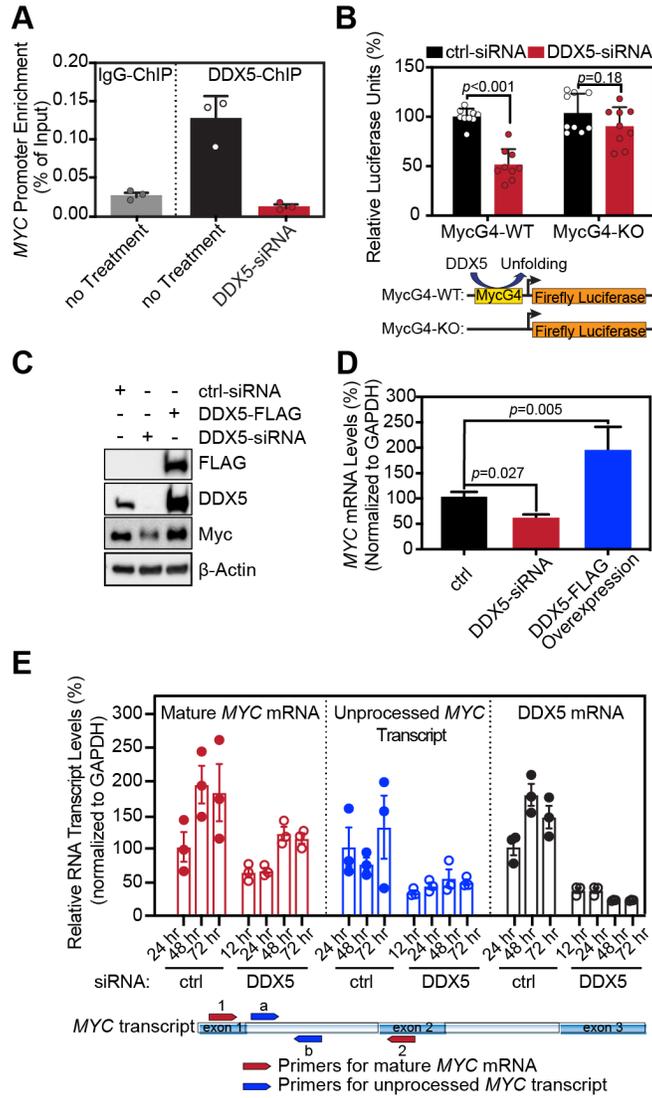
DDX5 level is low in normal cells but is overexpressed in multiple human tumors<sup>174, 175, 177-179, 195</sup>. In addition, DDX5 has been shown to activate *MYC* in tumors<sup>176, 177, 185</sup> and a positive feedback loop of DDX5 and Myc was suggested to contribute to tumorigenesis<sup>185</sup>. As DNA G4s have been shown to form in the *MYC* promoter and negatively regulate *MYC* transcription<sup>10, 20</sup>, we rationalized that DDX5 regulates *MYC* expression in a G4-dependent mechanism. To test this hypothesis, we first determined whether DDX5 directly interacts with the *MYC* G4-forming promoter region *in vivo* by chromatin immunoprecipitation (ChIP) experiments in MCF7 cells (**Figure 3.9A**). The human breast cancer MCF7 cell line has been shown to have a high-level of DDX5 expression<sup>185</sup>. The ChIP results showed that DDX5 directly interacts with the *MYC* promoter, as demonstrated by the enrichment of the *MYC* G4-forming promoter region in co-immunoprecipitated DNA (**Figure 3.9A**). Note that this enrichment was lost after knocking down DDX5 by DDX5-siRNA, indicating that the signal is not due to background. We further analyzed the published DDX5 ChIP-seq data from HeLa S3 cells<sup>189</sup>, which showed high expression levels of both Myc and DDX5 (**Figure 3.5**), and found a specific enrichment of DDX5 close to the TSS of the *MYC* gene, corresponding to the G4-forming region (**Figure 3.6B**).

We next employed luciferase assays using a MycG4-WT construct with the wild-type *MYC* promoter sequence or a MycG4-KO construct with a MycG4-knockout promoter sequence in front of the luciferase gene<sup>163</sup> in MCF7 cells. For the MycG4-WT construct, the promoter activities were markedly repressed by DDX5-siRNA. In contrast, a smaller effect was observed for the MycG4-KO construct (**Figure 3.9B**). These results indicate that DDX5-depletion inhibits the *MYC* promoter activity when the MycG4 is present.

We then examined the Myc expression levels in MCF7 cells in response to DDX5 depletion or DDX5 overexpression. The treatments of DDX5-siRNA lowered both *MYC* protein and mRNA levels in MCF7 cells (**Figures 3.9C and 3.9D**). To directly assess the effect of DDX5 on *MYC* transcription, we designed two sets of primers (**Figure 3.9E, bottom**) to amplify either the unprocessed, nascent or mature *MYC* RNA transcripts by qRT-PCR at various time points following DDX5 depletion (**Figure 3.9E and Figure 3.10**). Our results showed that both unprocessed and mature *MYC* transcripts were rapidly decreased within 12 hr after DDX5 depletion and this effect could last up to 72 hr, suggesting the direct involvement of DDX5 in the

Figure 3.9. DDX5 regulates *MYC* transcriptional activity in a G-quadruplex-dependent manner

**(A)** The interaction of DDX5 with the *MYC* promoter was probed by ChIP-qPCR with or without 24 hr treatment of DDX5-siRNA in MCF7 cells. DDX5 directly interacts with the *MYC* promoter at the MycG4 forming region (left), whereas the MycG4 forming region cannot be detected from the immunoprecipitated DNA/protein complex after DDX5 silencing (middle).  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.d.. **(B)** Luciferase reporter assays showing that DDX5 silencing by siRNA inhibits wild-type, but not G4-knockout, *MYC* promoter activity in MCF7 cells. Relative firefly luciferase activities produced by two reporters (pGL4.10-c-MYC-WT Pu43 or pGL4.10-c-MYC-KO Pu43<sup>163</sup>) at different conditions were normalized to the MYC-WT reporter in the presence of non-targeting control siRNA (ctrl-siRNA). Renilla luciferase activity produced by pRL-TK reporter was served as an internal control.  $n = 9$  biologically independent samples. Error bars represent mean  $\pm$  s.d.  $p$ -value was calculated by Sidak's multiple comparisons test after two-way ANOVA. **(C)** Representative western blot analysis showing positively correlated expressions of Myc and DDX5 in MCF7 cells. DDX5 was either knocked down using siRNA or overexpressed using DDX5-FLAG encoding vectors in MCF7 cells. The cells were transfected with respective plasmids for 48 hr before harvesting. Note: FLAG/DDX5, and Myc/ $\beta$ -actin were analyzed as two blots. The positive correlation between Myc and DDX5 expressions was further confirmed using 22Rv1 prostate cancer cell line, see **Figure 3.11**. **(D)** *MYC* transcription is directly regulated by DDX5. DDX5 was either knocked down using siRNA or overexpressed using DDX5-FLAG encoding vectors in MCF7 cells, *MYC* mRNA levels were measured using qRT-PCR at 48 hr after transfection. Error bars represent mean  $\pm$  s.d.  $p$ -value was calculated by Dunnett's multiple comparisons test after one-way ANOVA. **(E)** qRT-PCR results showing knocking-down DDX5 by siRNA decreases the levels of both unprocessed and mature *MYC* RNA transcripts at 12, 24, 48, and 72 hr time points. Unprocessed *MYC* RNA transcript was measured using the intron specific primers (primers a and b). Mature *MYC* mRNA levels were measured using a pair of primer flanking intron 1 (primers 1 and 2). All RNA transcript levels were measured at desired time points after the DDX5-siRNA treatments.  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.e.m.. The same data with normalization is shown in **Figure 3.10**.



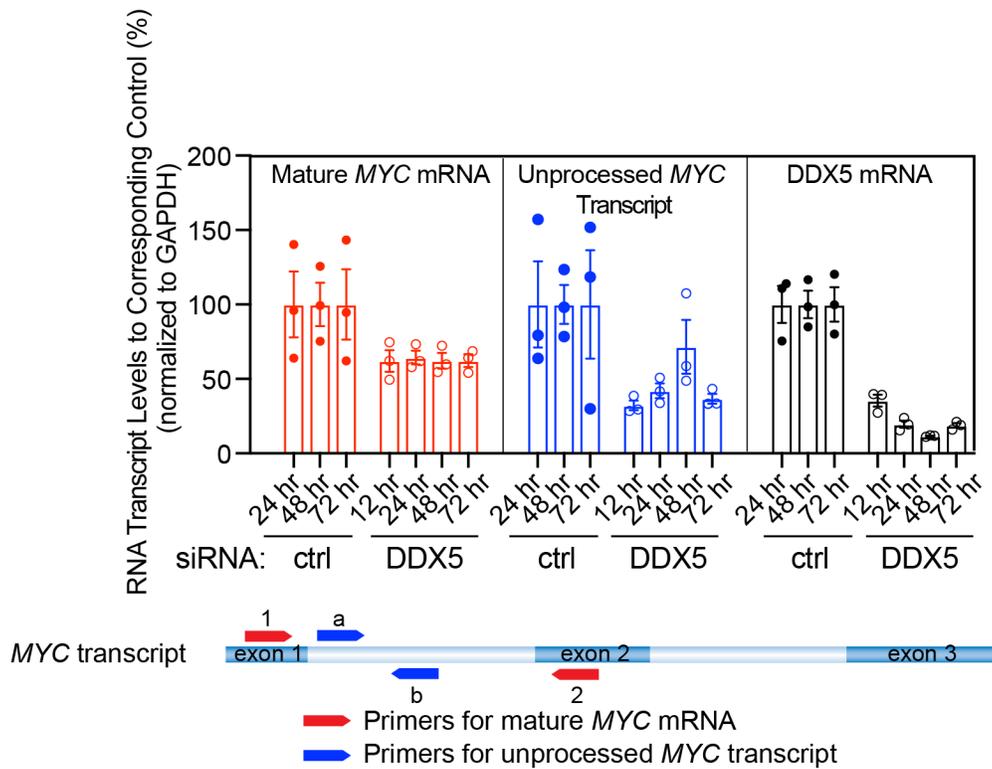


Figure 3.10. Normalized data from Figure 3.9E

Different RNA transcript levels that are normalized to the corresponding controls at different time points. Note: The RNA transcript levels from the 12 hr siRNA-DDX5 samples were normalized to the 24 hr control samples.

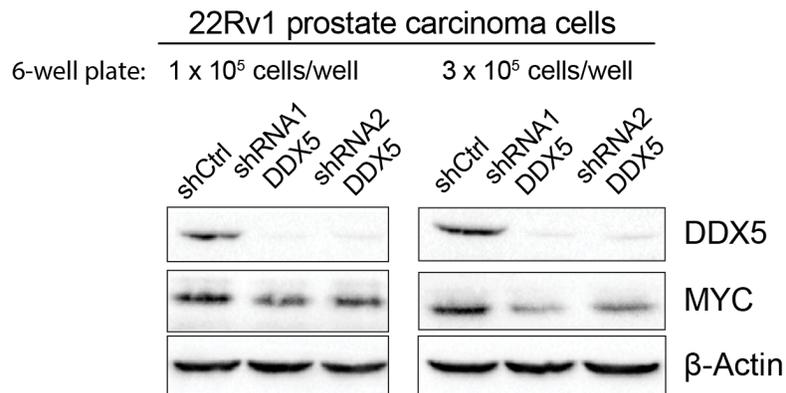


Figure 3.11. The effects of DDX5 depletion using two different shRNAs in 22Rv1 prostate carcinoma cells

To knockdown DDX5 (Gene ID: 1655) expression, two different SMARTvector Lentiviral shRNAs were transfected into 22rv1 cells using the Lipofectamine 2000 according to the manufacturer's instructions. Stable 22Rv1 cell lines carrying shDDX5 or shCtrl expression vectors were selected and then maintained with 4 µg/ml puromycin. Western blotting analysis were conducted to examine the effects of DDX5 depletion on *MYC* expression. All analyses were performed at 48 hr post shRNA induction.

transcriptional activation of the *MYC* gene. The suppression of *MYC* expression by DDX5 depletion was also shown in a prostate carcinoma cell line 22Rv1 using two different DDX5-silencing shRNAs (**Figure 3.11**). In contrast, the overexpression of DDX5 by a DDX5-overexpression plasmid (DDX5-FLAG) promoted *MYC* expression (**Figures 3.9C and 3.9D**).

We then analyzed the expression levels of DDX5 and *MYC* in multiple immortalized cell lines (**Figure 3.5**). We found a positive correlation between the expression of *MYC* and DDX5 in tumorigenic cell lines, which showed higher DDX5 and *MYC* expression levels than non-tumorigenic cell lines, suggesting the positive correlation between *MYC* and DDX5 promotes tumorigenesis. These results indicate DDX5 directly transactivates the *MYC* transcription through the G4-forming promoter region.

### **3.2.6 DDX5-induced MycG4-unfolding and *MYC* gene activation are inhibited by G-quadruplex-interactive compounds**

G4 structures can be recognized and stabilized by small molecules. TMPyP4 is a G4-interactive compound, whereas its positional isomer, TMPyP2, is a poor G4-interactive compound<sup>57, 144</sup> (**Figure 3.12A**). We examined the effects of TMPyP4 and TMPyP2 on the MycG4 unfolding activity of DDX5. Our results showed TMPyP4, but not TMPyP2, significantly inhibited the G4 unfolding activity of DDX5 (~ 70% inhibition) (**Figure 3.12B**). We then examined the effects of TMPyP4 and TMPyP2 on the direct binding of DDX5 to the MycG4. The results showed TMPyP4 disrupted the DDX5 binding to MycG4 in a dose-dependent manner, whereas TMPyP2 did not (**Figure 3.12C**). In addition, we examined Phen-DC3, another G4-interactive small molecule (**Figure 3.12A**), and found it disrupted the DDX5 binding to the MycG4 and inhibited MycG4 unfolding by DDX5 (**Figures 3.12B and 3.12C**). These data indicate that G4-interactive compounds can inhibit DDX5 G4-unfolding activity *in vitro* by preventing its association with G4 structure.

To investigate whether inhibition of DDX5 activity presents a new mechanism of action for G4-interactive small molecule in cancer cells, we examined the effects of TMPyP4 and TMPyP2 on the DDX5 and Myc levels in MCF7 cells. We found that treatments of both TMPyP4 and TMPyP2 showed no inhibitory effects on DDX5 expression up to 48 hr, whereas TMPyP4, but not TMPyP2, decreased Myc expression in both a time- and dose-dependent

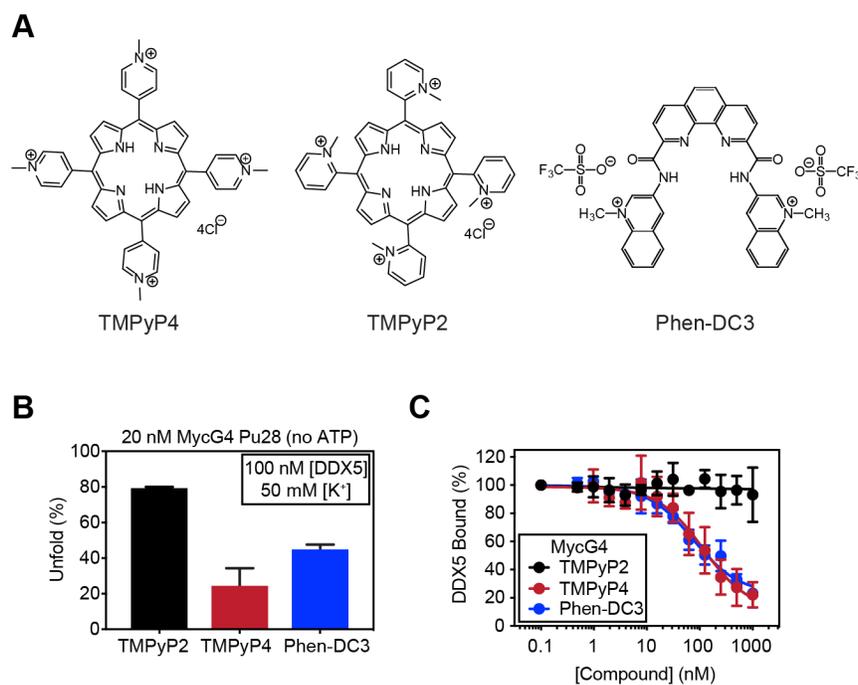


Figure 3.12. G-quadruplex-interactive compounds inhibit G-quadruplex-binding and -unfolding of DDX5

(A) Structures of TMPyP4, TMPyP2, and Phen-DC3. (B) G4-interactive compound TMPyP4 (250 nM) and Phen-DC3 (250 nM), but not poor-G4-interactive TMPyP2 (250 nM), inhibits MycG4 unfolding activity of DDX5. (C) Inhibition curves by ELISA showing that TMPyP4 and Phen-DC3, but not TMPyP2, disrupts the interaction of DDX5 to MycG4 DNA. DDX5 was co-incubated with increasing concentrations of either TMPyP2, TMPyP4, or Phen-DC3 for assessment of DDX5 binding to MycG4 in ELISA.  $n = 4-6$  independent measurements. Error bars represent mean  $\pm$  s.d.

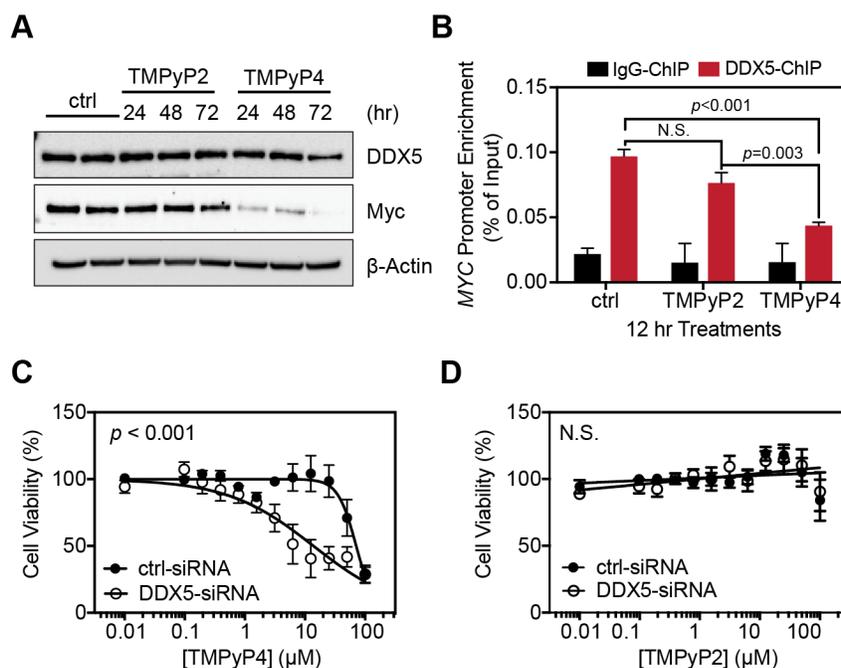


Figure 3.13. G-quadruplex-interactive molecules, such as TMPyP4, can disrupt transcriptional activation of DDX5 on *MYC* and lead to cancer cell death

(A) TMPyP4, but not TMPyP2, lowers Myc protein levels but does not affect DDX5 expression. The time of treatments with either TMPyP2 or TMPyP4 is listed at the top of the gels. (B) ChIP analysis of the interaction of DDX5 with the *MYC* promoter showing TMPyP4, but not TMPyP2, disrupts the interactions of DDX5 with the *MYC* promoter MycG4 forming region. The cells were either untreated (ctrl) or treated with 12  $\mu$ M TMPyP2 or TMPyP4 for 12 hr. For ChIP assay validation, see **Figure 3.6**.  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.d.  $p$ -value was calculated by Tukey's multiple comparisons test after two-way ANOVA. (C-D) siRNA knockdown of DDX5 increases sensitivity of MCF7 cells to the cell death caused by TMPyP4 (C), but not TMPyP2 (D). Data are mean  $\pm$  s.e.m. of 9 biologically independent samples. NS,  $p > 0.05$ ;  $p$ -value for the  $IC_{50}$  comparison was calculated by extra sum-of-squares F test.

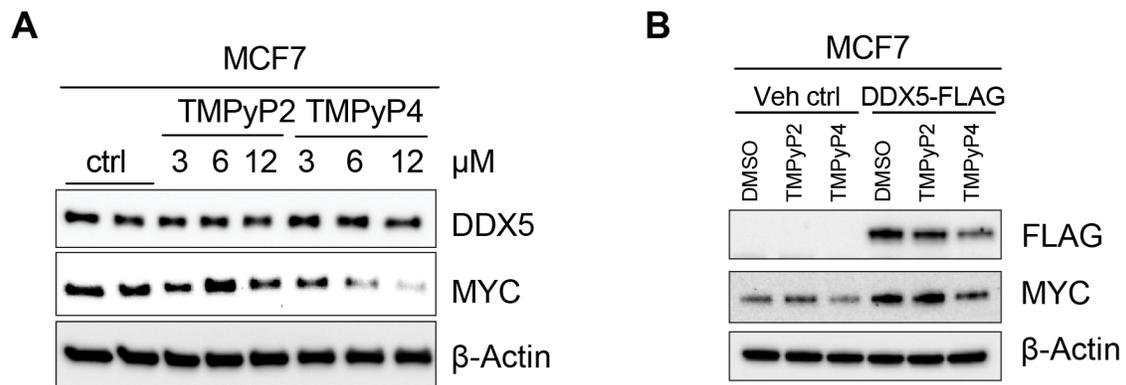


Figure 3.14. G4-interactive molecule TMPyP4 can disrupt transcriptional activation of DDX5 on *MYC*

(A) Treating MCF7 cells with TMPyP4, but not TMPyP2, for 24 hr led to a dose-dependent reduction of *MYC* expression, but no inhibitory effects were observed on DDX5 expression. (B) Overexpression of DDX5-FLAG is able to up-regulate *MYC* expression, which can be blocked by TMPyP4, but not TMPyP2.

manner (**Figure 3.13A** and **Figure 3.14A**). It is noted that while treating cells with TMPyP4 for 72 hr completely knocked out the Myc expression, the DDX5 expression was also slightly repressed (**Figure 3.13A**), likely reflecting previously reported positive feedback loop between MYC and DDX5<sup>185</sup>. Because 12  $\mu$ M of TMPyP4 induced significant Myc reduction within 48 hr (**Figure 3.14A**), this dosage was used in all the subsequent cellular studies.

We further examined the effects of TMPyP4 and TMPyP2 on the Myc levels when DDX5 was overexpressed (**Figure 3.14B**). Our results showed that while the overexpression of DDX5 up-regulated the Myc expression in MCF7 cells, TMPyP4 was still able to inhibit the Myc expression, suggesting G4-interactive small molecule is able to inhibit DDX5-induced MYC activation *in vivo*.

To confirm whether the reduced Myc expression by G4-interactive compounds is related to the inhibition of the DDX5 interactions with the MYC G4-forming promoter, ChIP analysis was carried out in MCF7 cells with 12 hr drug treatments. As a result, treating cells with TMPyP4, but not TMPyP2, led to a significant reduction (more than 50%,  $P < 0.001$  versus control) of DDX5 occupancy on MYC promoter (**Figure 3.13B**). It is important to note that DDX5 protein levels were not affected by the treatments of TMPyP4 or TMPyP2 up to 48 hr (**Figure 3.13A**). Taken together, these results indicate that the direct interactions of DDX5 with the MYC promoter G4 can be disrupted by G4-interactive compounds to suppress MYC expression.

### 3.2.7 DDX5 inhibition and MycG4 stabilization lead to synergistic lethality in cancer cells

We have demonstrated that DDX5 is a highly active MycG4 unfolding resolvase that activates MYC transcription in a G4-dependent manner. To investigate if DDX5 and MycG4 interaction can serve as a new molecular target for cancer intervention, we knocked down DDX5 prior to the addition of MycG4-stabilizing small molecule TMPyP4. We found that after DDX5 silencing, MCF7 cancer cells were more sensitive to the cell death induced by MycG4-interactive small molecule TMPyP4, but not TMPyP2 (**Figures 3.13C** and **3.13D**). This suggests that inhibition of DDX5 and stabilization of MYC promoter G4 can yield synergistic lethality in cancer cells supporting a role of DDX5 in active resolution of G4 structures *in vivo*. Altogether, these results demonstrate that G4-interactive ligands can target and block the DDX5 interactions

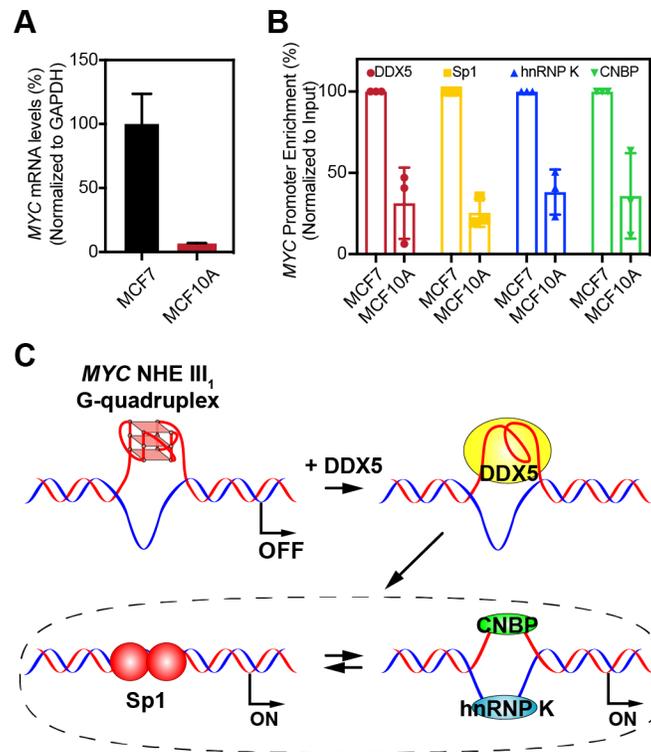


Figure 3.15. *MYC* transcriptional regulation via MycG4 in cancer cells

(A) The human breast adenocarcinoma cell line MCF7 has a high transcription level of *MYC* whereas the non-tumorigenic human breast epithelial cell line MCF10A has a low *MYC* level.  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.e.m. (B) DDX5, and previously reported transcription factors Sp1, hnRNP K, and CNBP, are more enriched at the *MYC* promoter G4-forming region in the MCF7 cell line compared to MCF10A cell line, suggesting the potential involvement in the active transcription of the *MYC* gene in tumorigenic cells.  $n = 3$  biologically independent samples. Error bars represent mean  $\pm$  s.e.m. (C) A model of transcriptional states associated with the *MYC* promoter G4 and *MYC* transcriptional regulation by DDX5. The DNA G-quadruplex that forms in the *MYC* promoter NHE III<sub>1</sub> region functions as a transcriptional silencer element by preventing the binding of transcription factors (top left); DDX5 unfolds the *MYC* promoter G-quadruplex (top right), likely enabling the binding of transcription factors, such as Sp1 to dsDNA or CNBP and hnRNP K to ssDNA<sup>80, 169</sup>, leading to transcriptional activation (bottom).

with the *MYC* promoter G4, thereby disrupt transcriptional activation of DDX5 on *MYC* and lead to *MYC* down-regulation and cancer cell death.

### 3.2.8 ChIP analysis of DDX5 and other transcription factors at the *MYC* promoter

We have shown that DDX5 is overexpressed in cancer cells (**Figure 3.5**) and enriched at the *MYC* promoter NHE III<sub>1</sub> (**Figure 3.9A** and **Figure 3.6B**). Moreover, depletion of DDX5 using siRNA can significantly decrease the transcriptional activation of the *MYC* gene in cancer cells (**Figures 3.9C-E** and **Figures 3.12** and **3.14**). Other transcription factors, including Sp1<sup>64</sup>, hnRNP K<sup>75</sup>, and CNBP<sup>140</sup>, have been reported to interact with the *MYC* promoter NHE III<sub>1</sub> for *MYC* transcriptional regulation. Tumorigenic MCF7 breast cancer cells were shown to have much higher *MYC* transcription levels as compared to non-transformed MCF10A breast epithelial cells (**Figure 3.15A**). We analyzed the enrichment of DDX5 and other transcription factors at the *MYC* promoter in MCF7 and MCF10A cells using ChIP analyses. The ChIP results showed that all of them, including DDX5, are enriched at the *MYC* promoter in MCF7 cells as compared to MCF10A cells (**Figure 3.15B**).

### 3.3 Discussion

We have discovered the G4-resolvase activity of DDX5, which actively unfolds the *MYC* promoter DNA G-quadruplex structure as well as other intramolecular DNA and RNA G4 structures. DDX5 is one of the founding members of the DEAD-box RNA helicase family<sup>172</sup> whose members have been shown to act only on RNA to date<sup>196</sup>. Thus, it was largely unexpected that DDX5 could unfold DNA G4 structures. We showed that DDX5 can recognize the G4 conformation through a structure-based mechanism since DDX5 showed similar binding activity to MycG4(DNA) and MycG4(RNA). Interestingly, we found that the unfolding of G4 by DDX5 occurred in the absence of ATP, whereas addition of ATP did not appreciably alter the unfolding activity. This shows that ATP hydrolysis is not directly coupled to the unfolding process.

We show that the unfolding mechanism of DDX5 is distinct from previously characterized G4 helicases, as DDX5 exhibits active G4-unfolding without requiring a single-stranded tail (as shown in this study). BLM and WRN, members of RecQ family helicases, were the first human helicases reported to resolve G4 DNA in human telomeres<sup>197</sup> and these enzymes function in a

3'→5' manner<sup>43</sup>. The DEAH-box helicase DHX9 and DHX36 (RHAU) were also shown to unfold both RNA and DNA G4 3'→5'<sup>198-200</sup>. Both DHX9 and DHX36 require a 3' single-stranded region, which is sequence-specific, for loading and G4-unfolding<sup>198-200</sup>. All of the reported active G4-helicases/resolvases are directional/processive and require a single-stranded tail at either the 3' or 5' end for loading and G4 unfolding<sup>43</sup>. Pot1 and RPA, two other proteins associated with G4 destabilization, were found to passively unfold G4 without a tail by trapping an unfolded, single-stranded DNA as it transits between conformations, thereby shifting the equilibrium to the unfolded state<sup>201-203</sup>. The DEAD-box RNA helicases DDX21, DDX1, and DDX3X were recently reported to unfold or interact with G4s, but only in RNA<sup>204-206</sup>. In these cases, the G4-unfolding mechanism is unknown. Thus, DDX5-mediated unfolding represents the first description of a G4-unfolding mechanism that is likely direct and also the first description of a DEAD-box RNA helicase activity on DNA.

Transcription-associated dynamic negative supercoiling at promoter sites proximal to the *MYC* transcription start site induces formation of the *MYC* promoter G4. This G4 is a transcription inhibitor by blocking binding of transcriptional activators such as Sp1 (double-stranded DNA) or CNBP and hnRNP K (single-stranded DNA)<sup>80, 169</sup> (**Figure 3.15C**). However, the *MYC* G-quadruplex structure is thermodynamically very stable under physiologically relevant salt conditions (melting temperature > 85 °C)<sup>143</sup>, thus requiring active unfolding *in vivo* for transcriptional activation. What may facilitate the unfolding has been a missing link. Here, we demonstrated that DDX5 actively unfolds the *MYC* promoter G4 and transactivates the *MYC* oncogene in a G4-dependent manner, defining a mechanism for how this structure can be resolved *in vivo* (**Figure 3.15C**). It has been shown that within the *MYC* promoter the G4 on the G-rich strand and i-motif secondary structure on the C-rich strand are mutually exclusive<sup>207</sup>. Resolution of the stable *MYC* G4 structure by DDX5 may allow the formation of i-motif on the C-rich strand and promote hnRNP K and CNBP binding, or the formation of ds DNA for Sp1 binding, to activate transcription.

Importantly, DDX5 is overexpressed in major human cancers and its overexpression promotes tumorigenesis and tumor progression<sup>174, 175, 177-179, 195</sup>. Our results suggest that DDX5 promotes cancer progression through *MYC* transactivation by unfolding of the *MYC* promoter G4. Significantly, we showed that DDX5 interactions with the *MYC* promoter and the DDX5-mediated unfolding of MycG4 can be inhibited by G4-interactive compounds, leading to *MYC* repression

and cancer cell death. This result indicates that small molecules targeting the interaction of DDX5 with the *MYC* promoter may be a future strategy for *MYC* downregulation.

In summary, we discovered DDX5 as a novel DNA G4 resolvase, elucidated a G4-dependent transactivation mechanism of *MYC* oncogene by DDX5, and established a potential new molecular target to downregulate *MYC* for cancer intervention.

### **3.4 Footnotes**

This chapter have been previously published as:

Wu, G., Xing, Z., Tran, E.\*, & Yang, D.\*, DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. *Proc Natl Acad Sci U S A.* 116, 20453-20461 (2019).  
(Copyright © 2019 National Academy of Sciences)

Author contributions: G.W., E.J.T., and D.Y. designed research; G.W. and Z.X. performed research; G.W., Z.X., E.J.T., and D.Y. analyzed data; and G.W. and D.Y. wrote the paper.

## CHAPTER 4. INDENOISOQUINOLINE TOPOISOMERASE INHIBITORS STRONGLY BIND AND STABILIZE THE *MYC* PROMOTER G-QUADRUPLEX AND DOWNREGULATE *MYC*

Reprinted with permission from

Kai-Bo Wang, Mohamed S. A. Elsayed, Guanhui Wu, Nanjie Deng, Mark Cushman and Danzhou Yang, Indenoisoquinoline Topoisomerase Inhibitors Strongly Bind and Stabilize the *MYC* Promoter G-Quadruplex and Downregulate *MYC*. *J Am Chem Soc* 141, 11059-11070 (2019)  
Copyright © 2019 American Chemical Society

### 4.1 Introduction

DNA is the target of many important anticancer agents, including human topoisomerase I inhibitors. Recently there has been significant progress in developing molecular-targeted therapies. A therapeutic advantage can be gained from DNA-targeted drugs combined with cancer-specific molecular targeting properties. Indenoisoquinolines are human topoisomerase I inhibitors with improved physicochemical and biological properties as compared to the traditional camptothecin topoisomerase I inhibitors that are clinically used for the treatment of various solid tumors.<sup>1-6</sup> Three indenoisoquinolines, indotecan (LMP400), indimitecan (LMP776), and LMP744 (**Figure 4.1A**), have entered phase I clinical trials in adults with relapsed solid tumors and lymphomas.<sup>7-14</sup> However, some indenoisoquinolines with potent anticancer activity did not show strong topoisomerase I inhibition,<sup>3, 15</sup> suggesting additional mechanism of action. Notably, high concentrations of some indenoisoquinoline compounds have been reported to target DNA outside of topoisomerase I action.<sup>6-7, 16-17</sup>

*MYC* is one of the most important oncogenes and is overexpressed in more than 80% of all types of cancer.<sup>18-19</sup> The transcription factor *MYC* protein is involved in cell proliferation, differentiation, and apoptosis, and plays a pivotal role in tumor initiation and progression as well as drug resistance.<sup>20-24</sup> *MYC* is found to be a general transcriptional “amplifier” in cancer cells.<sup>25-26</sup> Even a brief inhibition of *MYC* expression has been shown to permanently stop tumor growth and induce tumor regression *in vivo*,<sup>27</sup> because of the “oncogene addiction” of tumor cells.<sup>28</sup>

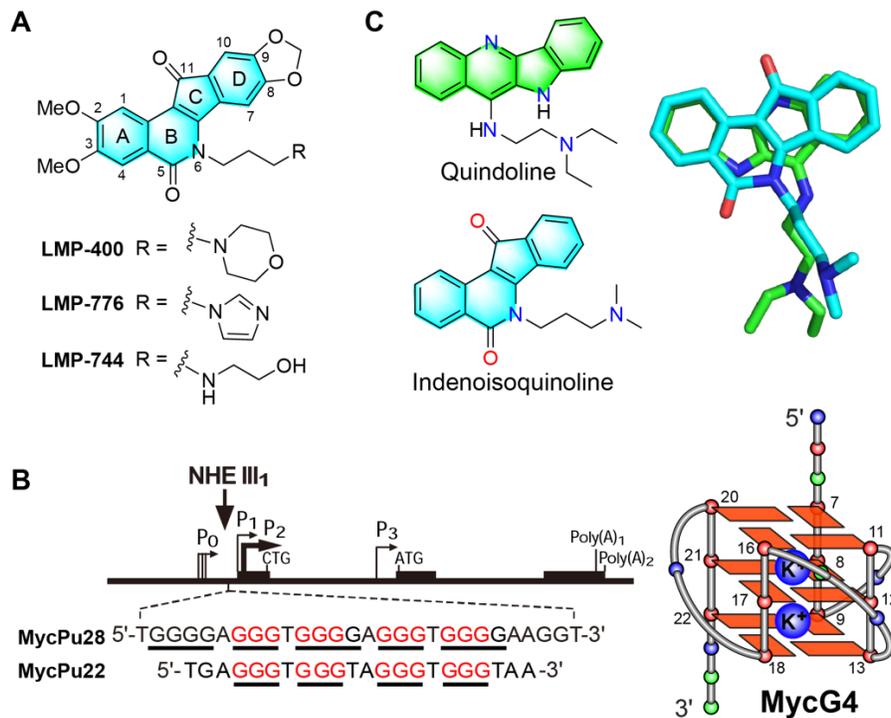


Figure 4.1. Chemical structures of indenoisoquinoline topoisomerase I inhibitors in Phase I clinical trials and quindoline, as well as the *MYC* promoter and *MYC* promoter G-quadruplex

(A) Indenoisoquinoline topoisomerase I inhibitors currently in clinical trials. (B) Left: The structure of the human *MYC* gene promoter. The G4-forming region NHE III<sub>1</sub> sequence is shown, with the guanine runs underlined. The guanine runs involved in the formation of the major MycG4 are highlighted in red. Right: The folding topology of MycG4 adopted by the MycPu22 sequence is a parallel-stranded 3-tetrad G-quadruplex, with the two stabilizing potassium cations shown. Red ball = guanine, green ball = adenine, blue ball = thymine, large blue ball = K<sup>+</sup>. (C) Left: a MycG4 stabilizer quindoline and a topoisomerase I inhibitor indenoisoquinoline. Right: overlay of the three-dimensional structures of quindoline and an indenoisoquinoline in their energy-minimized states.

Therefore, *MYC* is a hotly pursued therapeutic target. However, the *MYC* protein is not an easy drug target due to its short half-life and lack of a small molecule binding pocket.<sup>29-31</sup>

The nuclease hypersensitive element (NHE) III<sub>1</sub> in the *MYC* promoter, which controls 85-90% of *MYC* transcriptional activity, forms a DNA G-quadruplex (G4) under transcription-associated negative supercoiling and functions as a transcriptional silencer (**Figure 4.1B, left**).<sup>32-36</sup> DNA G-quadruplexes (G4s) are globular four-stranded secondary structures consisting of stacked Hoogsteen hydrogen-bonded G-tetrads stabilized by K<sup>+</sup> or Na<sup>+</sup>.<sup>37</sup> DNA G-quadruplexes found in promoter regions of key oncogenes have emerged as a promising new class of cancer-specific molecular targets for drug development.<sup>38-40</sup> Using a G4-specific antibody, G4 structures have been visualized in human cells at both telomeric and non-telomeric sites on chromosomes, and G4-loci increase after exposure of live cells to G4 ligands.<sup>41</sup> G4s detected in immortalized precancerous cells are at 10 times higher levels than in normal human cells, and G4-sites are found to be specifically enriched in regulatory, transcriptionally active regions of chromatin, particularly the *MYC* promoter region.<sup>42</sup> We previously determined the structures of the *MYC* promoter G-quadruplexes.<sup>43-44</sup> The major *MYC* promoter G-quadruplex (MycG4) is a parallel-stranded structure with three G-tetrads connected by three propeller loops (**Figure 4.1B, right**).<sup>32, 43, 45</sup> Significantly, stabilization of the *MYC* promoter G-quadruplex by small molecules suppresses *MYC* transcription.<sup>32, 36, 46</sup> For example, a quindoline anticancer agent was shown to stabilize the *MYC* G-quadruplex and downregulate *MYC*.<sup>46-47</sup> We have determined the molecular structure of the 2:1 quindoline-MycG4 complex, which shows specific recognition of the MycG4 by the crescent-shaped quindoline.<sup>48</sup> Interestingly, indenoisoquinolines are crescent-shaped and share structural similarity with the quindoline compound (**Figure 4.1C**), which is consistent with the report that 6-substituted indenoisoquinolines<sup>15</sup> bind the c-Kit promoter G4s.<sup>17</sup>

Herein, using fluorescence resonance energy transfer (FRET) assays, nuclear magnetic resonance (NMR), fluorescence-based binding assay and competition fluorescence displacement assay, circular dichroism (CD) spectroscopy, and gel electromobility shift assay (EMSA), we demonstrate that a large number of anticancer indenoisoquinolines strongly bind and stabilize MycG4 *in vitro*. Using cell-based western blotting and quantitative reverse transcription-polymerase chain reaction (qRT-PCR) assays, we show that MycG4-interactive indenoisoquinolines lower *MYC* mRNA and protein levels *in vivo*, indicating that targeting the *MYC* promoter G4 to downregulate *MYC* is a likely mechanism of action for the anticancer

activities of these particular indenoisoquinolines. Furthermore, some active indenoisoquinolines show both *MYC* downregulation and topoisomerase I inhibition, suggesting that dual-targeting of MycG4 and topoisomerase I could be a potential strategy for anticancer drug development.

## 4.2 Results

### 4.2.1 Indenoisoquinolines can induce and stabilize MycG4

To examine whether the indenoisoquinolines could induce and stabilize the MycG4, we conducted a FRET-quenching assay on indenoisoquinoline compounds. The full-length *MYC* promoter NHE III<sub>1</sub> G4 DNA (MycPu28, **Figure 4.1B**) was labeled with FAM (6-fluorescein) on the 3'-end and BHQ-1 (Black Hole-1 quencher) on the 5'-end (**Figure 4.2A, left**). The MycG4 structure adopted by MycPu22 (**Figure 4.1B**) is the major conformation formed by the wild-type MycPu28 in K<sup>+</sup> solution.<sup>32, 43, 45</sup> Dual-labeled MycPu28 was used for the FRET-quenching screening assay because of its higher FAM-fluorescence in the unfolded form which can be attributed to the longer distance between the FAM and BHQ quencher, and thus provided greater range for screening. We confirmed that very similar FRET-quenching effects were observed for MycPu22 and MycPu28 upon compound binding and G4-stabilization. The stable formation of G-quadruplexes requires the presence of K<sup>+</sup> or Na<sup>+</sup> cations in solution, with a preference of K<sup>+</sup> (**Figure 4.1B**). In the absence of K<sup>+</sup>, the MycPu28 is in the extended single-stranded form with its two ends far apart and shows high FAM-fluorescence (**Figure 4.2A, left**). In the presence of 100 mM K<sup>+</sup>, the G4 is folded and the FAM-fluorescence is quenched because the quencher and fluorophore at the two ends are in closer proximity (**Figure 4.2A, left**). Alternatively, the addition of G4-stabilizing ligands can induce G4 formation in the absence of K<sup>+</sup> and thereby lead to quenching FAM-fluorescence (**Figure 4.2A, left**).

We examined 56 indenoisoquinoline compounds using this FRET-quenching assay (**Figure 4.2A**). 100 mM K<sup>+</sup> buffer was used as a positive control, which decreased FAM-fluorescence by 39%. We found that 37 compounds decreased the FAM-fluorescence by more than 39%, indicating that these indenoisoquinolines can induce and stabilize MycG4. Some indenoisoquinolines decreased FAM-fluorescence more than 100 mM K<sup>+</sup>, which is likely due to greater stabilization of MycG4 or its flanking structures. However, it cannot be ruled out that some

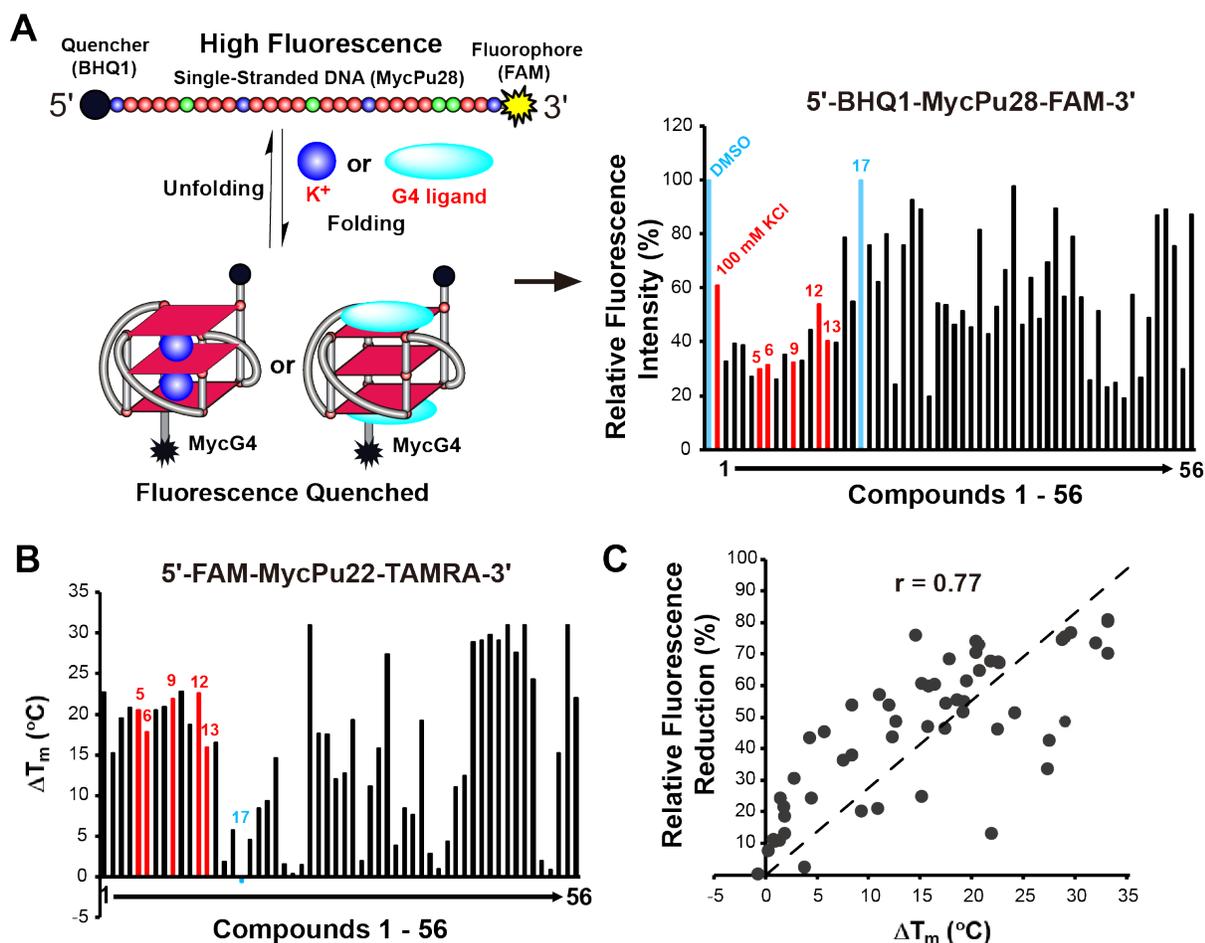


Figure 4.2. Indenoisoquinolines can induce and stabilize MycG4

(A) Left: schematic of the FRET-quenching assay used for compound screening. The FRET-quenching (shown as fluorophore in black color) caused by MycG4 folding can be induced by K<sup>+</sup> or MycG4-inducing compounds. Right: relative fluorescence intensities of the labeled MycG4 in the presence of DMSO, 100 mM K<sup>+</sup>, and indenoisoquinoline analogs as shown by FRET-quenching assay. Data shown are the average values of the two individual experiments. DMSO (negative control), 100 mM K<sup>+</sup> (positive control), and six indenoisoquinolines used for further studies are highlighted and labeled. Conditions: 1  $\mu$ M labeled DNA, 10  $\mu$ M compound, 25  $^{\circ}$ C, 50 mM Tris-acetate, pH 7. (B) Thermal stabilization values ( $\Delta T_m$ ) of MycG4 by indenoisoquinoline analogs as shown by FRET-melting assay. Data shown are the average values of the two individual experiments. The six representative indenoisoquinolines used for further studies are highlighted and labeled. Conditions: 150 nM labeled DNA, 1.5  $\mu$ M compound, 25  $^{\circ}$ C, pH 7, 10 mM K<sup>+</sup>. (C) Correlation of FRET-quenching and FRET-melting data. The Pearson correlation coefficient ( $r$ ) is shown.

indenoisoquinolines may interact with the FAM fluorophore directly to quench the FAM-fluorescence.

To confirm the stabilizing effect of indenoisoquinolines on MycG4, the  $T_m$  values of MycG4 were measured in the presence of indenoisoquinoline compounds in 10 mM  $K^+$  using dual-3'-FAM- and 5'-TAMRA-labeled MycPu22 DNA by FRET-melting experiments. MycPu22 DNA forms a single MycG4 structure and was used for NMR structure determination (**Figure 4.1B, right**).<sup>43</sup> Therefore, MycPu22 provides the best molecular system for MycG4 and was used in all the subsequent experiments. 10 mM  $K^+$  was used in the FRET-melting experiments because the melting temperature of MycG4 at 100 mM  $K^+$  is above 90 °C, making it impossible to determine an accurate melting temperature upon compound addition.<sup>49</sup> The FRET-melting results showed that forty-four of the fifty-six indenoisoquinolines increased the  $T_m$  values of MycG4 by more than 5 °C (**Figure 4.2B**). A clear positive correlation was observed between the indenoisoquinolines' ability to induce MycG4 formation and to increase its thermal stability (**Figure 4.2C**).

#### 4.2.2 Some indenoisoquinolines significantly lower *MYC* levels in cancer cells

G-quadruplex formed in the *MYC* promoter was found to function as a transcriptional silencer.<sup>32-34</sup> To determine the effects of indenoisoquinolines on the *MYC* protein level, a western blotting experiment was carried out using MCF-7 breast cancer cells treated with 44 indenoisoquinolines that increased the  $T_m$  value of MycG4 by more than 5 °C. MCF-7 cells were incubated with each compound at four concentrations (0.5, 1, 2, and 4  $\mu$ M) for 24 hours, and the *MYC* protein levels were measured (**Figure 4.3A**).

The human topoisomerase I inhibitory activities of the 44 indenoisoquinolines have been previously determined.<sup>3, 6-9, 15, 50-52</sup> Of the 44 compounds tested for their cytotoxicities in the NCI-60 cancer cell lines, the 31 most potent compounds had their mean graph midpoint (MGM) values determined based on the  $GI_{50}$  values obtained from NCI-60 cancer cell lines drug screen.<sup>53-55</sup> The topoisomerase I inhibitory activities were plotted against the anticancer activities of these 31 compounds (**Figure 4.3B**). Some of the more active compounds (with MGM values < 0.5  $\mu$ M) showed strong topoisomerase I inhibition. However, many of the active compounds were not strong topoisomerase I inhibitors. The *MYC* inhibition activities of these compounds were ranked in four groups, i.e., strong, medium, weak, and no inhibition (**Figure 4.3B**). Significantly, strong

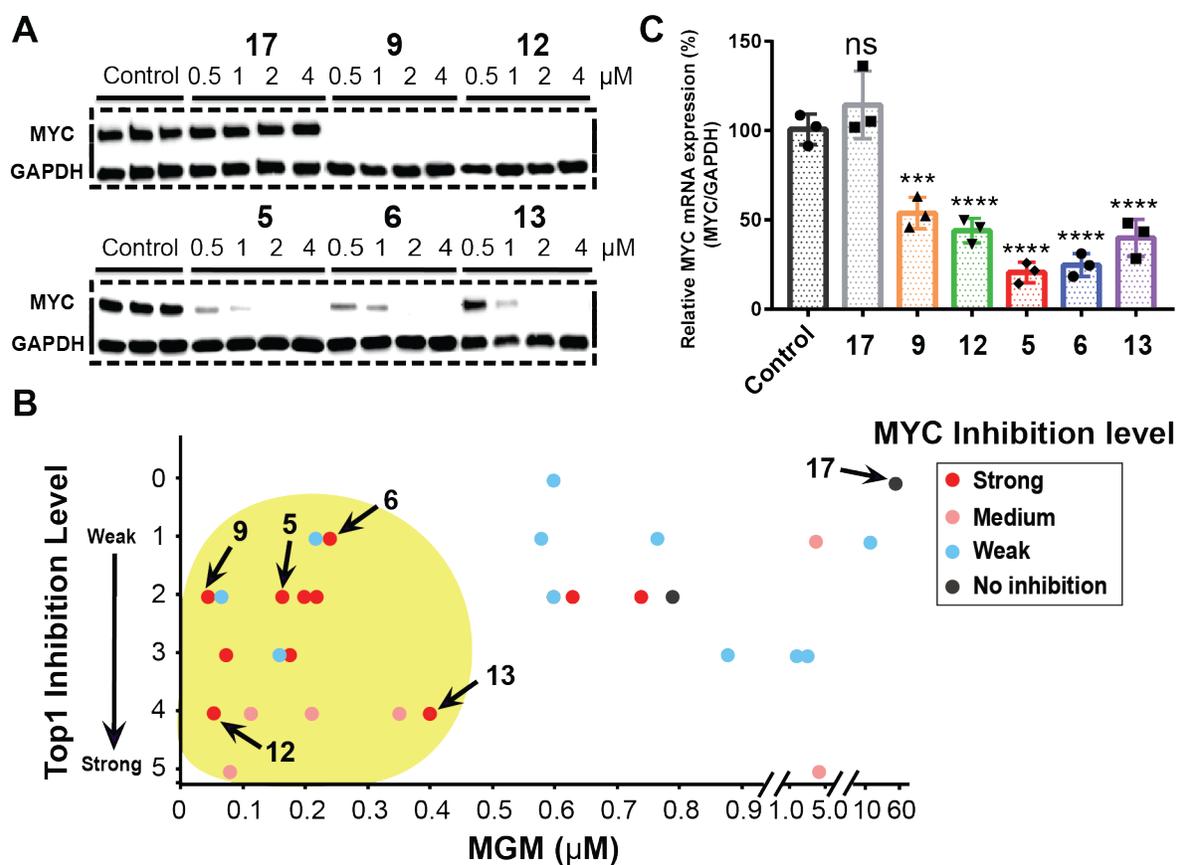


Figure 4.3. *MYC* inhibitory activities of indenoisoquinoline analogs

(A) *MYC* protein expression levels in the absence and presence of various concentrations of indenoisoquinolines (24 hr treatment) were obtained by western blotting experiments in MCF-7 breast cancer cells. GAPDH was used as an internal control. (B) Plot of the topoisomerase I inhibition levels against the MGM values of 31 indenoisoquinolines that were used to determine topoisomerase, *MYC*, and MGM activities. The yellow shaded area indicated the region of more active indenoisoquinolines. Based on the *MYC* downregulation shown in the western blotting results (Figures 4.3A), *MYC* inhibition levels were classified into four levels: strong inhibition, *MYC* expression inhibited at 0.5 to 1.0  $\mu\text{M}$ , red dots; medium inhibition, *MYC* expression inhibited at 2.0  $\mu\text{M}$ , or no clear dose-dependent *MYC* inhibition, pink dots; weak inhibition, *MYC* expression inhibited at 4.0  $\mu\text{M}$ , blue dots; no inhibition, no *MYC* expression inhibition up to 4.0  $\mu\text{M}$ , black dots. The relative topoisomerase I (Top1) inhibition levels of the compounds were previously determined and classified into six levels (0 – 5).<sup>3, 6-9, 15, 50-52</sup> The MGM values are the average of  $\text{GI}_{50}$  values across the entire panel of NCI-60 cancer cell lines; the  $\text{GI}_{50}$  values are the concentrations corresponding to 50% growth inhibition which were determined in the NCI-60 cancer cell lines drug screen. (C) *MYC* transcription levels in the absence and presence of indenoisoquinolines (6 hr treatment) were obtained by qRT-PCR experiments in MCF-7 cancer cells. DMSO was used as the negative control (no inhibition, 100%). The relative *MYC* mRNA levels were normalized with *GAPDH*. The experiments were run in triplicate. P values (\*\*P = 0.0004, \*\*\*\*P = 0.0001) for the comparison of DMSO control group were obtained by one-way ANOVA analysis.

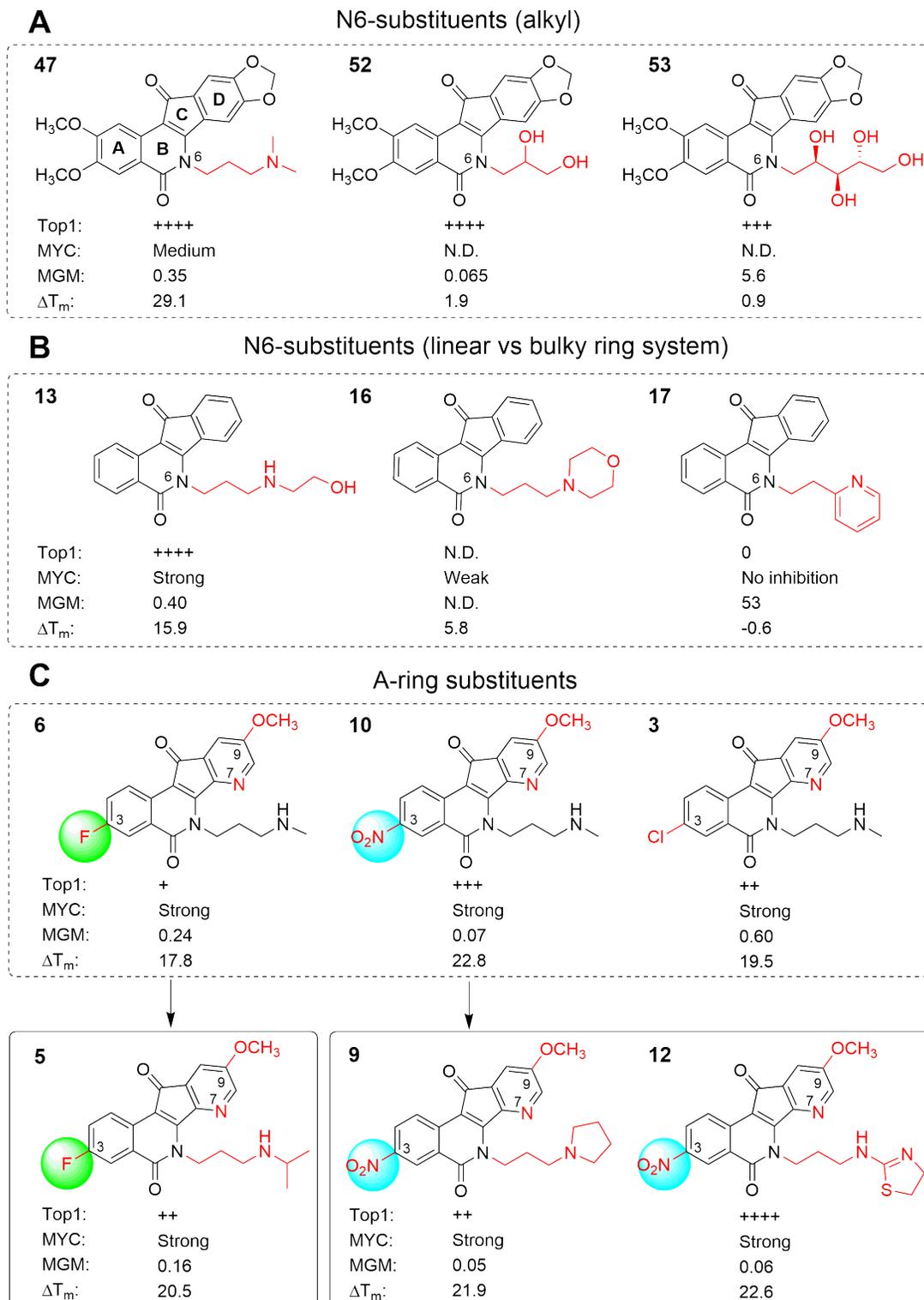


Figure 4.4. SAR of selected indenoisoquinolines. N.D., not determined

*MYC* inhibition was concentrated in compounds with potent anticancer activities, including those showing weak topoisomerase I inhibitory activity (**Figure 4.3B**). We selected compounds **5**, **6**, **9**, **12** and **13** for further investigation as they showed clear *MYC*-inhibitory effect (**Figures 4.3B and 4.4**). Compound **17** was used as a negative control (**Figures 4.3B and 4.4**).

To confirm the effect on the transcription of the *MYC* gene in cancer cells by the six selected indenoisoquinoline compounds, the *MYC* mRNA levels in MCF-7 cancer cells were measured by qRT-PCR. Consistent with the western blotting data, all five *MYC*-inhibiting compounds significantly lowered *MYC* mRNA levels at 6 hours post the treatments with 1  $\mu$ M indenoisoquinolines. The negative control compound **17** showed no reduction of *MYC* mRNA level (**Figure 4.3C**).

### 4.2.3 *MYC*-inhibiting indenoisoquinolines are strong MycG4 binding ligands

The binding interactions of six selected indenoisoquinolines with MycG4 were examined using  $^1\text{H}$  NMR titration experiments in  $\text{K}^+$ -containing solution. The free MycG4 DNA shows 12 imino proton peaks of guanines from the three G-tetrads (**Figure 4.5**).<sup>43, 48</sup> Upon respective addition of the five MycG4-interactive indenoisoquinolines, clear changes of the tetrad-guanine imino proton signals were observed, confirming the binding of these compounds to MycG4 (**Figure 4.5A-E**). The binding appeared to be in the medium-to-fast exchange rate on the NMR timescale, as shown by the broadening of DNA proton peaks at lower drug equivalence (0.5 and 1) and the sharpening at higher drug equivalence (2 and 3). Indenoisoquinolines appeared to bind at both ends of the MycG4, as shown by the imino proton peaks corresponding to both of the 3'- and 5'- tetrads being significantly shifted upon drug addition. Three *MYC*-inhibiting compounds, the 7-azaindenoisoquinolines **5** and **6**, and the indenoisoquinoline **13**, showed more specific binding to MycG4, where a well-defined complex was shown to form at the drug equivalence of 3, with a new set of 12 imino proton peaks. For compounds binding at intermediate exchange rates on the NMR time scale, a compound:DNA ratio higher than its binding stoichiometry is needed to push the equilibrium towards the formation of a stable drug-DNA complex, as shown by the sharp, well-resolved proton peaks.<sup>48,56</sup> In contrast, the negative control compound **17** did not show any binding as no change was observed in the  $^1\text{H}$  NMR spectra upon titration (**Figure 4.5F**). The MycG4 complexes of the five MycG4-interactive indenoisoquinolines were monomeric in nature as shown by native EMSA gels.

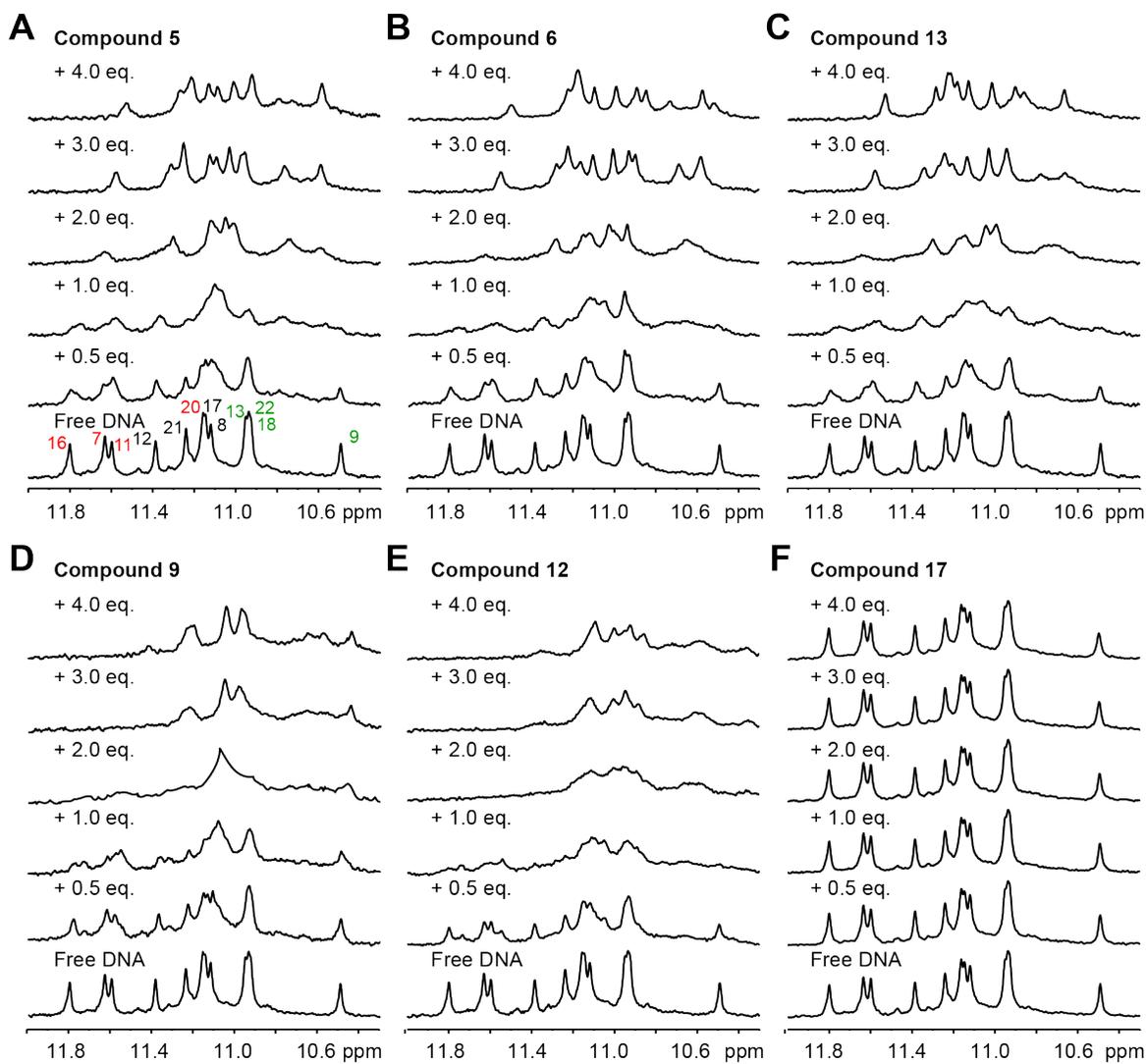


Figure 4.5. 1D  $^1\text{H}$  NMR titrations of MycPu22 DNA with indenoisoquinolines and 7-azaindenoisoquinolines

Imino proton regions of the titration spectra of MycG4 with compound **5** (A), **6** (B), **13** (C), **9** (D), **12** (E), and **17** (F) are shown. In Figure 5.5A, the imino proton signals from the 5' G-tetrad (Figure 5.1B) are labeled in red, the middle G-tetrad in black, and the 3' G-tetrad in green. Conditions: 150  $\mu\text{M}$  DNA, 25  $^\circ\text{C}$ , pH 7, 100 mM  $\text{K}^+$ .

CD titration experiments with MycG4 were also carried out for the six selected indenoisoquinolines. The free MycPu22 DNA in K<sup>+</sup> buffer showed the CD signature of a parallel G-quadruplex, with a positive peak at 264 nm and a negative peak at 242 nm.<sup>57</sup> Upon addition of indenoisoquinolines, the CD signature of a parallel G-quadruplex was maintained. The five MycG4-interactive compounds showed a slight decrease in intensity for both the positive peak at 264 nm and the negative peak at 242 nm, likely due to the ligand-induced capping structure formation by the flanking segments. The decrease in intensity in CD spectra of G4 upon ligand binding has been previously reported.<sup>58</sup> The negative control compound **17** showed no effect on the CD spectrum.

Binding affinities of these six indenoisoquinolines to MycG4 were measured using a 3'-TAMRA-labeled MycPu22 DNA.<sup>59</sup> The five MYC-inhibiting compounds showed strong binding with apparent binding affinity  $K_d$  values of 5.6 – 23.9 nM, whereas the negative control compound showed negligible binding. The indenoisoquinolines show negligible fluorescence in either the free or bound state.

#### 4.2.4 Molecular docking study of the binding of indenoisoquinoline **5** to MycG4

NMR titration data showed that 7-azaindenoisoquinoline **5** binds MycG4 to form a well-defined complex at both the 5'- and 3'-ends, as is evident by the significant shifting of the imino proton peaks of the 5'- and 3'-external tetrad guanines (**Figure 4.5A**). We have previously determined the NMR structure of the 2:1 quindoline:MycG4 complex in K<sup>+</sup> solution (PDB ID 2L7V), in which quindoline binds MycG4 at both ends to form a 5'-complex and 3'-complex.<sup>48</sup> As indenoisoquinolines are structurally similar to the quindoline compound (**Figure 4.1C**), we performed a molecular docking study to explore the possible binding modes of 7-azaindenoisoquinoline **5** with the MycG4 based on the NMR structure of the 2:1 quindoline:MycG4 complex. The docking program Glide was used in the standard precision (SP) mode: see Methods.<sup>60-61</sup> 7-Azaindenoisoquinoline **5** was docked to the binding sites at the two ends of the MycG4 using the 2:1 quindoline:MycG4 complex structure (**Figure 4.6**). Similar binding poses were predicted by the docking experiment for both the 5'- and 3'-sites. Docking studies gave docking scores for the 5'- and 3'-complexes at -6.69 and -6.08 kcal/mol, respectively. Figure 5.6 shows a representative model of the 2:1 7-azaindenoisoquinoline **5**:MycG4 complex.

The overall binding modes of indenoisoquinoline resembled those of quindoline in the NMR structure of the 2:1 quindoline:MycG4 complex, in which a flanking DNA base from the 5'- or 3'-flanking segment was recruited to form a ligand-base plane stacking over the external tetrads, except that no H-bond existed in the 3'-complex between the indenoisoquinoline and the recruited base. Notably, the tetracyclic ring scaffold of 7-azaindenoisoquinoline **5** with A- and D-ring substituents stacks very well with both the 5'- and 3'-external tetrads, making extensive stacking interactions. The positively charged amine side chain of indenoisoquinoline **5** resides in the MycG4 groove and forms intermolecular salt bridges with phosphate groups on the nucleotide backbone.

#### **4.2.5 Binding selectivity of MycG4-interactive indenoisoquinolines and 7-azaindenoisoquinolines**

Using a competition fluorescence displacement assay, we determined the binding selectivity of five indenoisoquinolines for MycG4 as compared to a parallel K-Ras promoter G4, a hybrid telomeric G4, and double-stranded (ds) DNA at 1 and 5 equivalents of each compound (**Figures 4.7**). The 3'-TAMRA labeled MycPu22 DNA was used as the fluorescence probe, whose fluorescence was quenched upon the binding of indenoisoquinolines. Upon addition of unlabeled, non-fluorescent competitors (e.g. other DNA G4s and dsDNA), the TAMRA-labeled MycPu22 DNA is displaced by the competitor DNA for indenoisoquinoline binding and the initial high TAMRA-fluorescence is restored. The competition fluorescence displacement assay allows for a straightforward assessment of selective binding towards MycG4 vs. the competitors, i.e. MycG4s (parallel), K-Ras G4 (parallel), telomeric G4 (hybrid), and dsDNA. One and five compound equivalents were used to assess the selectivity of the strongest binding site and other binding sites of each indenoisoquinoline. As shown in **Figure 5.7**, all five MycG4-interactive indenoisoquinolines showed marked binding selectivity for parallel G4s (MycG4s and K-Ras G4) over dsDNA (**Figure 4.7**), and this selectivity became more pronounced at higher compound ratio. Significantly, four 7-azaindenoisoquinolines, **5**, **6**, **9**, and **12**, showed remarkable selectivity for DNA G4s over dsDNA. However, indenoisoquinoline **13**, which has only N6-substitution but no A- and D-ring substituents, showed much less selectivity against dsDNA. This result suggested that substituents on the A- and D-rings are important for selective binding of G4s vs dsDNA. As shown in the modeling study, the substituents on the A- and D-rings of indenoisoquinolines likely

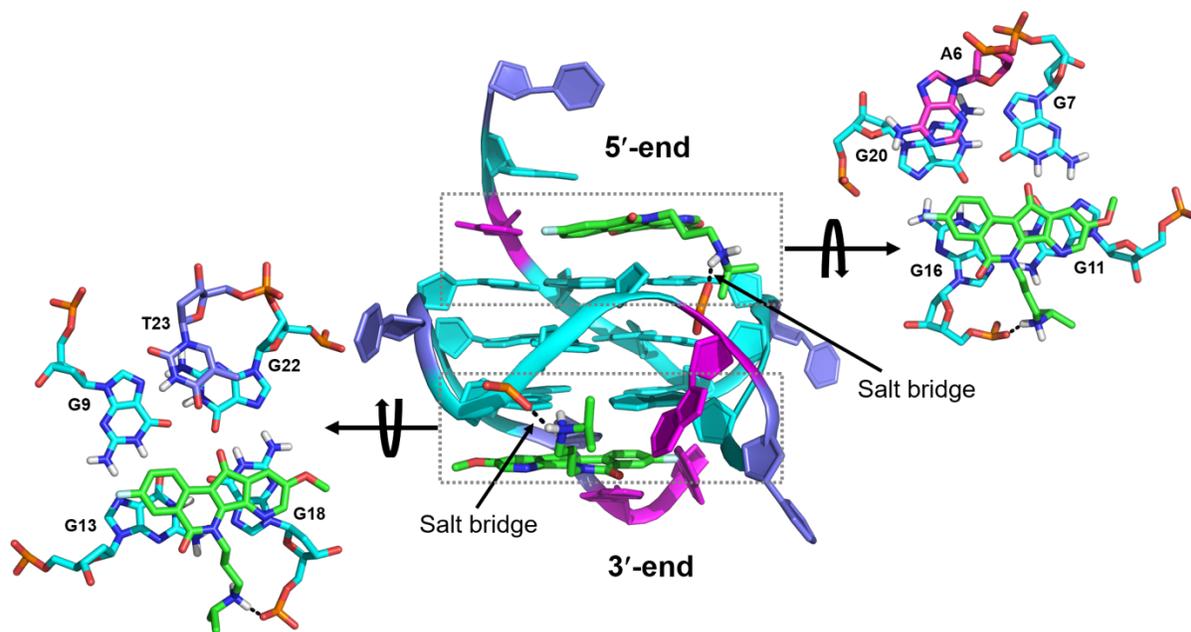


Figure 4.6. A model of the 2:1 complex of 7-azaindenoisoquinoline 5:MycG4 suggested by Glide docking in different views

Indenoisoquinoline 5 were shown as green sticks. Intermolecular salt bridges are shown as black dashed lines.

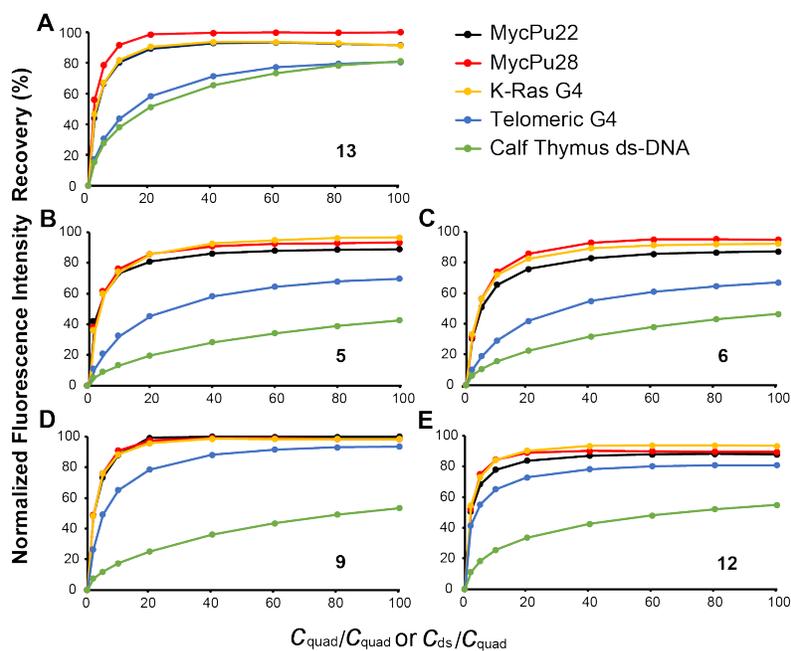


Figure 4.7. Binding selectivity of MycG4-interactive indenoisoquinolines

Competition fluorescence displacement experiments with increasing concentrations of unlabeled G4s and dsDNA added to 3'-TAMRA-labeled MycPu22 (20 nM) mixed with 1 equivalent of compound **13** (A), **5** (B), **6** (C), **9** (D), and **12** (E). The normalized TAMRA fluorescence intensities at 580 nm were plotted as a function of molar ratio of added G4 DNA (in 3 G-tetrads) or calf thymus dsDNA (in 11 bp) to labeled MycPu22 DNA. The fluorescence intensity of free 3'-TAMRA labeled MycPu22 was defined as 100%, and 1:1 mixture of 3'-TAMRA labeled MycPu22 and indenoisoquinoline was defined as 0%. Conditions: 20 °C, pH 7, 100 mM K<sup>+</sup>.

contribute to binding MycG4 by more optimal stacking interactions with the external G-tetrads. On the other hand, the increased size of the indenoisoquinoline ring system may hinder intercalation in dsDNA due to possible steric collision with the DNA backbone. Interestingly, the 3-fluoro-substituted 7-azaindenoisoquinolines **5** and **6** showed marked selectivity for parallel G4s over hybrid G4, whereas the 3-nitro-substituted 7-azaindenoisoquinolines **9** and **12** showed much less selectivity, suggesting that the 3-nitro-group may contribute to a less-specific interaction. The less-specific interaction of 7-azaindenoisoquinolines **9** and **12** was also supported by the NMR titration data showing less well-defined MycG4 complexes formed with **9** and **12** (**Figure 4.5**). Albeit with low selectivity against dsDNA, 6-substituted indenoisoquinoline **13** showed selectivity for parallel G4s over hybrid G4. 6-Substituted indenoisoquinolines were previously reported to bind to the c-Kit promoter G4s which were also primarily parallel.<sup>17</sup>

#### 4.2.6 Structure-activity relationship of MycG4 binding by indenoisoquinolines

To understand the factors that govern indenoisoquinoline recognition for MycG4, indenoisoquinoline analogues were analyzed for their MycG4 interactions and MYC inhibitory activity. Clear trends could be established to generate structure-activity relationships for MycG4 binding (**Figure 4.4**). It was shown that N6-substituents play a critical role in MycG4 binding and stabilization (**Figures 4.4A-B**). For example, indenoisoquinoline **47** with an N6-dimethylaminopropyl moiety, showed medium MycG4 stabilizing activity, whereas indenoisoquinolines **52** and **53**, which lack the aminopropyl side chain structure, were found to be poor MycG4 binders and stabilizers. These results suggested that an alkyl amine-containing side chain at N6 of ring B is important for MycG4 binding (**Figure 4.4A**), likely due to the favorable electrostatic interactions between the positively charged N-containing side chain and the negatively charged phosphate backbone in the groove of MycG4 at physiological pH 7.4. However, this favorable interaction (compound **13**) appeared to be weakened by a more bulky N-containing ring-system (compound **16**), and abolished by an aromatic N-containing ring-system (compound **17**, reduced positive charge for N) (**Figure 4.4B**), suggesting that the bulky nitrogen-containing group may sterically hinder the binding.

Significantly, the newer generation 9-methoxy-7-azaindenoisoquinolines, which were developed with improved water solubility and increased charge-transfer properties,<sup>62-63</sup> appeared

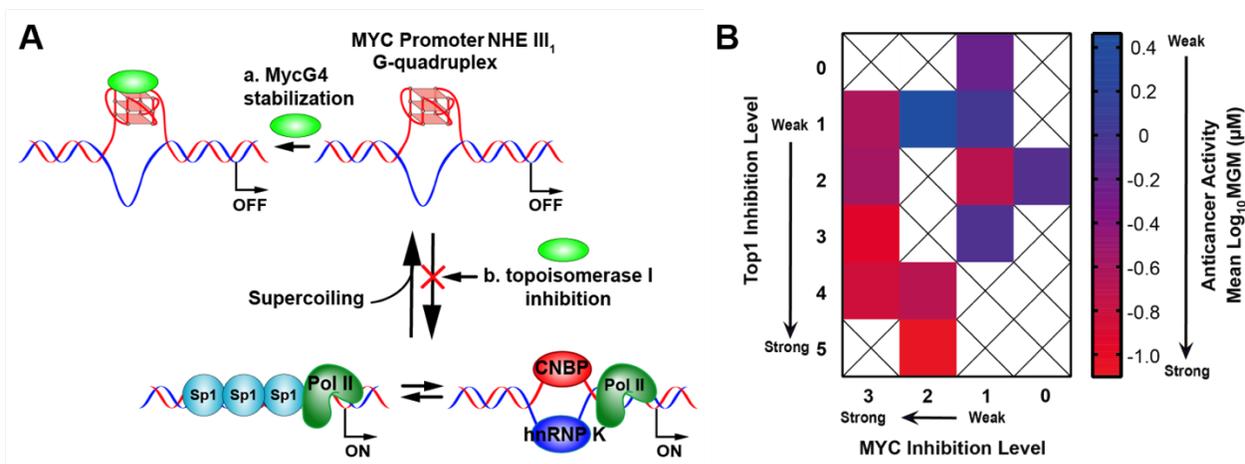


Figure 4.8 Dual inhibition of *MYC* and topoisomerase I by indenoisoquinolines

(A) A schematic model showing the potential mechanisms of *MYC* suppression by indenoisoquinolines by (a) stabilization of MycG4 in the *MYC* promoter to inhibit transcription, and (b) inhibition of topoisomerase I to maintain negative supercoiling for G4 formation. (B) A heat map showing the synergistic effect of *MYC* inhibition and topoisomerase I inhibition on the anticancer activities of 29 indenoisoquinolines. The 29 indenoisoquinolines are grouped by their *MYC* inhibition levels and topoisomerase I inhibition levels. The anticancer activity for each group is determined by the mean( $\log_{10}$ MGM) value of the grouped compounds, which is displayed as color gradient in the heat map. The MGM values are the approximate average of  $GI_{50}$  values across the entire panel of NCI-60 cancer cell lines for each compound. The synergistic effect of *MYC* inhibition and topoisomerase I inhibition is reflected by the increased anticancer activities (redder color) towards the bottom left corner with strong *MYC* and topoisomerase I inhibitory activities.

to bind MycG4 well and show potent MYC-inhibitory activity (**Figure 4.4C**). 7-Azaindenoisoquinolines with small substituents, such as 3-fluoro-, 3-nitro-, and 3-chloro, on the A-ring were found to be strong MycG4 binders and stabilizers and showed potent MYC-inhibitory activity.

### 4.3 Conclusion

This study demonstrates that anticancer indenoisoquinolines and 7-azaindenoisoquinolines strongly bind and stabilize MycG4 and lower *MYC* levels in cancer cells as revealed by various biophysical, biochemical, computer modeling, and cell-based experiments. Significantly, a large number of active indenoisoquinolines and 7-azaindenoisoquinolines caused strong *MYC* downregulation. Indenoisoquinoline analogs are clinically useful anticancer drugs and present a promising scaffold for MycG4-targeting anticancer drug development (**Figure 4.8A**). The insights into structure-activity-relationships of MycG4 recognition by indenoisoquinolines were also elucidated. In addition, some active indenoisoquinolines and 7-azaindenoisoquinolines were shown to cause both *MYC* downregulation and topoisomerase I inhibition. The analysis of indenoisoquinoline analogues for their *MYC* inhibitory activity, topoisomerase I inhibitory activity, and anticancer activity led to the discovery of a synergistic effect of *MYC* inhibition and topoisomerase I inhibition on anticancer activity (**Figures 4.8B**). Notably, topoisomerase I specifically relaxes the transcription-induced negative supercoiling,<sup>4</sup> which is the key to the formation of the *MYC* promoter G4 (**Figure 4.8A**), thus dual targeting of MycG4 and topoisomerase I could be an effective mechanism of action for cancer intervention. Collectively, the results uncover a novel mechanism of action of the clinically useful indenoisoquinoline scaffold as a new family of drugs targeting MycG4 for *MYC* downregulation. Furthermore, the study suggests that dual targeting of *MYC* and topoisomerase I may serve as a novel strategy for anticancer drug development.

### 4.4 Footnotes

This chapter has been previously published as:

Wang, K.B., Elsayed, M.S.A., Wu, G., Deng, N., Cushman, M. & Yang, D. Indenoisoquinoline topoisomerase inhibitors strongly bind and stabilize the *MYC* promoter G-quadruplex and

downregulate *MYC*. *J Am Chem Soc* 141, 11059-11070 (2019). (Copyright © 2019 American Chemical Society)

Author contributions: K.B.W., M.S.A.E, G.W., M.C., and D.Y. designed research; K.B.W., M.S.A.E., G.W., and N.D. performed research; K.B.W., M.S.A.E., G.W., N.D., M.C., and D.Y. analyzed data; and K.B.W. and D.Y. wrote the paper.

## CHAPTER 5. REVIEW OF THE MOLECULAR RECOGNITION OF THE HYBRID-TYPE G-QUADRUPLEXES IN HUMAN TELOMERES

Reprinted with permission from

Guanhui Wu, Luying Chen, Wenting Liu and Danzhou Yang, Molecular Recognition of the Hybrid-Type G-Quadruplexes in Human Telomeres. *Molecules* 24 (2019). Copyright © 2019 MDPI

### 5.1 Introduction

Structural data of small-molecule complexes of the human telomeric G-quadruplexes can provide important information for understanding specific recognition by small molecules, and for structure-based rational drug design targeting human telomeric G-quadruplexes. Only a handful solution structures are available of ligand complexes of human telomeric G-quadruplexes<sup>208-211</sup>. This part will focus on the two recent solution structural studies. We use NMR spectroscopy to elucidate the solution structures of a 1:1 complex between a medicinal natural product epiberberine (**Figure 5.1A**) and the hybrid-2 telomeric G-quadruplex<sup>211</sup>, and 1:1 and 4:2 complexes between a Pt-containing small molecule (**Figure 5.1B**) and the hybrid-1 telomeric G-quadruplex<sup>210</sup>.

### 5.2 Structural conformation of telomeric G-quadruplexes

Previous studies showed the human telomeric G-overhang predominantly forms two hybrid-type G-quadruplex structures (**Figure 5.2**) in physiologically relevant K<sup>+</sup>-containing solution, named hybrid-1<sup>212-214</sup> and hybrid-2<sup>85, 215</sup>. These two structures coexist in K<sup>+</sup> solution in an equilibrium<sup>216</sup>. They are unique in strand orientation, G-tetrad arrangements, loop arrangements, as well as 5'- and 3'-capping, compared to the parallel structures formed in the crystalline form<sup>49</sup> and those predominantly found in the oncogene promoters<sup>6</sup>. The 26-mer wtTel26 sequence adopts hybrid-2 conformation<sup>85</sup> (**Figure 5.2, left**). The first, third and fourth G-tracts are parallel, and the second G-tract is antiparallel. In addition, the first, second, and third G-tracts are connected with two TTA lateral loops, whereas the third and fourth G-tracts are linked with a TTA propeller loop, i.e. a lateral-lateral-propeller loop arrangement. The glycosidic torsion angles of guanine nucleotides in a G-tetrad adopt either *syn* or *anti* conformations. In the hybrid-2 structure,

the top G-tetrad has reversed glycosidic conformations (*syn:anti:syn:syn*) from those of the bottom two G-tetrads (*anti:syn:anti:anti*) (**Figure 5.2, left**). On the other hand, the 26-mer Tel26 sequence with modified 3'- and 5'-flanking segments adopts hybrid-1 conformation (**Figure 5.2, right**)<sup>213</sup>. The first, second and fourth G-tracts go in one direction and the third G-tract in the opposite direction, with a side-lateral-lateral loop arrangement. The top G-tetrad has reversed glycosidic conformations (*syn:syn:anti:syn*) relative to the other two G-tetrads (*anti:anti:syn:anti*) (**Figure 5.2, right**).

The two hybrid-type telomeric G-quadruplexes have unique capping structures, determined by the flanking and loop sequences together with the folding topology. In the hybrid-2 structure (**Figure 5.2, left**), the 5'-flanking residues, the second TTA lateral loop and the A21 residue of the third TTA reversal loop are above the top G-tetrad, but are not well structured. In contrast, a well-defined T:A:T triad is formed above the 3'-end of hybrid-2 telomeric G-quadruplex between T8 and A9 of the first TTA lateral loop, as well as the T25 of the 3'-end flanking segment. The 3'-capping structure is specific to the hybrid-2 structure and is not possible to form in the hybrid-1 structure. Additionally, this capping structure contributes to the stability for the hybrid-2 structure as demonstrated by mutational analysis<sup>213</sup>. In the hybrid-1 structure (**Figure 5.2, right**), both 5'-end and 3'-end capping structures are well-defined. The 5'-end capping structure, an adenine-triple, is formed by the A3 of the 5'-flanking residues, the A9 residue of the first TTA strand-reversal side loop, and the A21 residue of the third TTA lateral loop. The 3'-capping structure, an A:T base pair, is formed by the A25 of 3'-end flanking segment and the T14 of the second lateral loop form.

### 5.3 Molecular recognition of human telomeric hybrid-2 G-quadruplex by epiberberine

Protoberberines are a class of isoquinoline alkaloids with antitumor activities<sup>217, 218</sup>. Notably, berberine and some of its derivatives have been shown to stabilize telomeric G-quadruplexes and inhibit telomerase activity<sup>219</sup>. Epiberberine (EPI) was found to exhibit great fluorescence enhancement induced by human telomeric sequences in K<sup>+</sup> solution<sup>220</sup>. Our data show that EPI specifically binds the hybrid-2 human telomeric G-quadruplex structure, and we have determined the solution structure of the 1:1 complex of EPI and hybrid-2 human telomeric G-quadruplex by NMR<sup>211</sup>. The NMR solution structure of the complex shows that the EPI molecule (**Figure 5.1A**) forms a well-defined 1:1 complex with the hybrid-2 human telomeric G-quadruplex. This complex is stabilized by extensive hydrogen-bonding, base-stacking and electrostatic

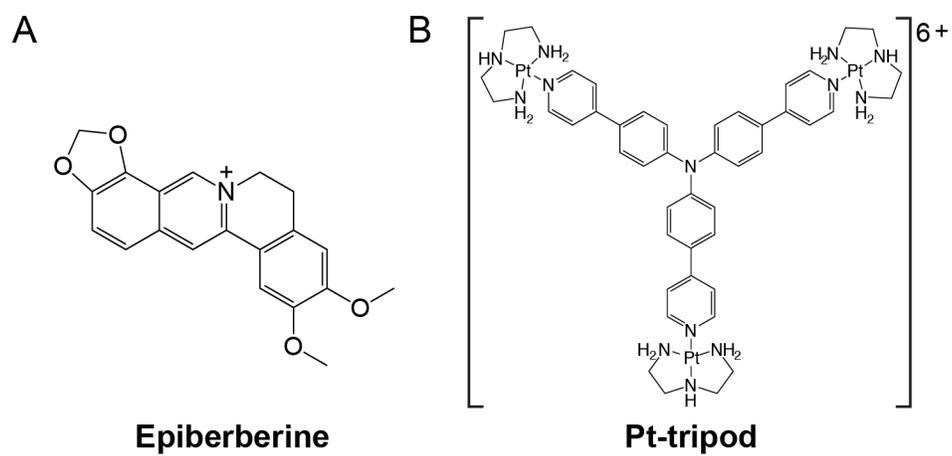


Figure 5.1. The chemical structures of epiberberine and Pt-tripod

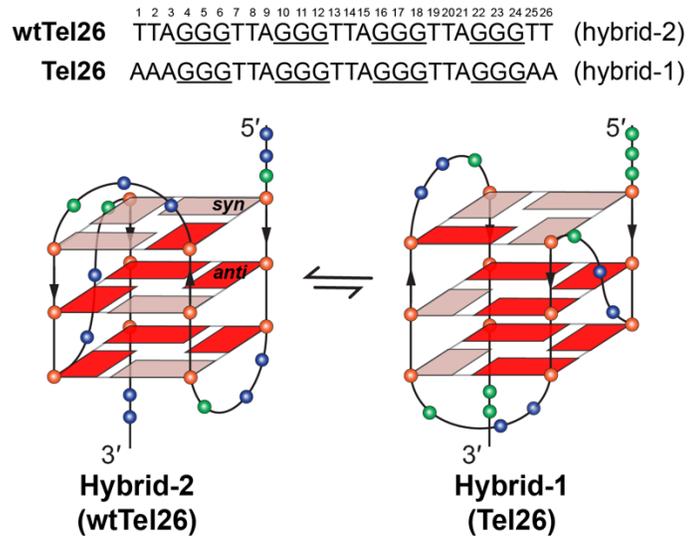


Figure 5.2. The folding pattern of hybrid-2 and hybrid-1 telomeric G-quadruplexes

Human telomeric sequences predominantly fold into two hybrid-type G-quadruplexes with an equilibrium between hybrid-1 (PDB ID: 2HY9) and hybrid-2 (PDB ID: 2JPZ) forms. The human telomeric sequences used to determine hybrid-2 (wtTel26) and hybrid-1 (Tel26) structures are shown on the top. Strand polarities are indicated as black arrows. Glycosidic conformations of guanine nucleotides are marked as follow: *syn*-pink; and *anti*-red. Different nucleotides are represented as follow: thymine-blue sphere; guanine-red sphere; adenine-green sphere.

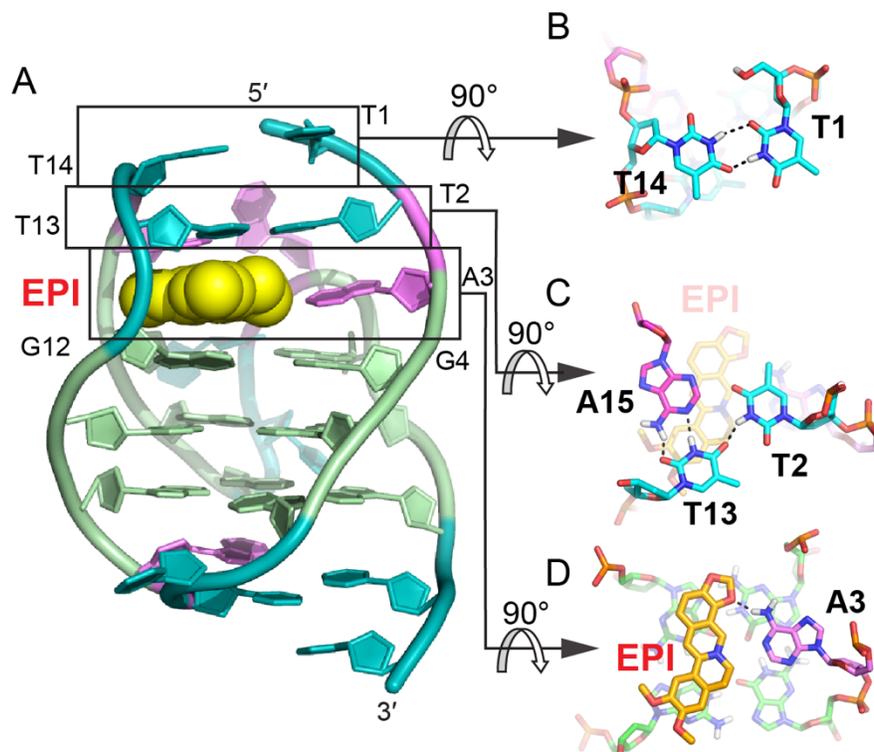


Figure 5.3. The specific recognition of human hybrid-2 telomeric G-quadruplex by EPI (PDB ID: 6CCW)

(A) Cartoon representation of 1:1 EPI/hybrid-2 complex. Different nucleotides are marked as follow: thymine-cyan, adenine-pink, guanine-green. The EPI molecule is shown as the yellow spheres. (B-D) Top view of the EPI:A3 quasi-triad plane (B), T2:T13:A15 triad plane (C), and T1:T14 pair (D). Potential hydrogen bonds are shown in black dashed lines.

interactions (**Figure 5.3A**). In the free telomeric hybrid-2 structure, the flanking segment and the T13-T14-A15 lateral loop at the 5' end are disordered. Upon binding, EPI stacks on the 5'-external G-tetrad, and extensive rearrangement occurs in the 5'-end flanking residues and lateral loop to form a multi-layer drug-binding pocket. The A3 base from the 5'-flanking strand is recruited by EPI to form a 'quasi-triad plane' which is intercalated between the 5'-G-tetrad and two additional capping layers (**Figure 5.3A**). Specifically, EPI stacks over the tetrad G12 and G16, with its N7 centered over the 5'-tetrad (**Figure 5.3D**), and A3 covers the center of the G4-G22 edge.

Importantly, a hydrogen bond is formed between A3/NH6 and the oxygen of the methylenedioxy ring of the EPI, which stabilizes the EPI-A3 plane. The positively charged N7 atom (**Figure 5.3D**) of the EPI is in proximity to the O6 atoms of the 5' external tetrad guanines, likely stabilizing the complex analogous to the K<sup>+</sup> cations within the central channel of the G-core. Above the EPI:A3 plane, a T:T:A triad is formed from T2 of the 5' flanking strand and T13 and A15 from the lateral loop. The triad is stabilized by a hydrogen-bonding network, with two hydrogen-bond interactions between T13 and A15 to form a reversed Watson-Crick base pair (**Figure 5.3C**) and one hydrogen bond between T2 and T13. In fact, the stable formation of the T:T:A triad results in the appearance of a signature NMR imino proton peak of T13 arising from the unique triad conformation of the specific EPI-binding pocket within the hybrid-2 telomeric G-quadruplex. Finally, this triad is capped by a hydrogen-bonded T1:T14 base pair which stacks over the triad and further stabilizes the overall complex (**Figure 5.3B**).

The optimal recognition of EPI is specific to the hybrid-2 folding topology and the human telomeric DNA loop sequence TTA. In the EPI-hybrid-2 complex, the second TTA loop adopts a lateral loop conformation above the 5'-tetrad, which provides ideal orientation for the pairing interaction with the 5'-flanking segment to form the highly stable capping structures of T:T:A triad and T:T pair (**Figure 5.4A** and **5.4B**). In the hybrid-1 structure, the third TTA loop at the 5'-end (**Figure 5.4C**) is offset by 90° relative to the second lateral loop at the 5'-end of the hybrid-2 structure, and thus cannot form the optimal stable capping structures for the EPI binding pocket. In the basket-type telomeric G-quadruplex structure, the diagonal loop and the 3'-flanking segments are both at the same end of the 5'-flanking (**Figure 5.4D**), sterically hindering the EPI binding. Human telomeric G-quadruplexes bear inherent structure polymorphism, predominantly

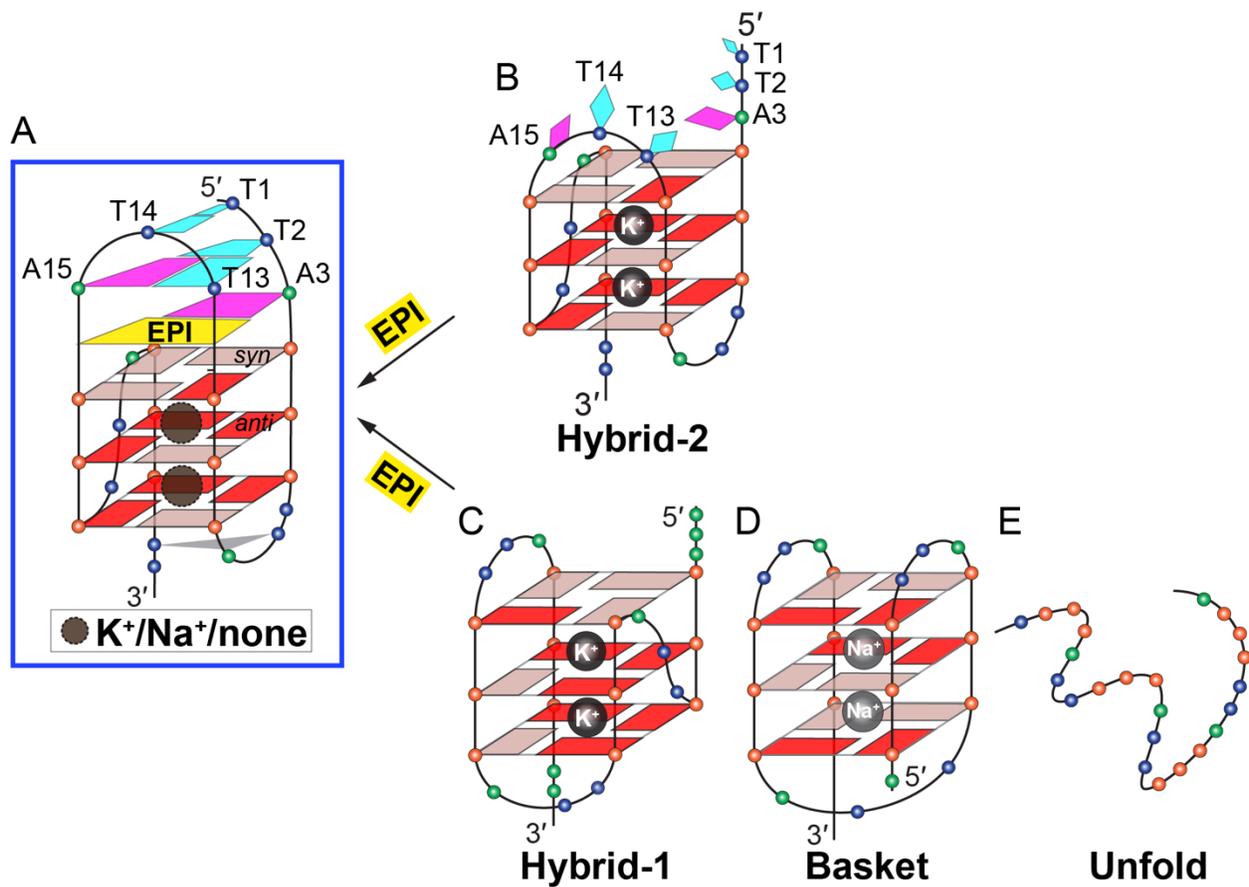


Figure 5.4. Conversion of human telomeric G-quadruplex structures to the hybrid-2 form induced by EPI

(A-B) EPI binding induces extensive rearrangement of previously disordered 5'-flanking and lateral loop segments (B) to form a well-defined four-layer binding pocket (A) specific to hybrid-2 telomeric G-quadruplex. (C-E) Other human telomeric G-quadruplex forms, including hybrid-1 (C) and basket (D), or the free telomeric sequence in the absence of salt (E) can be converted to the hybrid-2 form as the addition of EPI molecules. The specific recognition of human hybrid-2 telomeric G-quadruplex by EPI (PDB ID: 6CCW).

with two hybrid-type G-quadruplexes in equilibrium between hybrid-2 and hybrid-1 structures under physiologically relevant solution conditions. Remarkably, EPI is able to convert hybrid-1, basket and unfolded telomeric G-quadruplex structures to the hybrid-2 form independent of available solution cations (**Figure 5.4**); this is the only such compound reported to date. The highly specific and selective binding of EPI to the hybrid-2 structure likely shifts the equilibrium between different forms in solution, resulting in the overall conversion of other telomeric G-quadruplex structures to the hybrid-2 form. The EPI-hybrid-2 complex structure reveals several features which enable specific recognition of the hybrid-2 human telomeric G-quadruplex. EPI contains a crescent-shaped asymmetric stacking moiety that can only stack over two tetrad guanines, allowing the recruitment of a flanking base partner to co-stack over the 5'-G-tetrad and lock the EPI position. The pairing with the recruited adenine and the central location of positively charged N7 together anchor the position and orientation of EPI above the 5' external G-tetrad. In addition, the appropriately positioned hydrogen-bond acceptors in the methylenedioxy ring E of EPI enable it to optimally hydrogen-bond with the recruited adenine (**Figure 5.3A and 5.3D**). This is highlighted by the observation that the structurally similar berberine alkaloids berberine, palmatine, coptisine, which only differ in the positions of the methylenedioxy ring and methoxy groups, are unable to form a well-defined complex with the telomeric G-quadruplex. This molecular-level recognition information can only be obtained from detailed structural study and is important for rational drug design of improved analogs targeting the hybrid-2 human telomeric G-quadruplex.

#### **5.4 Molecular recognition of human telomeric hybrid-1 G-quadruplex by Pt-tripod in monomeric and dimeric complexes**

Recently, a series of platinum (II) compounds was found to bind telomeric G-quadruplex and suppress telomerase activity<sup>221, 222</sup>. Among them, a Pt-based tripod (Pt-tripod) showed strong *in vitro* and *in vivo* anticancer activity upon light irradiation. This tripod is a non-planar compound with a central tertiary amine connecting three arms each bearing two aromatic rings and a cationic platinum complex with a three-fold symmetry (**Figure 5.1B**). Interestingly, Pt-tripod binds to the intramolecular hybrid-1 human telomeric G-quadruplex formed by the Tel26 sequence and forms well-defined monomeric and dimeric Pt-tripod-Tel26 complexes dependent on the drug-DNA ratio<sup>210</sup>.

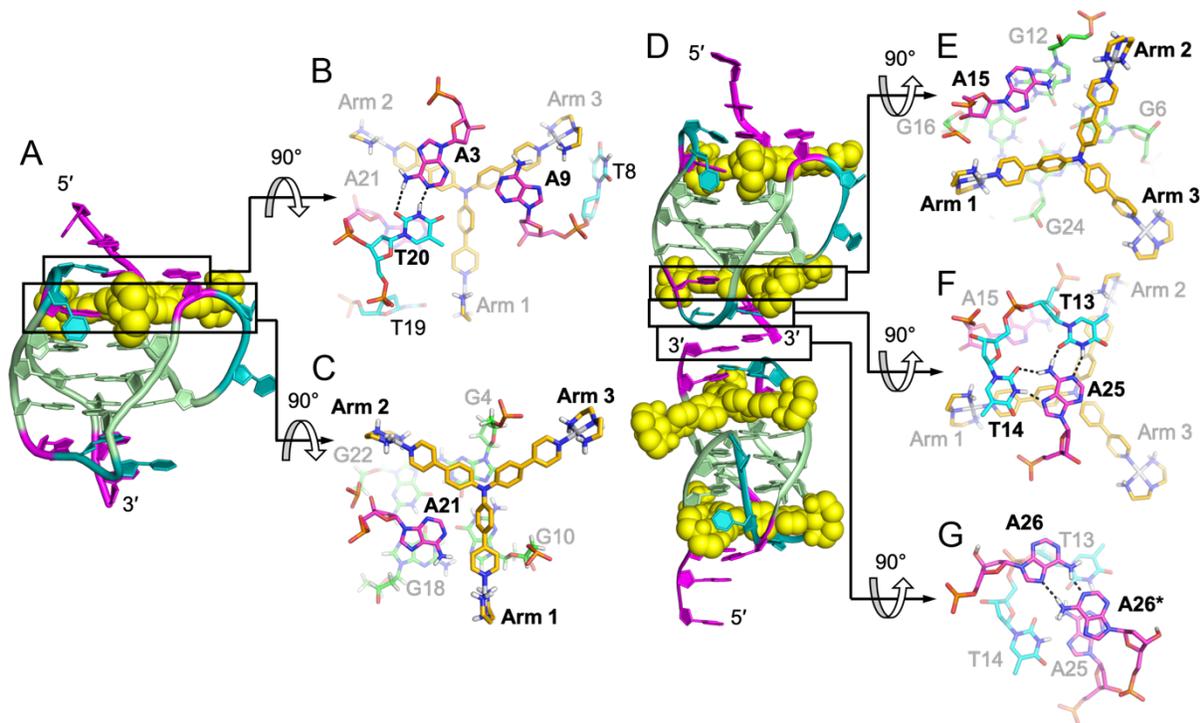


Figure 5.5. Solution structure of 1:1 and 4:2 Pt-tripod/hybrid-1 G-quadruplex complexes (PDB ID: 5Z80 and 5Z8F)

(A) Cartoon representation of 1:1 Pt-tripod/hybrid-1 complex. Different nucleotides are marked as follow: thymine-cyan, adenine-pink, guanine-green. The Pt-tripod molecule is shown as the yellow spheres. (B-C) Top views of the A3:A9:T20 triad plane (B) and Pt-tripod:A21 plane at 5' site (C). (D) Cartoon representation of 4:2 Pt-tripod/hybrid-1 complex. (E-G) Structural details of the 3' site binding pocket. Bottom views of the Pt-tripod:A15 plane (E), T13:T14:A25 triad plane (F), and the A26:A26\* (A26 from each hybrid-1) base pair (G). Potential hydrogen bonds are shown in black dashed lines.

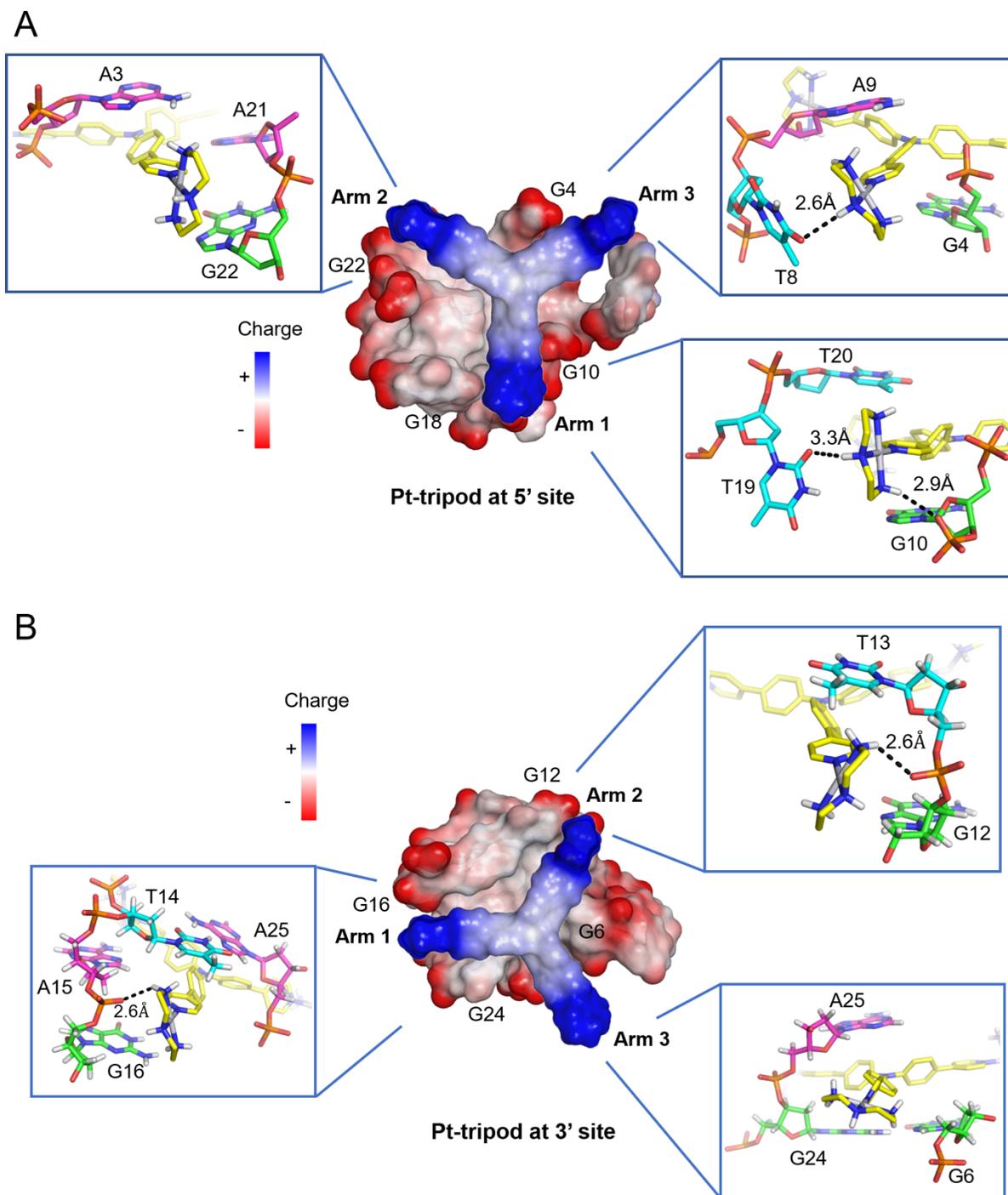


Figure 5.6. Arm-groove interactions of Pt-tripod at 5' and 3' sites in the 4:2 Pt-tripod/hybrid-1 complex

(A) Top view of Pt-tripod at the 5' site of hybrid-1 without capping structure (middle). The binding pocket surface is color coded according to the charge. The interaction details between each arm and groove are shown in enlarged view (corners). (B) The same information is shown for Pt-tripod at the 3' site. Potential hydrogen bonds are shown in black dashed lines.

At the 1:1 ratio, Pt-tripod binds to the 5'-end of the hybrid-1 Tel26 to form a well-defined 1:1 Pt-tripod–Tel26 complex (**Figure 5.5A**). In the 1:1 complex, the ligand binding induces a large conformational rearrangement at the 5' end. Pt-tripod intercalates between A3 of the 5'-flanking segment and G4 of the 5'-tetrad, and recruits A21 from the third lateral loop to form a Pt-tripod-A21 paired plane above the 5'-external tetrad (**Figure 5.5C**), replacing the adenine triad formed by A3:A9:A21 in the free Tel26 G-quadruplex. The Pt-tripod covers G10 and G4 by its arm 1 and arm 2, respectively, with the central tertiary amine nitrogen right above the G4 and G10 edge, while A21 stacks over the G18 and G22. Above the Pt-tripod-A21 plane, A3 and T20 form a hydrogen-bonded reversed Watson-Crick base pair, while A9 from the first side loop also positions in the same plane to cover the aromatic moiety of arm 3 (**Figure 5.5B**).

At the 3:1 Pt-tripod–DNA ratio, Pt-tripod binds the hybrid-1 Tel26 to predominately form a well-defined 4:2 Pt-tripod–Tel26 3'–3' dimeric complex (**Figure 5.5D**) with a twofold symmetry. In the 4:2 Pt-tripod–Tel26 complex, Pt-tripod binds at both the 5'- and 3'-ends of Tel26 G-quadruplex. In addition to the 5'-end complex, the second ligand induces a large conformational rearrangement at the 3'-end to form a well-defined binding pocket, in which Pt-tripod intercalates between A25 and G24 of the 3'-tetrad and recruits A15 from the second lateral loop to form a Pt-tripod-A21 paired plane above the 3'-external tetrad (**Figure 5.5E**). Here, arm 1 and arm 2 of the Pt-tripod cover G24 and G6, respectively, with the central tertiary amine nitrogen above the G6 and G24 edge, while A15 stacks over G12 and G16. While the 4:2 dimeric complex is induced by Pt-tripod binding, the binding of Pt-tripod at the 3'-end also appears to be facilitated by the 3'-end interlocking interface in the dimeric complex, including a well-defined hydrogen-bonded T:A:T triad (**Figure 5.5F**) and a A:A base pair (**Figure 5.5G**). This stable three-layer binding pocket at the 3' dimeric interface gives rise to a less dynamic binding of Pt-tripod as compared to the 5'-complex. The T14:A25 base pair observed at the 3'-end in the free Tel26 G-quadruplex is completely rearranged, with the glycosidic conformation of T14 changed from *anti* to *syn* and A25 from *syn* to *anti*. It is noted that this unique 3' dimeric interface appears to be specific to the modified 3'-flanking sequence of Tel26.

A unique feature of the two Pt-tripod–Tel26 complexes is the anchoring of the three positively charged arms of Pt-tripod in three G-quadruplex grooves, which defines and locks the position of Pt-tripod in the complexes (**Figure 5.6**). Interestingly, the anchoring of the Pt-tripod arms appears to avoid the lateral loops, which may cause potential steric hindrance. Each arm of

Pt-tripod stretches into a groove without a lateral loop and thereby positions its outer platinum complexes for interactions with the loop residues and DNA backbone. These include electrostatic interactions of the positively charged platinum moiety of the three arms with the negatively charged phosphate backbone, and hydrogen-bonding interactions and  $\pi$ - $\pi$  interactions of the three arms with loop and flanking residues. As the tertiary amine nitrogen of Pt-tripod is located above the edge of the external G-tetrad in both complexes, the aromatic moieties of arm 1 and arm 2 of both ligands stack over a guanine of the external G-tetrad, whereas the arm 3 almost completely extends into the groove without many stacking interactions with the external G-tetrads. In the 5'-end complex (**Figure 5.6A**), however, the arm 3 of Pt-tripod extensively interacts with the propeller side loop, with A9 stacking over its first aromatic ring and T8 hydrogen-bonded with its terminal moiety. While A21 of the third lateral loop at the 5'-end is recruited to pair with Pt-tripod, T19 of the same loop interacts with arm 1 through hydrogen bonding with T20 further stacking over A21. Arm 2 is intercalated between A3 of the flanking segment and G4 of the 5'-tetrad, with its platinum-moiety interacting with the phosphate backbone in the groove. In the 3'-end complex (**Figure 5.6B**), arm 1 and arm 2 of Pt-tripod show electrostatic and potential hydrogen bonding interactions with the phosphate backbone.

## 5.5 Insights obtained from the complex structures

EPI and Pt-tripod both induce extensive conformational rearrangements of flanking and loop residues after binding to the human telomeric G-quadruplexes. The EPI-hybrid-2 G-quadruplex and 1:1 and 4:2 Pt-tripod-hybrid-1 G-quadruplex complexes all include ligand-induced well-defined multi-layer binding pockets involving the external tetrad, flanking and loop residues. Such rearrangements are more extensive than most G-quadruplex-interactive-ligand complexes, particularly those observed with the parallel-stranded G-quadruplexes that lack the lateral loops. Therefore, the importance of interactions with flanking segments and various loop types in the hybrid-type human telomeric G-quadruplexes for binding specificity is highlighted. On the other hand, both EPI and Pt-tripod recruit a DNA base to form a ligand-base plane covering the external G-tetrad. Both the EPI and the Pt-tripod scaffolds cover only two guanine bases, and by recruiting an adenine residue, a full tetrad coverage is achieved. Such a phenomenon has also been observed for ligand complexes of the parallel-stranded *MYC* promoter G-quadruplex<sup>150, 154</sup>. The engagement of the target-DNA residue anchors the specific orientation of the ligand and might be the basis for

the specific recognition of a ligand. In contrast to quadruplex ligands with large aromatic systems covering most of the tetrad, the asymmetric crescent shaped EPI or the tri-armed Pt-tripod facilitate this target-base recruitment.

Comparing how EPI and Pt-tripod bind in the 5'-complexes of the two hybrid-type human telomeric quadruplexes reveals similarities and critical differences in stacking, hydrogen bond, and electrostatic interactions (**Figure 5.7**). Similarly, both drugs stack upon two guanine bases of the 5'-tetrad and recruit an adenine base to form a ligand-base plane covering the 5'-tetrad. However, their binding positions and interactions are different. The more-compact and more-planar EPI intercalates between the second lateral loop and the 5'-tetrad and recruits A3 from the 5'-flanking segment in the hybrid-2 structure. EPI forms a clear hydrogen bond with A3 in the ligand-base plane, while the T13 and A15 of the second lateral loop provide an optimal capping covering the long axis of EPI. In contrast, the tri-armed Pt-tripod intercalates within the 5'-flanking segment between A3 and the 5'-tetrad, and recruits A21 from the lateral loop in the hybrid-1 structure using two arms in a more shape-complementary manner. Thereby, the EPI molecule is opposite to the 5'-flanking strand anchoring towards the A3, while the Pt-tripod is opposite to and restricted by the lateral loop above the 5' external tetrad. Both ligand-base planes are covered by a triad of flanking and loop residues, each of which including a reverse Watson-Crick AT base pair. Stacking interactions are more predominant for the planar EPI in its interactions with the 5'-tetrad and the capping triad than for the non-planar Pt-tripod (**Figures 5.7**). In the EPI complex, the capping triad is more stable and well-defined as the three bases are all connected through hydrogen-bonds, and is covered by an additional layer of T:T base pair. This extensive binding pocket defines the stable complex formation of EPI and hybrid-2 structure, which promotes the EPI-induced conversion of other human telomeric quadruplex topologies towards the hybrid-2 folding. In contrast, in the capping triad of the Pt-tripod 5'-complex, the A3:T20 base pair is distant from the A9 residue, which likely stabilizes the arm 3 by stacking with its aromatic moiety (**Figure 5.7B**).

The positive charge of the two drugs contributes to their respective binding interactions in different ways. The positively charged nitrogen of EPI is located above the central pore close to the negatively polarized carbonyl functional groups of the 5'-tetrad, analogous to the centrally coordinated cations. For Pt-tripod, the central moiety is uncharged and also moved to the tetrad

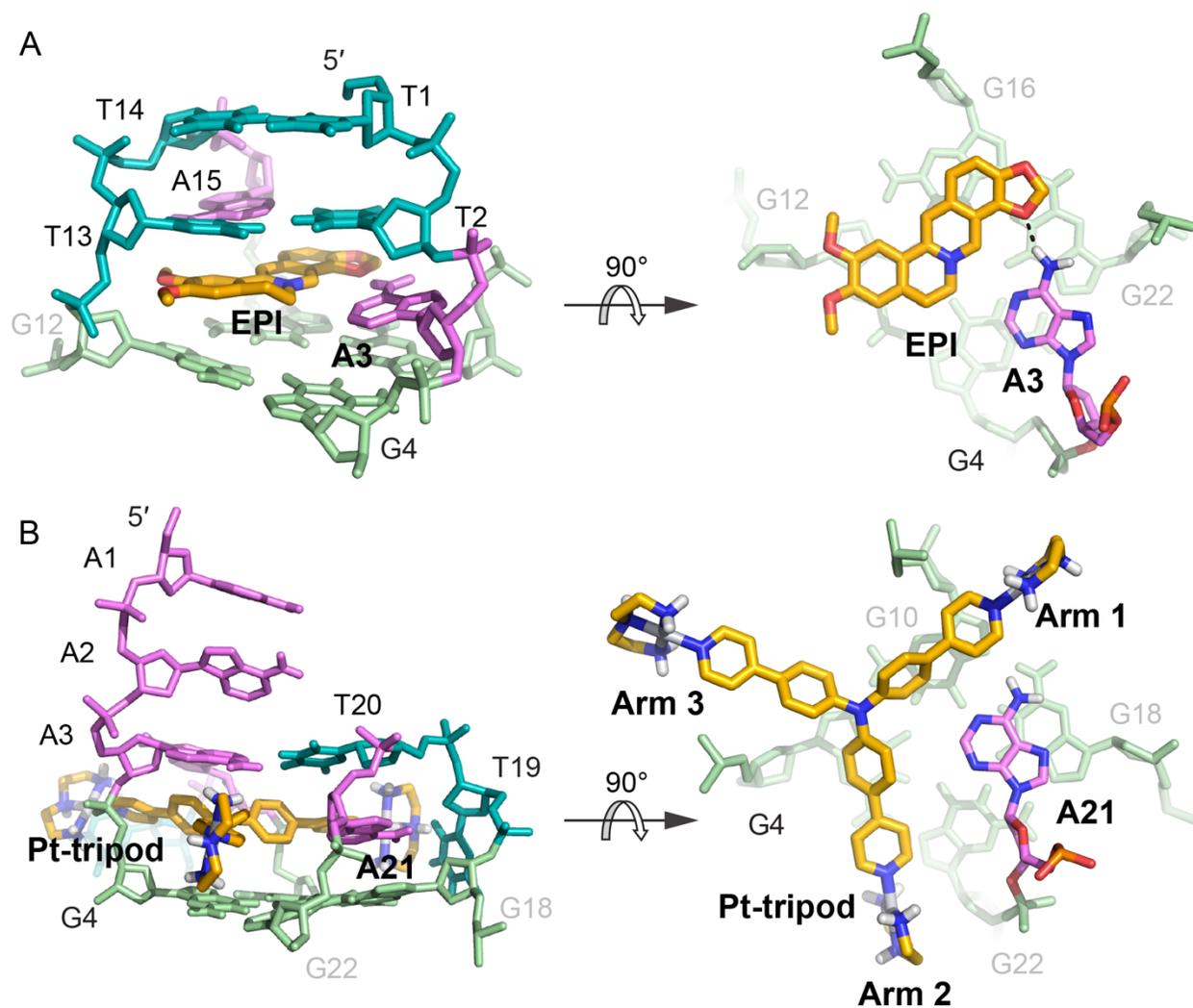


Figure 5.7. Comparison of the 5'-binding modes between EPI and Pt-tripod with human telomeric hybrid-type G-quadruplexes

- (A) Upon binding, the 5'-flanking A3 is recruited by EPI and forms a quasi-triad plane (right). The multi-layer drug-binding pocket for EPI is formed by the 5'-end flanking residues and TTA lateral loop. (left). Potential hydrogen bond is shown as the black dashed line. (B) Upon binding, A21 in the second lateral loop is recruited by Pt-tripod and forms a quasi-triad plane (right), stacking on top of the 5'-external G-tetrad and locking the position of Pt-tripod. Each arm of Pt-tripod stretches into different grooves. The Pt-tripod:A21 plane is capped by the A3:A9:T20 triad (left).

edge. Instead, its long arms position the three terminal positively charged Pt(II) complexes for interactions with negatively charged phosphate backbone of three different G-quadruplex grooves. Although cationic drugs might have general binding to other nucleic acids structures, this groove-anchoring mode may be used to enable further G-quadruplex selectivity. For example, cationic side chain might be added in an EPI derivative to similarly anchor it to the grooves of the hybrid-2 quadruplex.

In summary, the reviewed high-resolution complex structures of the hybrid-type human telomeric G-quadruplexes with the EPI and the Pt-tripod small molecules advance our knowledge about quadruplex-ligand interactions. Both drugs induce a previously unprecedented rearrangement of the capping residues to form extensive multi-layer binding pockets, with each drug recruiting a G-quadruplex DNA residue to form a ligand-base plane and define its respective binding position. Additional features observed for each drug are the EPI-induced conversion of alternative human telomeric quadruplex topologies towards the hybrid-2 type and the Pt-tripod-induced 3'-3' dimerization to form a 4:2 complex. The identified mechanisms of molecular recognition will provide insights into designing improved cancer therapeutics targeting the human telomeric G-quadruplexes.

## 5.6 Footnotes

This chapter has been previously published as:

Wu, G., Chen, L., Liu, W., & Yang, D., Molecular Recognition of the Hybrid-Type G-Quadruplexes in Human Telomeres. *Molecules*. 2019; 24(8):1578. (Copyright © 2019 MDPI)

Author contributions: G.W., L. C., W. L., and D. Y. wrote the paper.

## CHAPTER 6. A DNA POLYMERASE STOP ASSAY FOR CHARACTERIZATION OF G-QUADRUPLEX FORMATION AND IDENTIFICATION OF G-QUADRUPLEX-INTERACTIVE COMPOUNDS

Reprinted with permission from

Guanhui Wu. & Haiyong Han, A DNA polymerase stop assay for characterization of G-quadruplex formation and identification of G-quadruplex-interactive compounds. *Methods Mol Biol* 2035, 223-231 (2019). Copyright © 2019, Springer Science Business Media, LLC, part of Springer Nature

### 6.1 Introduction

The formation of DNA G-quadruplexes (G4s) can influence critical nuclear processes, such as transcription<sup>10, 11, 20</sup>, DNA replication<sup>9</sup>, and DNA repair<sup>12, 164, 223</sup>. The G4 structures arise from DNA sequences containing consecutive runs of guanine, which are predominantly restricted in telemetric regions and regulatory regions of cancer-related genes in genomes<sup>25</sup>. Stabilization of G4 structures using G4-interactive ligands can disrupt the interaction between single-stranded or double-stranded DNA binding proteins and their target sequences, thereby inhibiting gene transcription and impairing cellular homeostasis<sup>80, 152</sup>. G-quadruplexes can also induce replication fork stalls during DNA replication and eventually result in DNA double-strand breaks<sup>223, 224</sup>. Since they are more prevalent in cancer cells than in normal cells<sup>20</sup>, G4 structures have gained significant attraction as new targets for cancer drug discovery.

Biophysical techniques, such as Förster resonance energy transfer (FRET) and circular dichroism (CD) spectroscopy have been widely used to characterize G4 formation in particular short G-rich oligonucleotides, but their use is limited in the determination of G4-forming region in a long DNA sequence especially when a sequence consisting of more than four runs of contiguous guanines. In this chapter, we describe a biochemical method using DNA polymerases for investigating the G4 formation in a relatively long (70-80 nt) DNA strand<sup>225</sup>. DNA polymerases are enzymes that synthesize DNA strands during DNA replication. They move along the DNA template strand in the 3'-5' direction and create a new strand from a 5' to 3' direction. In the polymerase stop assay, the template DNA containing G4 forming region is annealed with

radiolabeled primers. The primer-annealed DNA templates are then utilized by *Taq* DNA polymerase for primer extension. The presence of G4 structures in the template strand can cause DNA polymerase to stall immediately before these structures (**Figure 6.1A**). Resolving these paused products on a denaturing PAGE gel can provide a simple and rapid way to identify DNA secondary structures in the template strand. This assay has also been widely used to test/identify potential G-quadruplex-interactive small molecules<sup>225-227</sup>.

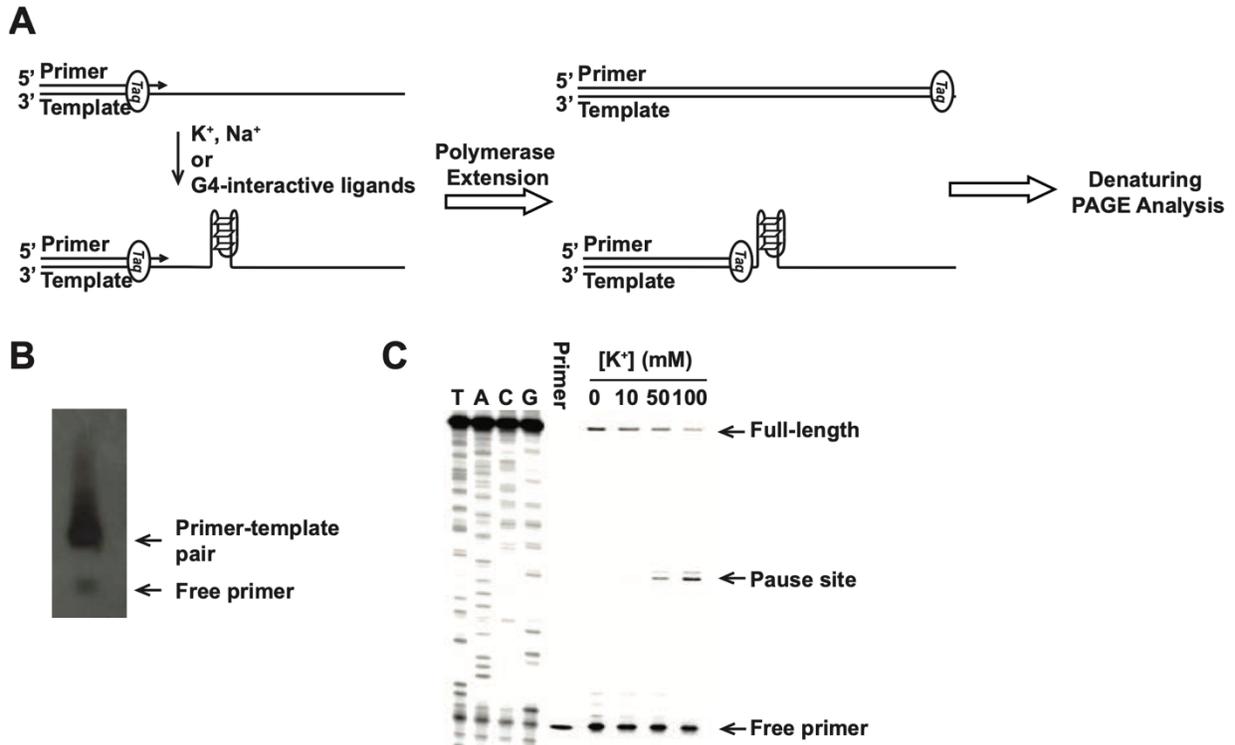


Figure 6.1. Overview of DNA polymerase stop assay

(A) Schematic diagram showing the strategy for characterizing G-quadruplexes formed in a G-rich sequence using the polymerase stop assay. (B) Autoradiogram showing the separation of the primer-template pair from the excess labeled primer on an 8% native PAGE. (C) DNA polymerase stop assay showing increasing concentrations of  $K^+$  induce the formation of G-quadruplex structures within the template strand and result in DNA polymerase to stall immediately before the quadruplex (Pause site). Lanes T, A, G, and C represent the sequencing reactions using the same DNA template to indicate the exact site where the DAN polymerase reaction pauses.

## 6.2 Materials

All solutions are prepared with ultrapure deionized water and stored at room temperature unless indicated otherwise.

### I. Radioactive 5'-end labeling with T4 polynucleotide kinase

- i. Oligonucleotide primer: stock solution 20  $\mu\text{M}$  (*see Note 1*)
- ii. T4 polynucleotide kinase (PNK) (10 units/ $\mu\text{L}$ , NEB, Ipswich, MA, USA).
- iii. T4 PNK kinase reaction buffer (10X): 70 mM Tris-HCl (pH 7.6), 100 mM  $\text{MgCl}_2$ , 50 mM DTT (NEB).
- iv. Adenosine 5'-gamma  $^{32}\text{P}$  triphosphate (10 mCi/mL, BLU502A, PerkinElmer, Waltham, MA, USA).
- v. Micro Bio-Spin 30 Columns (Bio-rad, Hercules, CA, USA).

### II. Template DNA

An oligonucleotide serves as the DNA template for the primer extension: 20  $\mu\text{M}$  (*see Note 2*)

### III. Purification of primer-template pair by native polyacrylamide gel electrophoresis (PAGE)

- i. TBE (Tris/Borate/EDTA) electrophoresis buffer (10X): 0.89 M Tris-HCl (pH 8.0), 0.89 M boric acid, 20 mM EDTA. Filter through 0.45  $\mu\text{m}$  membrane and store at room temperature.
- ii. 40% acrylamide/bis-acrylamide, 29:1 (3.3% crosslinker, Bio-rad).
- iii. UltraPure™ N,N,N',N'-Tetramethylethylenediamine TEMED: Store at 4 °C (Invitrogen, Carlsbad, CA, USA).
- iv. Ammonium persulfate (APS): prepare 10% solution in water and store at 4 °C up to 1 month (Bio-rad).
- v. Gel loading buffer (10X): 50% glycerol by volume, 0.005% bromophenol blue (w/v). Store at -20 °C.
- vi. 0.3 M sodium acetate solution (pH 5.2)
- vii. Ethanol, molecular biology grade (Fisher Bioreagents, Waltham, MA, USA)

- viii. Costar spin-X centrifuge tube filter (0.45  $\mu\text{m}$ , Corning Inc, Corning, NY, USA)

#### **IV. DNA polymerase reaction**

- i. *Taq* DNA Polymerase (5 units/ $\mu\text{L}$ , NEB).
- ii. DNA polymerase buffer (10X): 500 mM Tris-HCl (pH 7.6), 10 mM  $\text{MgCl}_2$ , 50 mM DTT.
- iii. dNTP mix (10 mM each) (Thermo Scientific, Waltham, MA, USA).

#### **V. Sequencing reaction**

Thermo sequenase Cycle Sequencing Kit (Affymetrix, USB, Santa Clara, CA, USA).

#### **VI. Denaturing PAGE**

- i. TBE electrophoresis buffer (10X): 0.89 M Tris-HCl (pH 8.0), 0.89 M boric acid, 20 mM EDTA. Filter through 0.45  $\mu\text{m}$  membrane and store at room temperature.
- ii. 40% acrylamide/bis-acrylamide, 29:1 (3.3% crosslinker, Bio-rad).
- iii. UltraPure™ N,N,N',N'-Tetramethylethylenediamine TEMED (Invitrogen) stored at 4 °C.
- iv. Ammonium persulfate (APS): prepare 10% solution (w/v) in water and store at 4 °C up to 1 month (Bio-rad).
- v. Urea (Crystalline Power, Fisher BioReagents).
- vi. Alkaline gel loading dye (2X): 80% formamide by volume, 10 mM NaOH, 0.005% bromophenol blue (w/v). Store at -20 °C.

### **6.3 Methods**

#### **I. Radioactive 5'-end labeling with T4 polynucleotide kinase**

Carry out the following steps in a laboratory suited for working with radioactive materials. Wear appropriate personal protective equipment and use proper shields.

- i. Prepare a 40  $\mu\text{L}$  reaction mixture by adding 4  $\mu\text{L}$  of 20  $\mu\text{M}$  primer oligonucleotide (20  $\mu\text{M}$  stock solution), 12  $\mu\text{L}$   $\gamma\text{-}^{32}\text{P}$  ATP (10 mCi/mL), 6  $\mu\text{L}$  T4 polynucleotide kinase (PNK), 4  $\mu\text{L}$  10 $\times$  kinase buffer, and 14  $\mu\text{L}$  water.
- ii. Incubate the reaction mixture at 37  $^{\circ}\text{C}$  for 1 h.
- iii. Heat inactivate PNK by incubating at 65 $^{\circ}\text{C}$  for 20 minutes.
- iv. After the completion of the reaction, use Micro Bio-Spin 30 Columns to remove unincorporated radioactive  $\gamma\text{-}^{32}\text{P}$  ATP from labeled DNA by following the protocol provided by the column manufacturer. Briefly, centrifuge the column at 1,000  $\times$  g for 2 mins and remove packing buffer, and then place the column back in the same tube and centrifuge the column for an additional 1 min at 1,000  $\times$  g to remove residual packing buffer. Next, place the column in a clean 1.5 mL tube. Apply the entirety of the reaction mixture (40  $\mu\text{L}$ ) to the center of the gel bed and then centrifuge for 4 mins at 1,000  $\times$  g to collect the purified 5'-end-labeled oligonucleotide (*see Note 3*).

## II. Annealing $^{32}\text{P}$ -labeled primers to the template DNA

- i. Set up a 50  $\mu\text{L}$  annealing reaction by mixing an equimolar amount (15 pmol) of  $^{32}\text{P}$ -labeled primer DNAs and the template DNA together in  $\text{H}_2\text{O}$  (*see Note 4*).
- ii. Anneal  $^{32}\text{P}$ -labeled primers to the template strands by heating samples at 95  $^{\circ}\text{C}$  for 5 mins using a heating block and then allow it cool down slowly to room temperature. This may take 2-3 hrs (Cooled samples can be stored at 4  $^{\circ}\text{C}$  for several days).
- iii. Set up a native 8% polyacrylamide gel (10-well) of 20 cm  $\times$  16 cm  $\times$  0.8 mm using 60 mL of gel solution with the following procedure. Before preparing the gel solution, clean plates and spacer carefully in order to prevent air bubbles from forming in the gel and assemble the glass plates in gel caster. Prepare gel solution by mixing different solutions in the following order: 42 mL water, 6 mL TBE buffer (10X), 12 mL of 40% acrylamide/bisacrylamide (29:1), 600  $\mu\text{L}$  of 10% ammonium persulfate, and 60  $\mu\text{L}$  TEMED. After addition of TEMED, work quickly to pour the gel solution to the space between the glass plates and immediately insert an appropriate comb into the gel before the acrylamide

polymerizes. Allow the acrylamide to polymerize for 60 mins at room temperature.

- iv. Pre-run the gel at 150 V for 30 mins.
- v. Add 5  $\mu\text{L}$  of 10X gel loading buffer to the annealed reaction mixture, mix by gently tapping the side of the tube. Load 50  $\mu\text{L}$  per well of the mixture onto the pre-run native polyacrylamide gel, and run the gel at 150 V.
- vi. After the bromophenol blue dye has migrated about half way down the gel, stop the electrophoresis, carefully separate the two gel plates and visualize the location of DNA bands by autoradiography (*see Note 5*). **Figure 6.1B** shows an example of an autoradiogram obtained after exposure of X-ray film to a gel.
- vii. Cut out the DNA template/primer band from the gel with a razor blade and grind the gel pieces using a spatula in a 1.5 mL tube.
- viii. Elute DNA by incubating the gel pieces in 250-500  $\mu\text{L}$  of 0.3 M sodium acetate solution on a tube rotator at room temperature overnight.
- ix. Next day, transfer the gel solution to a Costar Spin-X Centrifuge tube filter (0.45  $\mu\text{m}$ ), centrifuge at 3,200 rpm for 10 mins and collect the eluted solution.
- x. Add 3.5-4 volumes of ethanol to the eluent and incubate at  $-20\text{ }^{\circ}\text{C}$  overnight or at  $-80\text{ }^{\circ}\text{C}$  for at least an hour. Precipitate DNA by centrifuging the tubes at  $17,900 \times g$  for 10 mins and gently remove supernatant using a 1 mL transfer pipette (*see Note 6*).
- xi. Wash the DNA pellet once with 100  $\mu\text{L}$  of 70% ice-cold ethanol, and centrifuge sample at 13,000 rpm for 5 mins. After removing the supernatant using a 1 mL transfer pipette, allow the samples to air-dry for 5-10 mins.
- xii. Quantitate radioactivity of the samples using a Geiger counter or a scintillating counter and then re-suspend the samples in a proper amount of  $\text{H}_2\text{O}$  to get a concentration about 10,000 cpm/ $\mu\text{L}$  (*see Note 7*).

### III. Sequencing reactions

The template DNA is sequenced and resolved in a gel alongside the DNA polymerase stop assay products to indicate the stoppage site of the primer extension reaction (the 3' end of the G4 structure). Sequencing reactions are carried out using the Thermo

Sequenase Cycle Sequencing Kit. The following procedure is modified from the instructions provided by the kit manufacturer (Affymetrix).

- i. Label 4 PCR tubes with A, T, G, and C.
- ii. Add 8  $\mu\text{L}$  of the ddGTP termination mix into the tube labeled with G. Similarly fill the A, T, and C tubes with 8  $\mu\text{L}$  of the corresponding ddATP, ddTTP, and ddCTP termination mixes, respectively. Cap tightly the tubes to prevent evaporation.
- iii. In a 0.5 mL microcentrifuge tube, prepare a 35  $\mu\text{L}$  reaction mixture containing 0.2  $\mu\text{g}$  or 0.1 pmol template DNA, 4  $\mu\text{L}$  reaction buffer, 1 pmol labeled primer (from Section 3.1), 4  $\mu\text{L}$  Thermo Sequenase DNA Polymerase, and  $\text{H}_2\text{O}$ .
- iv. Transfer 8  $\mu\text{L}$  of the reaction mixture to the tube labeled G, mix gently. Similarly transfer 8  $\mu\text{L}$  of the mixture to the each of A, T, and C tubes. Overlay each vial with 5-10  $\mu\text{L}$  mineral oil and place them in a thermal cycler.
- v. Start the cycling program. The following cycle parameters are recommended by the kit manufacturer: 95  $^\circ\text{C}$  for 30 secs, 55  $^\circ\text{C}$  for 30 sec, 72  $^\circ\text{C}$  for 60-120 secs (40-60 cycles) (*see Note 8*).
- vi. When the cycling reaction completes, add 8  $\mu\text{L}$  alkaline gel loading dye to each tube to terminate reaction and store at 4  $^\circ\text{C}$ .

#### IV. DNA polymerase reaction

- i. Prepare 20  $\mu\text{L}$  reaction mixtures containing 2  $\mu\text{L}$  DNA primer/template mix from Section 3.2 (10,000-20,000 cpm), 2  $\mu\text{L}$  dNTP mix (10 mM), 2  $\mu\text{L}$  cation solution (LiCl, KCl or NaCl in a desired concentration), 2  $\mu\text{L}$  tested G4-interactive small molecule in a desired concentration, 2  $\mu\text{L}$  of 10X DNA polymerase buffer and 10  $\mu\text{L}$  water, and then incubate the reaction mixtures at a room temperature for 1-3 hr to allow G4 formation (*see Note 9*).
- ii. After incubation, add 0.5 -1  $\mu\text{L}$  *Taq* DNA polymerase to each reaction mixture and mix by gently tapping the tubes.
- iii. Spin down samples for a few seconds and place them in a thermal cycler at a temperature calculated based on the  $T_m$  of the primer for 30 mins (*see Note 10*).

- iv. When polymerase extension reaction completes, immediately add 10  $\mu$ L alkaline gel loading dye to each tube.
- v. Concentrate the DNA sequencing samples and polymerase extension samples by evaporation using a SpeedVac concentrator (reduce the volume to 10  $\mu$ L or less which may take 4-5 hours).
- vi. Prepare a 16% 0.4 mm thick sequencing gel using 60 mL of gel solution (6 mL TBE buffer (10X), 8 M urea (~30 g), 24 mL of 40% acrylamide/bisacrylamide (29:1), and water) (*see Note 11*).
- vii. Pre-run and warm the gel at 2,000 V with a maximal current of 40 mA for about 30 mins.
- viii. Heat the sequencing samples from Section 3.3 and the polymerase extension samples at 95 °C for 3 mins and cool down by placing the tubes on ice.
- ix. Quantitate the radioactivity of each sample using a Geiger counter or a scintillation counter.
- x. Load samples (4,000-5,000 cpm per sample to each well) onto the denaturing PAGE gel.
- xi. Run the gel at 2,000 V with a maximal current of 40 mA until the dye front runs 1/2 to 2/3 of the way down the gel (*see Note 12*).
- xii. After the gel electrophoresis, detach the gel plates from the electrophoresis apparatus, and carefully separate the two plates, lift one plate and leave the gel attached to the other plate.
- xiii. Carefully place a piece of a thin chromatography paper on top of the gel, and gently pull back the paper from a corner to transfer the gel to the paper.
- xiv. Cover the wet gel with plastic wrap on top.
- xv. Place the gel sandwich on a gel dryer (e.g. Bio-Rad Gel Dryer Model 583) and dry for 45-60 mins with a vacuum.
- xvi. Place dried gel in a phosphor screen cassette and expose the gel to a phosphor screen for 1-2 days. Scan the screen using a Phosphor Imager. **Figure 6.1C** shows an example of the results from a polymerase stop experiment.

## 6.4 Notes

1. The primer for primer extension and sequencing reaction is usually 18-20 nt long and is complementary to the 3' of the DNA template sequence containing the G-rich region.
2. The DNA template is typically 70-80 nt in length and contains a G-rich region close to the 5' end of the template.
3. Columns containing radioactive material should be disposed properly.
4. The annealing reaction can be prepared as follows: 7.5  $\mu\text{L}$  labeled primer (2  $\mu\text{M}$ ), 0.75  $\mu\text{L}$  template DNA (20  $\mu\text{M}$ ), and 41.75  $\mu\text{L}$  water.
5. The boundary of the gel should be marked directly on the X-ray film during the exposure. After developing the film, place the film underneath the gel and align the gel with the film based on previously marked gel boundary.
6. The solvent containing radioactive material should be disposed properly.
7. The optimal radioactivity of the resuspended samples is between 10,000 - 20,000 cpm/ $\mu\text{L}$ . If the total radioactivity of purified sample is not high enough, the concentration can be reduced to 5,000 - 10,000 cpm/ $\mu\text{L}$ , but the total volume to be added in Section 3.4.1 needs to be scaled down accordingly.
8. The annealing temperature is usually 2 degrees below the  $T_m$  of the primer.
9. G-quadruplexes are very stable in the solution containing a high concentration of potassium. To determine the effects of tested compounds on G4 stability, low concentration of potassium or other cations, such as lithium or sodium, can be used in the polymerase reaction. Additionally, a series dilution of the tested compound may be evaluated to determine the optimal condition.
10. The temperature can be set between 37-60  $^{\circ}\text{C}$  depending on the objective of the experiment and the  $T_m$  of the primer. The higher the temperature, the less stable the G4 structures are.
11. Before preparing the gel solution, clean plates and spacer very carefully in order to prevent air bubbles from forming in the gel and then assemble the glass plates in gel caster. After addition of TEMED, pour quickly the gel solution and immediately insert the appropriate comb into the gel before the acrylamide polymerizes and allow the acrylamide to polymerize for 60 min at room

temperature. To completely dissolve urea crystalline powder, the solution can be heated up using a microwave oven for 30 secs with caution.

12. It is important that the current does not exceed 40 mA as the gel can overheat.

## 6.5 Footnotes

This chapter has been previously published as:

Wu, G. & Han, H. A DNA polymerase stop assay for characterization of G-quadruplex formation and identification of G-quadruplex-interactive compounds. *Methods Mol Biol* 2035, 223-231 (2019). (Copyright © 2019, Springer Science Business Media, LLC, part of Springer Nature)

Author contributions: G.W. and H.H. wrote the paper.

## CHAPTER 7. CONCLUSION

DNA G-quadruplex (G4) is a class of secondary structures that are implicated in all nuclear processes involving DNA. Research in the past 30 years has demonstrated a wide range of biological functions and therapeutic potential of DNA G4 structures in cancer. Recent advances in cellular G4 detection have established a substantial body of evidence for the existence of G4 structures in human cells. More experimental data shows G4-formation in the regulatory region of the human genome, which can be regulated by proteins, providing new insights into the biological functions of G4 structures. Those fundamental insights, together with an increased number of high-resolution G4 structures, made selective therapeutic targeting G4 structures to treat cancer possible. *MYC*, a critical transcriptional factor, is commonly overexpressed in cancer cells. The *MYC* promoter G4 (MycG4) is the first and most extensively studied promoter G4 and functions as a transcription silencer. My doctoral research mainly focuses on the genomic MycG4 to i) understand the importance of protein interactions with MycG4 in *MYC* transcriptional regulation, ii) explore the potentials of MycG4 as a target for cancer treatment, and iii) establish a multifaceted approach to facilitate characterizing protein and ligand interactions of MycG4 in the lab. Additionally, I have worked on other genomic G4s, including telomeric G4s, BCL-2 and PDGFR- $\beta$  promoter G4s.

The dynamic regulation of *MYC* transcription requires the active unfolding of MycG4 structure, while the MycG4 is very stable with a melting temperature of over 85 °C in K<sup>+</sup> solution<sup>143-145</sup>. It is unknown what drives MycG4 unfolding and promotes *MYC* transcriptional activation in cancer cells. My doctoral studies fill the gap. We have discovered that DDX5 unfolds MycG4 with great proficiency and thereby transactivates *MYC* expression (**Publication #4**). To understand the functions of DDX5 on MycG4, we have characterized the conformation of MycG4-DNA within the DDX5-MycG4 complex using FRET spectroscopy, DNA footprinting, and CD spectroscopy. While DDX5 is known as a dsRNA helicase, our results show that DDX5 is a highly active DNA G4-resolvase that neither requires extended loading of ss-tails nor ATP hydrolysis for G4-unfolding. Strikingly, our protein-binding-ELISA experiments reveal specific and high-affinity binding to G4 structures regardless of whether the substrate is DNA or RNA. To elucidate the cellular functions of DDX5, we have analyzed publicly available DDX5 ChIP-seq data and

found that chromatin binding sites of DDX5 are G-rich sequences in cancer cells. Furthermore, our ChIP-qPCR and Western Blot results show that DDX5 is enriched at the *MYC* promoter and activates *MYC* transcription. Importantly, the DDX5 interaction with the promoter MycG4 and DDX5-mediated *MYC* gene transcriptional activation can be inhibited by G4-interactive small-molecules. Furthermore, knock-down of DDX5 expression using DDX5-specific siRNA in cancer cells result in downregulation of *MYC* expression and sensitization to G4-interactive small molecules. Our results thus uncover a novel function of DDX5 in resolving DNA and RNA G4s and establish a new molecular target to suppress *MYC* for cancer intervention.

The MycG4 transcriptional silencer has emerged as an attractive cancer-specific molecular target for drug development. Small molecules that bind and stabilize MycG4 can repress *MYC* transcription<sup>6, 10, 80</sup>. Indenoisoquinolines are human topoisomerase I inhibitors in clinical testing. However, some indenoisoquinolines with potent anticancer activity do not exhibit strong topoisomerase I inhibition<sup>3, 15</sup>, suggesting a separate mechanism of action. Our recent studies have discovered a novel mechanism of action of indenoisoquinolines (**Publication #5**). Using FRET, NMR, and CD spectroscopy, we have demonstrated a number of indenoisoquinolines can strongly bind and stabilize MycG4. Furthermore, they can lower *MYC* levels in cancer cells. Molecular docking studies suggest the potential binding model of indenoisoquinolines to MycG4. The analysis of indenoisoquinoline analogues for their *MYC* inhibitory activity, topoisomerase I inhibitory activity, and anticancer activity reveals a synergistic effect of *MYC* inhibition and topoisomerase I inhibition on anticancer activity, suggesting dual targeting of *MYC* and topoisomerase I may serve as a novel strategy for anticancer drug development. In addition to indenoisoquinolines, I have tested the *MYC*-inhibitory activity of small molecule BMVC using Western Blot, RT-qPCR, and MTS assay and showed that BMVC represses *MYC* in cancer cells (**Publication #3**).

Beyond *MYC*, I have also worked on ligand interactions of other genomic G4s, including those formed in human telomeres (**Publications #6 and #9**), the BCL-2 promoter (**Publication #11**), and the PDGFR- $\beta$  promoter (**Publications #2 and #10**). The results reveal different G4 structures are formed and can be specifically targeted by different small molecules, generating distinct cellular effects. These analyses illustrate the potential of the G4 structures as molecular targets for cancer treatments.

To facilitate the characterization of protein and ligand interactions with MycG4, I have established DNA polymerase stop (**Publications #7**) and DMS footprinting (**Publication #8**) experiments in the lab. The DNA polymerase stop assay utilizes DNA polymerase stalling at G-quadruplex structures when stabilized by either a G4-interactive ligand or protein. This assay can determine the binding specificity of small molecules and proteins to the MycG4 over duplex DNA. In the DMS footprinting experiment, the exposed N7s of guanine nucleotides are methylated by DMS, cleaved by subsequent piperidine treatment, and then visualized on a sequencing gel. Because G4 and other ssDNA conformations have distinct patterns of base pairing and unpairing, they display different characteristic profiles of sensitivity to DMS, allowing to monitor the G4-conformation change induced by ligands and proteins. In collaboration with Charles Vinson's lab at NCI, we have developed custom G4 microarrays to determine the binding preferences of proteins and small molecules to thousands of G4s (**Publication #1**). In this method, thousands of G4 oligos were attached to a glass surface and our results demonstrate that the microarray platform can be used to access the binding preferences of molecules to G4 structures on a large scale.

In summary, my doctoral studies illustrate how proteins and small molecules interact with MycG4 and are involved in the G4-dependent transcriptional regulation. Understanding those molecular interactions and cellular effects sheds light on the great potential of those unique secondary structures to become a new class of molecular receptors for the next wave of molecularly targeted onco-therapeutics.

## REFERENCES

- [1] Bang, I. (1910) Untersuchungen über die Guanylsäure, *Biochemische Zeitschrift* 26, 293-311.
- [2] Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) Helix formation by guanylic acid, *Proc Natl Acad Sci U S A* 48, 2013-2018.
- [3] Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model, *Cell* 59, 871-880.
- [4] Sen, D., and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA, *Nature* 344, 410-414.
- [5] Hud, N. V., Smith, F. W., Anet, F. A. L., and Feigon, J. (1996) The Selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA Quadruplexes Is Dominated by Relative Free Energies of Hydration: A Thermodynamic Analysis by 1H NMR, *Biochemistry* 35, 15383-15390.
- [6] Chen, Y., and Yang, D. Z. (2012) Sequence, stability, and structure of G-quadruplexes and their interactions with drugs, *Curr Protoc Nucleic Acid Chem* 50, 17.15.11–17.15.17.
- [7] Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., and Blackburn, E. H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs, *Cell* 51, 899-908.
- [8] Sundquist, W. I., and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, *Nature* 342, 825-829.
- [9] Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature* 334, 364-366.
- [10] Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proc Natl Acad Sci U S A* 99, 11593-11598.
- [11] Gray, L. T., Vallur, A. C., Eddy, J., and Maizels, N. (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD, *Nat Chem Biol* 10, 313-318.
- [12] Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures, *Nat Rev Genet* 13, 770-780.
- [13] Piazza, A., Boule, J. B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M. P., and Nicolas, A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*, *Nucleic Acids Res* 38, 4337-4348.
- [14] Ribeyre, C., Lopes, J., Boule, J. B., Piazza, A., Guedin, A., Zakian, V. A., Mergny, J. L., and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS Genet* 5, e1000475.
- [15] Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S., and Kurreck, J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif, *RNA* 14, 1290-1296.
- [16] Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, *Nat Chem Biol* 3, 218-221.
- [17] Huang, H., Zhang, J., Harvey, S. E., Hu, X., and Cheng, C. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPf, *Genes Dev* 31, 2296-2309.

- [18] Yang, S. Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A. G., Wong, J. M. Y., and Monchaud, D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells, *Nat Commun* 9, 4730.
- [19] Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome, *Nucleic Acids Res* 35, 406-413.
- [20] Hansel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., Tannahill, D., and Balasubramanian, S. (2016) G-quadruplex structures mark human regulatory chromatin, *Nat Genet* 48, 1267-1272.
- [21] Huppert, J. L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs, *Nucleic Acids Res* 36, 6260-6268.
- [22] Rouleau, S. G., Garant, J. M., Bolduc, F., Bisaillon, M., and Perreault, J. P. (2018) G-Quadruplexes influence pri-microRNA processing, *RNA biology* 15, 198-206.
- [23] Rouleau, S., Glouzon, J. S., Brumwell, A., Bisaillon, M., and Perreault, J. P. (2017) 3' UTR G-quadruplexes regulate miRNA binding, *RNA* 23, 1172-1179.
- [24] Fay, M. M., Lyons, S. M., and Ivanov, P. (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms, *J. Mol. Biol.* 429, 2127-2147.
- [25] Chen, Y., and Yang, D. (2012) Sequence, stability, and structure of G-quadruplexes and their interactions with drugs, *Curr Protoc Nucleic Acid Chem Chapter 17*, Unit17 15.
- [26] Verdun, R. E., and Karlseder, J. (2007) Replication and protection of telomeres, *Nature* 447, 924-931.
- [27] O'Sullivan, R. J., and Karlseder, J. (2010) Telomeres: protecting chromosomes against genome instability, *Nat Rev Mol Cell Biol* 11, 171-181.
- [28] Hackett, J. A., Feldser, D. M., and Greider, C. W. (2001) Telomere dysfunction increases mutation rate and genomic instability, *Cell* 106, 275-286.
- [29] Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C. B., and Lansdorp, P. M. (1994) Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age, *Proc Natl Acad Sci U S A* 91, 9857-9860.
- [30] Maciejowski, J., and de Lange, T. (2017) Telomeres in cancer: tumour suppression and genome instability, *Nat Rev Mol Cell Biol* 18, 175-186.
- [31] Chai, W., Shay, J. W., and Wright, W. E. (2005) Human telomeres maintain their overhang length at senescence, *Mol Cell Biol* 25, 2158-2168.
- [32] Levy, M. Z., Allsopp, R. C., Futcher, A. B., Greider, C. W., and Harley, C. B. (1992) Telomere end-replication problem and cell aging, *J. Mol. Biol.* 225, 951-960.
- [33] Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation, *Cell* 144, 646-674.
- [34] Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts, *Cell* 43, 405-413.
- [35] Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., Coviello, G. M., Wright, W. E., Weinrich, S. L., and Shay, J. W. (1994) Specific association of human telomerase activity with immortal cells and cancer, *Science* 266, 2011-2015.
- [36] Shay, J. W., and Bacchetti, S. (1997) A survey of telomerase activity in human cancer, *Eur J Cancer* 33, 787-791.
- [37] Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., and Reddel, R. R. (1997) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines, *Nat Med* 3, 1271-1274.

- [38] Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S., and Reddel, R. R. (1995) Telomere elongation in immortal human cells without detectable telomerase activity, *The EMBO Journal* 14, 4240-4248.
- [39] Henson, J. D., and Reddel, R. R. (2010) Assaying and investigating Alternative Lengthening of Telomeres activity in human cells and cancers, *FEBS Lett* 584, 3800-3811.
- [40] Dilley, R. L., Verma, P., Cho, N. W., Winters, H. D., Wondisford, A. R., and Greenberg, R. A. (2016) Break-induced telomere synthesis underlies alternative telomere maintenance, *Nature* 539, 54-58.
- [41] Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M. (1991) Inhibition of telomerase by G-quartet DNA structures, *Nature* 350, 718-720.
- [42] Rhodes, D., and Lipps, H. J. (2015) G-quadruplexes and their regulatory roles in biology, *Nucleic Acids Res* 43, 8627-8637.
- [43] Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M., Jr., and Mergny, J. L. (2016) G-quadruplexes and helicases, *Nucleic Acids Res* 44, 1989-2006.
- [44] Sun, D., Thompson, B., Cathers, B. E., Salazar, M., Kerwin, S. M., Trent, J. O., Jenkins, T. C., Neidle, S., and Hurley, L. H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound, *J Med Chem* 40, 2113-2116.
- [45] Hurley, L. H., Wheelhouse, R. T., Sun, D., Kerwin, S. M., Salazar, M., Fedoroff, O. Y., Han, F. X., Han, H., Izbicka, E., and Von Hoff, D. D. (2000) G-quadruplexes as targets for drug design, *Pharmacology & Therapeutics* 85, 141-158.
- [46] Mergny, J. L., and Helene, C. (1998) G-quadruplex DNA: a target for drug design, *Nat Med* 4, 1366-1367.
- [47] Neidle, S., and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery, *Nat Rev Drug Discov* 1, 383-393.
- [48] Yang, D., and Okamoto, K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs, *Future Med Chem* 2, 619-646.
- [49] Parkinson, G. N., Lee, M. P., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature* 417, 876-880.
- [50] Riou, J. F., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Megnin-Chanet, F., Helene, C., and Mergny, J. L. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands, *Proc Natl Acad Sci U S A* 99, 2672-2677.
- [51] Gowan, S. M., Heald, R., Stevens, M. F., and Kelland, L. R. (2001) Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes, *Mol Pharmacol* 60, 981-988.
- [52] Harrison, R. J., Reszka, A. P., Haider, S. M., Romagnoli, B., Morrell, J., Read, M. A., Gowan, S. M., Incles, C. M., Kelland, L. R., and Neidle, S. (2004) Evaluation of by disubstituted acridone derivatives as telomerase inhibitors: the importance of G-quadruplex binding, *Bioorg Med Chem Lett* 14, 5845-5849.
- [53] Incles, C. M., Schultes, C. M., Kempfski, H., Koehler, H., Kelland, L. R., and Neidle, S. (2004) A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells, *Mol Cancer Ther* 3, 1201-1206.
- [54] Shay, J. W., Reddel, R. R., and Wright, W. E. (2012) Cancer. Cancer and telomeres--an ALternative to telomerase, *Science* 336, 1388-1390.
- [55] Hu, J., Hwang, S. S., Liesa, M., Gan, B., Sahin, E., Jaskelioff, M., Ding, Z., Ying, H., Boutin, A. T., Zhang, H., Johnson, S., Ivanova, E., Kost-Alimova, M., Protopopov, A., Wang, Y.

- A., Shirihai, O. S., Chin, L., and DePinho, R. A. (2012) Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer, *Cell* 148, 651-663.
- [56] Wu, G., Chen, L., Liu, W., and Yang, D. (2019) Molecular Recognition of the Hybrid-Type G-Quadruplexes in Human Telomeres, *Molecules* 24, 1578.
- [57] Grand, C. L., Han, H., Munoz, R. M., Weitman, S., Von Hoff, D. D., Hurley, L. H., and Bearss, D. J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo, *Mol Cancer Ther* 1, 565-573.
- [58] Bedrat, A., Lacroix, L., and Mergny, J. L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter, *Nucleic Acids Res* 44, 1746-1759.
- [59] Kouzine, F., Wojtowicz, D., Baranello, L., Yamane, A., Nelson, S., Resch, W., Kieffer-Kwon, K. R., Benham, C. J., Casellas, R., Przytycka, T. M., and Levens, D. (2017) Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome, *Cell Syst* 4, 344-356 e347.
- [60] Broxson, C., Beckett, J., and Tornaletti, S. (2011) Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene, *Biochemistry* 50, 4162-4172.
- [61] Zheng, K. W., Xiao, S., Liu, J. Q., Zhang, J. Y., Hao, Y. H., and Tan, Z. (2013) Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control, *Nucleic Acids Res* 41, 5533-5541.
- [62] Eddy, J., Vallur, A. C., Varma, S., Liu, H., Reinhold, W. C., Pommier, Y., and Maizels, N. (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes, *Nucleic Acids Res* 39, 4975-4983.
- [63] Kendrick, S., and Hurley, L. H. (2010) The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements, *Pure Appl Chem* 82, 1609-1621.
- [64] DesJardins, E., and Hay, N. (1993) Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human c-myc promoters, *Mol Cell Biol* 13, 5710-5724.
- [65] Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?, *Nat Rev Drug Discov* 10, 261-275.
- [66] Law, M. J., Lower, K. M., Voon, H. P., Hughes, J. R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., Abbott, A., Wilder, S. P., Taylor, S., Santos, G. M., Cross, J., Ayyub, H., Jones, S., Ragoussis, J., Rhodes, D., Dunham, I., Higgs, D. R., and Gibbons, R. J. (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner, *Cell* 143, 367-378.
- [67] Shioda, N., Yabuki, Y., Yamaguchi, K., Onozato, M., Li, Y., Kurosawa, K., Tanabe, H., Okamoto, N., Era, T., Sugiyama, H., Wada, T., and Fukunaga, K. (2018) Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome, *Nat Med* 24, 802-813.
- [68] Cree, S. L., Fredericks, R., Miller, A., Pearce, F. G., Filichev, V., Fee, C., and Kennedy, M. A. (2016) DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro, *FEBS Lett* 590, 2870-2883.
- [69] Mao, S. Q., Ghanbarian, A. T., Spiegel, J., Martinez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hansel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018) DNA G-quadruplex structures mold the DNA methylome, *Nat Struct Mol Biol* 25, 951-957.

- [70] Schwarzbauer, K., Bodenhofer, U., and Hochreiter, S. (2012) Genome-wide chromatin remodeling identified at GC-rich long nucleosome-free regions, *PLoS One* 7, e47924.
- [71] Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T. M., and Levens, D. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force, *Nat Struct Mol Biol* 20, 396-403.
- [72] Naughton, C., Avlonitis, N., Corless, S., Prendergast, J. G., Mati, I. K., Eijk, P. P., Cockroft, S. L., Bradley, M., Ylstra, B., and Gilbert, N. (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures, *Nat Struct Mol Biol* 20, 387-395.
- [73] Kouzine, F., Sanford, S., Elisha-Feil, Z., and Levens, D. (2008) The functional response of upstream DNA to dynamic supercoiling in vivo, *Nat Struct Mol Biol* 15, 146-154.
- [74] Wolfl, S., Wittig, B., Dorbic, T., and Rich, A. (1997) Identification of processes that influence negative supercoiling in the human c-myc gene, *Biochim Biophys Acta* 1352, 213-221.
- [75] Tomonaga, T., and Levens, D. (1996) Activating transcription from single stranded DNA, *Proc Natl Acad Sci U S A* 93, 5830-5835.
- [76] Wittig, B., Dorbic, T., and Rich, A. (1991) Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei, *Proc Natl Acad Sci U S A* 88, 2259-2263.
- [77] Sekibo, D. A. T., and Fox, K. R. (2017) The effects of DNA supercoiling on G-quadruplex formation, *Nucleic Acids Res* 45, 12069-12079.
- [78] Zheng, K. W., He, Y. D., Liu, H. H., Li, X. M., Hao, Y. H., and Tan, Z. (2017) Superhelicity Constrains a Localized and R-Loop-Dependent Formation of G-Quadruplexes at the Upstream Region of Transcription, *ACS Chem Biol* 12, 2609-2618.
- [79] Xia, Y., Zheng, K. W., He, Y. D., Liu, H. H., Wen, C. J., Hao, Y. H., and Tan, Z. (2018) Transmission of dynamic supercoiling in linear and multi-way branched DNAs and its regulation revealed by a fluorescent G-quadruplex torsion sensor, *Nucleic Acids Res* 46, 7418-7424.
- [80] Brooks, T. A., and Hurley, L. H. (2009) The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics, *Nat Rev Cancer* 9, 849-861.
- [81] Howard, F. B., Frazier, J., and Miles, H. T. (1977) Stable and metastable forms of poly(G), *Biopolymers* 16, 791-809.
- [82] Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures, *Nucleic Acids Res* 36, 5482-5515.
- [83] Gonzalez, V., and Hurley, L. H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, *Biochemistry* 49, 9706-9714.
- [84] Qin, Y., and Hurley, L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions, *Biochimie* 90, 1149-1171.
- [85] Dai, J., Carver, M., Punchihewa, C., Jones, R. A., and Yang, D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence, *Nucleic Acids Res* 35, 4927-4940.
- [86] Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter, *J Am Chem Soc* 126, 8710-8716.

- [87] Ambrus, A., Chen, D., Dai, J., Jones, R. A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, *Biochemistry* 44, 2048-2058.
- [88] Mathad, R. I., Hatzakis, E., Dai, J., and Yang, D. (2011) c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability, *Nucleic Acids Res* 39, 9023-9033.
- [89] Dickerhoff, J., Onel, B., Chen, L., Chen, Y., and Yang, D. (2019) Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length, *ACS Omega* 4, 2533-2539.
- [90] Agrawal, P., Hatzakis, E., Guo, K., Carver, M., and Yang, D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K<sup>+</sup>: insights into loop interactions of the parallel G-quadruplexes, *Nucleic Acids Res* 41, 10584-10592.
- [91] Agrawal, P., Lin, C., Mathad, R. I., Carver, M., and Yang, D. (2014) The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K<sup>+</sup> solution, *J Am Chem Soc* 136, 1750-1753.
- [92] Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M., and Yang, D. (2016) A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription, *J Am Chem Soc* 138, 2563-2570.
- [93] Kerkour, A., Marquevielle, J., Ivashchenko, S., Yatsunyk, L. A., Mergny, J. L., and Salgado, G. F. (2017) High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation, *J Biol Chem* 292, 8082-8091.
- [94] Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter, *J Am Chem Soc* 129, 4386-4392.
- [95] Hsu, S. T., Varnai, P., Bugaut, A., Reszka, A. P., Neidle, S., and Balasubramanian, S. (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics, *J Am Chem Soc* 131, 13399-13409.
- [96] Kuryavyi, V., Phan, A. T., and Patel, D. J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter, *Nucleic Acids Res* 38, 6757-6773.
- [97] Tong, X., Lan, W., Zhang, X., Wu, H., Liu, M., and Cao, C. (2011) Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence, *Nucleic Acids Res* 39, 6753-6763.
- [98] Chen, Y., Agrawal, P., Brown, R. V., Hatzakis, E., Hurley, L., and Yang, D. (2012) The major G-quadruplex formed in the human platelet-derived growth factor receptor beta promoter adopts a novel broken-strand structure in K<sup>+</sup> solution, *J Am Chem Soc* 134, 13220-13223.
- [99] Onel, B., Carver, M., Agrawal, P., Hurley, L. H., and Yang, D. (2018) The 3'-end region of the human PDGFR-beta core promoter nuclease hypersensitive element forms a mixture of two unique end-insertion G-quadruplexes, *Biochim Biophys Acta Gen Subj* 1862, 846-854.
- [100] Sengar, A., Vandana, J. J., Chambers, V. S., Di Antonio, M., Winnerdy, F. R., Balasubramanian, S., and Phan, A. T. (2019) Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter, *Nucleic Acids Res* 47, 1564-1572.
- [101] Guédin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010) How long is too long? Effects of loop size on G-quadruplex stability  
10.1093/nar/gkq639, *Nucleic Acids Res* 38, 7858-7868.

- [102] Dai, J., Chen, D., Jones, R. A., Hurley, L. H., and Yang, D. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region, *Nucleic Acids Res* 34, 5133-5144.
- [103] Chaires, J. B., Trent, J. O., Gray, R. D., Dean, W. L., Buscaglia, R., Thomas, S. D., and Miller, D. M. (2014) An improved model for the hTERT promoter quadruplex, *PLoS One* 9, e115580.
- [104] Micheli, E., Martufi, M., Cacchione, S., De Santis, P., and Savino, M. (2010) Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives, *Biophys Chem* 153, 43-53.
- [105] Lim, K. W., Lacroix, L., Yue, D. J., Lim, J. K., Lim, J. M., and Phan, A. T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter, *J Am Chem Soc* 132, 12331-12342.
- [106] Palumbo, S. L., Ebbinghaus, S. W., and Hurley, L. H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands, *J Am Chem Soc* 131, 10878-10891.
- [107] Borah, S., Xi, L., Zaug, A. J., Powell, N. M., Dancik, G. M., Cohen, S. B., Costello, J. C., Theodorescu, D., and Cech, T. R. (2015) Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer, *Science* 347, 1006-1010.
- [108] Dani, C., Blanchard, J. M., Piechaczyk, M., El Sabouty, S., Marty, L., and Jeanteur, P. (1984) Extreme instability of myc mRNA in normal and transformed human cells, *Proc Natl Acad Sci U S A* 81, 7046-7050.
- [109] Ramsay, G., Evan, G. I., and Bishop, J. M. (1984) The protein encoded by the human proto-oncogene c-myc, *Proc Natl Acad Sci U S A* 81, 7742-7746.
- [110] Hessmann, E., Schneider, G., Ellenrieder, V., and Siveke, J. T. (2016) MYC in pancreatic cancer: novel mechanistic insights and their translation into therapeutic strategies, *Oncogene* 35, 1609-1618.
- [111] Nesbit, C. E., Tersak, J. M., and Prochownik, E. V. (1999) MYC oncogenes and human neoplastic disease, *Oncogene* 18, 3004-3016.
- [112] Meyer, N., and Penn, L. Z. (2008) Reflecting on 25 years with MYC, *Nat Rev Cancer* 8, 976-990.
- [113] Jain, M., Arvanitis, C., Chu, K., Dewey, W., Leonhardt, E., Trinh, M., Sundberg, C. D., Bishop, J. M., and Felsher, D. W. (2002) Sustained loss of a neoplastic phenotype by brief inactivation of MYC, *Science* 297, 102-104.
- [114] Soucek, L., Whitfield, J., Martins, C. P., Finch, A. J., Murphy, D. J., Sodik, N. M., Karnezis, A. N., Swigart, L. B., Nasi, S., and Evan, G. I. (2008) Modelling Myc inhibition as a cancer therapy, *Nature* 455, 679-683.
- [115] Soucek, L., Whitfield, J. R., Sodik, N. M., Masso-Valles, D., Serrano, E., Karnezis, A. N., Swigart, L. B., and Evan, G. I. (2013) Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice, *Genes Dev* 27, 504-513.
- [116] Blackwood, E. M., and Eisenman, R. N. (1991) Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc, *Science* 251, 1211-1217.
- [117] Ji, H., Wu, G., Zhan, X., Nolan, A., Koh, C., De Marzo, A., Doan, H. M., Fan, J., Cheadle, C., Fallahi, M., Cleveland, J. L., Dang, C. V., and Zeller, K. I. (2011) Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation, *PLoS One* 6, e26057.

- [118] Margolin, A. A., Palomero, T., Sumazin, P., Califano, A., Ferrando, A. A., and Stolovitzky, G. (2009) ChIP-on-chip significance analysis reveals large-scale binding and regulation by human transcription factor oncogenes, *Proc Natl Acad Sci U S A* 106, 244-249.
- [119] Van Dang, C., and McMahon, S. B. (2010) Emerging Concepts in the Analysis of Transcriptional Targets of the MYC Oncoprotein: Are the Targets Targetable?, *Genes Cancer* 1, 560-567.
- [120] Sabo, A., and Amati, B. (2018) BRD4 and MYC-clarifying regulatory specificity, *Science* 360, 713-714.
- [121] Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J. J., Herzog, V. A., Reichholf, B., Cisneros, D. A., Hoffmann, T., Schlapansky, M. F., Bhat, P., von Haeseler, A., Kocher, T., Obenauf, A. C., Popow, J., Ameres, S. L., and Zuber, J. (2018) SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis, *Science* 360, 800-805.
- [122] Lin, C. Y., Loven, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., Lee, T. I., and Young, R. A. (2012) Transcriptional amplification in tumor cells with elevated c-Myc, *Cell* 151, 56-67.
- [123] Horiuchi, D., Anderton, B., and Goga, A. (2014) Taking on challenging targets: making MYC druggable, *Am Soc Clin Oncol Educ Book*, e497-502.
- [124] McKeown, M. R., and Bradner, J. E. (2014) Therapeutic strategies to inhibit MYC, *Cold Spring Harb Perspect Med* 4.
- [125] Chen, H., Liu, H., and Qing, G. (2018) Targeting oncogenic Myc as a strategy for cancer treatment, *Signal Transduct Target Ther* 3, 5.
- [126] Prochownik, E. V., and Vogt, P. K. (2010) Therapeutic Targeting of Myc, *Genes Cancer* 1, 650-659.
- [127] Yang, Z., Yik, J. H., Chen, R., He, N., Jang, M. K., Ozato, K., and Zhou, Q. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4, *Mol Cell* 19, 535-545.
- [128] Peterlin, B. M., and Price, D. H. (2006) Controlling the elongation phase of transcription with P-TEFb, *Mol Cell* 23, 297-305.
- [129] Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., Kastiris, 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., and Mitsiades, C. S. (2011) BET bromodomain inhibition as a therapeutic strategy to target c-Myc, *Cell* 146, 904-917.
- [130] Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philpott, M., Munro, S., McKeown, M. R., Wang, Y., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. A., Wiest, O., Kung, A. L., Knapp, S., and Bradner, J. E. (2010) Selective inhibition of BET bromodomains, *Nature* 468, 1067-1073.
- [131] Spencer, C. A., and Groudine, M. (1991) Control of c-myc regulation in normal and neoplastic cells, *Adv Cancer Res* 56, 1-48.
- [132] Wierstra, I., and Alves, J. (2008) The c-myc promoter: still MysterY and challenge, *Adv Cancer Res* 99, 113-333.

- [133] Hu, H. M., Kanda, K., Zhang, L., and Boxer, L. M. (2007) Activation of the c-myc p1 promoter in Burkitt's lymphoma by the hs3 immunoglobulin heavy-chain gene enhancer, *Leukemia* 21, 747-753.
- [134] Broome, H. E., Reed, J. C., Godillot, E. P., and Hoover, R. G. (1987) Differential promoter utilization by the c-myc gene in mitogen- and interleukin-2-stimulated human lymphocytes, *Mol Cell Biol* 7, 2988-2993.
- [135] Chang, Y., Spicer, D. B., and Sonenshein, G. E. (1991) Effects of IL-3 on promoter usage, attenuation and antisense transcription of the c-myc oncogene in the IL-3-dependent Ba/F3 early pre-B cell line, *Oncogene* 6, 1979-1982.
- [136] Ruppert, C., Goldowitz, D., and Wille, W. (1986) Proto-oncogene c-myc is expressed in cerebellar neurons at different developmental stages, *EMBO J* 5, 1897-1901.
- [137] Avigan, M. I., Strober, B., and Levens, D. (1990) A far upstream element stimulates c-myc expression in undifferentiated leukemia cells, *J Biol Chem* 265, 18538-18545.
- [138] Liu, J., Kouzine, F., Nie, Z., Chung, H. J., Elisha-Feil, Z., Weber, A., Zhao, K., and Levens, D. (2006) The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression, *EMBO J* 25, 2119-2130.
- [139] Simonsson, T., Pecinka, P., and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc, *Nucleic Acids Res* 26, 1167-1172.
- [140] Michelotti, E. F., Tomonaga, T., Krutzsch, H., and Levens, D. (1995) Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene, *J Biol Chem* 270, 9494-9499.
- [141] Tomonaga, T., and Levens, D. (1995) Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator, *J Biol Chem* 270, 4875-4881.
- [142] Takimoto, M., Tomonaga, T., Matunis, M., Avigan, M., Krutzsch, H., Dreyfuss, G., and Levens, D. (1993) Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro, *J Biol Chem* 268, 18249-18258.
- [143] Hatzakis, E., Okamoto, K., and Yang, D. (2010) Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes, *Biochemistry* 49, 9152-9160.
- [144] Seenisamy, J., Rezler, E. M., Powell, T. J., Tye, D., Gokhale, V., Joshi, C. S., Siddiqui-Jain, A., and Hurley, L. H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4, *J Am Chem Soc* 126, 8702-8709.
- [145] Mathad, R. I., and Yang, D. (2011) G-quadruplex structures and G-quadruplex-interactive compounds, *Methods Mol Biol* 735, 77-96.
- [146] Gray, R. D., Trent, J. O., and Chaires, J. B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex, *J. Mol. Biol.* 426, 1629-1650.
- [147] Stump, S., Mou, T. C., Sprang, S. R., Natale, N. R., and Beall, H. D. (2018) Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene, *PLoS One* 13, e0205584.
- [148] Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv Drug Deliv Rev* 46, 3-26.
- [149] Ou, T. M., Lu, Y. J., Zhang, C., Huang, Z. S., Wang, X. D., Tan, J. H., Chen, Y., Ma, D. L., Wong, K. Y., Tang, J. C., Chan, A. S., and Gu, L. Q. (2007) Stabilization of G-quadruplex

- DNA and down-regulation of oncogene c-myc by quindoline derivatives, *J Med Chem* 50, 1465-1474.
- [150] Dai, J., Carver, M., Hurley, L. H., and Yang, D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design, *J Am Chem Soc* 133, 17673-17680.
- [151] Boddupally, P. V. L., Hahn, S., Beman, C., De, B., Brooks, T. A., Gokhale, V., and Hurley, L. H. (2012) Anticancer Activity and Cellular Repression of c-MYC by the G-Quadruplex-Stabilizing 11-Piperazinylquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter, *J Med Chem* 55, 6076-6086.
- [152] Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H., and Brooks, T. A. (2011) Demonstration that drug-targeted downregulation of MYC in non-hodgkins lymphoma is directly mediated through the promoter G-quadruplex, *J Biol Chem* 286, 41018-41027.
- [153] Felsenstein, K. M., Saunders, L. B., Simmons, J. K., Leon, E., Calabrese, D. R., Zhang, S., Michalowski, A., Gareiss, P., Mock, B. A., and Schneekloth, J. S., Jr. (2016) Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression, *ACS Chem Biol* 11, 139-148.
- [154] Calabrese, D. R., Chen, X., Leon, E. C., Gaikwad, S. M., Phyo, Z., Hewitt, W. M., Alden, S., Hilimire, T. A., He, F., Michalowski, A. M., Simmons, J. K., Saunders, L. B., Zhang, S., Connors, D., Walters, K. J., Mock, B. A., and Schneekloth, J. S., Jr. (2018) Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex, *Nat Commun* 9, 4229.
- [155] Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei, *Proc Natl Acad Sci U S A* 98, 8572-8577.
- [156] Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H. J. (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates, *Nat Struct Mol Biol* 15, 598-604.
- [157] Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo, *Nat Struct Mol Biol* 12, 847-854.
- [158] Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells, *Nat Chem* 5, 182-186.
- [159] Henderson, A., Wu, Y., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., Brosh, R. M., Jr., Sen, D., and Lansdorp, P. M. (2014) Detection of G-quadruplex DNA in mammalian cells, *Nucleic Acids Res* 42, 860-869.
- [160] Zhang, H., Xing, Z., Mani, S. K., Bancel, B., Durantel, D., Zoulim, F., Tran, E. J., Merle, P., and Andrisani, O. (2016) RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis, *Hepatology* 64, 1033-1048.
- [161] Xing, Z., Wang, S., and Tran, E. J. (2017) Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with *S. cerevisiae* ortholog Dbp2 in transcriptional control and glucose metabolism, *RNA* 23, 1125-1138.
- [162] Liu, J. Q., Chen, C. Y., Xue, Y., Hao, Y. H., and Tan, Z. (2010) G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates, *J Am Chem Soc* 132, 10521-10527.

- [163] Shan, C., Yan, J. W., Wang, Y. Q., Che, T., Huang, Z. L., Chen, A. C., Yao, P. F., Tan, J. H., Li, D., Ou, T. M., Gu, L. Q., and Huang, Z. S. (2017) Design, Synthesis, and Evaluation of Isaindigotone Derivatives To Downregulate c-myc Transcription via Disrupting the Interaction of NM23-H2 with G-Quadruplex, *J Med Chem* 60, 1292-1308.
- [164] Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S. P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes, *Nat Chem Biol* 8, 301-310.
- [165] Marcu, K. B., Patel, A. J., and Yang, Y. (1997) Differential regulation of the c-MYC P1 and P2 promoters in the absence of functional tumor suppressors: implications for mechanisms of deregulated MYC transcription, *Curr Top Microbiol Immunol* 224, 47-56.
- [166] Hay, N., Bishop, J. M., and Levens, D. (1987) Regulatory elements that modulate expression of human c-myc, *Genes Dev* 1, 659-671.
- [167] Lavelle, C. (2009) Forces and torques in the nucleus: chromatin under mechanical constraints, *Biochemistry and cell biology = Biochimie et biologie cellulaire* 87, 307-322.
- [168] Sun, D., and Hurley, L. H. (2009) 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* 52, 2863-2874.
- [169] Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I., and Levens, D. (1996) Heterogeneous nuclear ribonucleoprotein K is a transcription factor, *Mol Cell Biol* 16, 2350-2360.
- [170] Linder, P., and Jankowsky, E. (2011) From unwinding to clamping - the DEAD box RNA helicase family, *Nat Rev Mol Cell Biol* 12, 505-516.
- [171] Lane, D. P., and Hoeffler, W. K. (1980) SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000, *Nature* 288, 167-170.
- [172] Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989) RNA helicase activity associated with the human p68 protein, *Nature* 339, 562-564.
- [173] Stevenson, R. J., Hamilton, S. J., MacCallum, D. E., Hall, P. A., and Fuller-Pace, F. V. (1998) Expression of the 'dead box' RNA helicase p68 is developmentally and growth regulated and correlates with organ differentiation/maturation in the fetus, *The Journal of pathology* 184, 351-359.
- [174] Fuller-Pace, F. V. (2013) DEAD box RNA helicase functions in cancer, *RNA biology* 10, 121-132.
- [175] Yang, L., Lin, C., and Liu, Z. R. (2005) Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation, *Molecular cancer research : MCR* 3, 355-363.
- [176] Wang, Z., Luo, Z., Zhou, L., Li, X., Jiang, T., and Fu, E. (2015) DDX5 promotes proliferation and tumorigenesis of non-small-cell lung cancer cells by activating beta-catenin signaling pathway, *Cancer Sci* 106, 1303-1312.
- [177] Shin, S., Rossow, K. L., Grande, J. P., and Janknecht, R. (2007) Involvement of RNA helicases p68 and p72 in colon cancer, *Cancer Res* 67, 7572-7578.
- [178] Clark, E. L., Coulson, A., Dalglish, C., Rajan, P., Nicol, S. M., Fleming, S., Heer, R., Gaughan, L., Leung, H. Y., Elliott, D. J., Fuller-Pace, F. V., and Robson, C. N. (2008) The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer, *Cancer Res* 68, 7938-7946.

- [179] Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., and Stillman, B. (2012) DDX5 regulates DNA replication and is required for cell proliferation in a subset of breast cancer cells, *Cancer Discov* 2, 812-825.
- [180] Janknecht, R. (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17), *American journal of translational research* 2, 223-234.
- [181] Fuller-Pace, F. V. (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators, *Biochim Biophys Acta* 1829, 756-763.
- [182] Fuller-Pace, F. V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation, *Nucleic Acids Res* 34, 4206-4215.
- [183] Cloutier, S. C., Ma, W. K., Nguyen, L. T., and Tran, E. J. (2012) The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription, *J Biol Chem* 287, 26155-26166.
- [184] Ma, W. K., Paudel, B. P., Xing, Z., Sabath, I. G., Rueda, D., and Tran, E. J. (2016) Recruitment, Duplex Unwinding and Protein-Mediated Inhibition of the Dead-Box RNA Helicase Dbp2 at Actively Transcribed Chromatin, *J. Mol. Biol.* 428, 1091-1106.
- [185] Tago, K., Funakoshi-Tago, M., Itoh, H., Furukawa, Y., Kikuchi, J., Kato, T., Suzuki, K., and Yanagisawa, K. (2015) Arf tumor suppressor disrupts the oncogenic positive feedback loop including c-Myc and DDX5, *Oncogene* 34, 314-322.
- [186] Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa, *Cell* 125, 287-300.
- [187] Kim, B. G., Evans, H. M., Dubins, D. N., and Chalikian, T. V. (2015) Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter, *Biochemistry* 54, 3420-3430.
- [188] Hellman, L. M., and Fried, M. G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, *Nature protocols* 2, 1849-1861.
- [189] Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D., and Felsenfeld, G. (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA, *Genes Dev* 24, 2543-2555.
- [190] Machanick, P., and Bailey, T. L. (2011) MEME-ChIP: motif analysis of large DNA datasets, *Bioinformatics* 27, 1696-1697.
- [191] Liu, F., Putnam, A., and Jankowsky, E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding, *Proc Natl Acad Sci U S A* 105, 20209-20214.
- [192] Tijerina, P., Mohr, S., and Russell, R. (2007) DMS footprinting of structured RNAs and RNA-protein complexes, *Nature protocols* 2, 2608-2623.
- [193] Kang, H. J., Kendrick, S., Hecht, S. M., and Hurley, L. H. (2014) 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* 136, 4172-4185.
- [194] Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H., and Luhrmann, R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs, *J. Mol. Biol.* 356, 1248-1262.
- [195] Wang, S. J., Zhang, C., You, Y., and Shi, C. M. (2012) Overexpression of RNA helicase p68 protein in cutaneous squamous cell carcinoma, *Clinical and experimental dermatology* 37, 882-888.

- [196] Gilman, B., Tijerina, P., and Russell, R. (2017) Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs, *Biochem Soc Trans* 45, 1313-1321.
- [197] Crabbe, L., Verdun, R. E., Haggblom, C. I., and Karlseder, J. (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity, *Science* 306, 1951-1953.
- [198] Chakraborty, P., and Grosse, F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes, *DNA repair* 10, 654-665.
- [199] Chen, M. C., Tippiana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., and Ferre-D'Amare, A. R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36, *Nature* 558, 465-469.
- [200] Yangyuoru, P. M., Bradburn, D. A., Liu, Z., Xiao, T. S., and Russell, R. (2018) The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism, *J Biol Chem* 293, 1924-1932.
- [201] Zaug, A. J., Podell, E. R., and Cech, T. R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro, *Proc Natl Acad Sci USA* 102, 10864-10869.
- [202] Hwang, H., Buncher, N., Opresko, P. L., and Myong, S. (2012) POT1-TPP1 regulates telomeric overhang structural dynamics, *Structure* 20, 1872-1880.
- [203] Salas, T. R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J. L., Favre, A., and Saintome, C. (2006) Human replication protein A unfolds telomeric G-quadruplexes, *Nucleic Acids Res* 34, 4857-4865.
- [204] McRae, E. K. S., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S. A. (2017) Human DDX21 binds and unwinds RNA guanine quadruplexes, *Nucleic Acids Res* 45, 6656-6668.
- [205] Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A. E., Sattentau, Q., Meinhart, A., and Proudfoot, N. J. (2018) RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination, *Mol Cell* 70, 650-662 e658.
- [206] Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., and Balasubramanian, S. (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts, *Nucleic Acids Res* 46, 11592-11604.
- [207] Sutherland, C., Cui, Y., Mao, H., and Hurley, L. H. (2016) A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1, *J Am Chem Soc* 138, 14138-14151.
- [208] Chung, W. J., Heddi, B., Tera, M., Iida, K., Nagasawa, K., and Phan, A. T. (2013) Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative, *J Am Chem Soc* 135, 13495-13501.
- [209] Wirmer-Bartoschek, J., Bendel, L. E., Jonker, H. R., Grün, J. T., Papi, F., Bazzicalupi, C., Messori, L., Gratteri, P., and Schwalbe, H. (2017) Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding, *Angewandte Chemie* 129, 7208-7212.
- [210] Liu, W., Zhong, Y. F., Liu, L. Y., Shen, C. T., Zeng, W., Wang, F., Yang, D., and Mao, Z. W. (2018) Solution structures of multiple G-quadruplex complexes induced by a platinum(II)-based tripod reveal dynamic binding, *Nat Commun* 9, 3496.
- [211] Lin, C., Wu, G., Wang, K., Onel, B., Sakai, S., Shao, Y., and Yang, D. (2018) Molecular Recognition of the Hybrid-2 Human Telomeric G-quadruplex by Epiberberine: Insights

- into Conversion of Telomeric G-quadruplex Structures, *Angew Chem Int Ed* 57, 10888-10893.
- [212] Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, *Nucleic Acids Res* 34, 2723-2735.
- [213] Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R. A., and Yang, D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation, *Nucleic Acids Res* 35, 2440-2450.
- [214] Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006) Structure of the human telomere in K<sup>+</sup> solution: an intramolecular (3 + 1) G-quadruplex scaffold, *J Am Chem Soc* 128, 9963-9970.
- [215] Phan, A. T., Luu, K. N., and Patel, D. J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K<sup>+</sup> solution, *Nucleic Acids Res* 34, 5715-5719.
- [216] Dai, J., Carver, M., and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures, *Biochimie* 90, 1172-1183.
- [217] Ortiz, L. M., Lombardi, P., Tillhon, M., and Scovassi, A. I. (2014) Berberine, an epiphany against cancer, *Molecules* 19, 12349-12367.
- [218] Tang, J., Feng, Y., Tsao, S., Wang, N., Curtain, R., and Wang, Y. (2009) Berberine and Coptidis rhizoma as novel antineoplastic agents: a review of traditional use and biomedical investigations, *J Ethnopharmacol* 126, 5-17.
- [219] Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., Savino, M., Schultes, C., and Neidle, S. (2006) Natural and synthetic G-quadruplex interactive berberine derivatives, *Bioorg Med Chem Lett* 16, 1707-1711.
- [220] Zhang, L., Liu, H., Shao, Y., Lin, C., Jia, H., Chen, G., Yang, D., and Wang, Y. (2015) Selective lighting up of epiberberine alkaloid fluorescence by fluorophore-switching aptamer and stoichiometric targeting of human telomeric DNA G-quadruplex multimer, *Anal Chem* 87, 730-737.
- [221] Zheng, X. H., Chen, H. Y., Tong, M. L., Ji, L. N., and Mao, Z. W. (2012) Platinum squares with high selectivity and affinity for human telomeric G-quadruplexes, *Chem Commun (Camb)* 48, 7607-7609.
- [222] Xu, C. X., Shen, Y., Hu, Q., Zheng, Y. X., Cao, Q., Qin, P. Z., Zhao, Y., Ji, L. N., and Mao, Z. W. (2014) Stabilization of human telomeric G-quadruplex and inhibition of telomerase activity by propeller-shaped trinuclear Pt(II) complexes, *Chem Asian J* 9, 2519-2526.
- [223] Zimmer, J., Tacconi, E. M. C., Folio, C., Badie, S., Porru, M., Klare, K., Tumiati, M., Markkanen, E., Halder, S., Ryan, A., Jackson, S. P., Ramadan, K., Kuznetsov, S. G., Biroccio, A., Sale, J. E., and Tarsounas, M. (2016) Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds, *Mol Cell* 61, 449-460.
- [224] Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N. J., Santos, N. D., Silvester, J., Wei, V., Garcia, J., Kabeer, F., Lai, D., Soriano, P., Banath, J., Chiu, D. S., Yap, D., Le, D. D., Ye, F. B., Zhang, A., Thu, K., Soong, J., Lin, S. C., Tsai, A. H., Osako, T., Algara, T., Saunders, D. N., Wong, J., Xian, J., Bally, M. B., Brenton, J. D., Brown, G. W., Shah, S. P., Cescon, D., Mak, T. W., Caldas, C., Stirling, P. C., Hieter, P., Balasubramanian, S., and Aparicio, S. (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours, *Nat Commun* 8, 14432.

- [225] Han, H., Hurley, L. H., and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds, *Nucleic Acids Res* 27, 537-542.
- [226] Brown, R. V., Wang, T., Chappeta, V. R., Wu, G., Onel, B., Chawla, R., Quijada, H., Camp, S. M., Chiang, E. T., and Lassiter, Q. R. (2017) The consequences of overlapping G-quadruplexes and i-motifs in the platelet-derived growth factor receptor  $\beta$  core promoter nuclease hypersensitive element can explain the unexpected effects of mutations and provide opportunities for selective targeting of both structures by small molecules to downregulate gene expression, *J. Am. Chem. Soc.* 139, 7456-7475.
- [227] Qin, Y., Fortin, J. S., Tye, D., Gleason-Guzman, M., Brooks, T. A., and Hurley, L. H. (2010) Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression, *Biochemistry* 49, 4208-4219.

## PUBLICATIONS

1. Sreejana Ray, Desiree Tillo, Robert E. Boer, Nima Assad, Mira Barshai, **Guanhui Wu**, Yaron Orenstein, Danzhou Yang, John S. Schneekloth Jr. and Charles Vinson, Custom DNA microarrays reveal diverse binding preferences of proteins and small molecules to thousands of G-quadruplexes. *ACS Chem. Biol.* 15, 4, 925-935 (2020) (*see abstracts*)
2. Kai-Bo Wang, Jonathan Dickerhoff, **Guanhui Wu** and Danzhou Yang. PDGFR- $\beta$  promoter forms a vacancy G-quadruplex that can be filled in by dGMP: solution structure and molecular recognition of guanine metabolites and drugs. *J. Am. Chem. Soc.* 142, 11, 5204-5211 (2020). (*see abstracts*)
3. Wenting Liu, Clement Lin, **Guanhui Wu**, Jixun Dai, Ta-Chau Chang and Danzhou Yang, Solution structures of 1:1 and 2:1 complexes of a fluorescent carbazole BMVC and *MYC* promoter G-quadruplex: Insights into conformation adjustment towards G4-recognition. *Nucleic Acids Res.* 47, 11931-11942 (2019) (*see abstracts*)
4. **Guanhui Wu**, Zheng Xing, Elizabeth J. Tran\* and Danzhou Yang\*, DDX5 helicase resolves G-quadruplex and is involved in *MYC* gene transcriptional activation. *Proc Natl Acad Sci U S A.* 116, 20453-20461 (2019)
5. Kai-Bo Wang, Mohamed S. A. Elsayed, **Guanhui Wu**, Nanjie Deng, Mark Cushman and Danzhou Yang, Indenoisoquinoline topoisomerase inhibitors strongly bind and stabilize the *MYC* promoter G-quadruplex and downregulate *MYC*. *J Am Chem Soc* 141, 11059-11070 (2019)
6. **Guanhui Wu**, Luying Chen, Wenting Liu and Danzhou Yang, Molecular recognition of the hybrid-type G-quadruplexes in human telomeres. *Molecules* 24 (2019)
7. **Guanhui Wu** and Haiyong Han, A DNA polymerase stop assay for characterization of G-quadruplex formation and identification of G-quadruplex-interactive compounds. *Methods Mol Biol* 2035, 223-231 (2019)
8. Buket Onel, **Guanhui Wu**, Daekyu Sun, Clement Lin and Danzhou Yang, Electrophoretic mobility shift assay and dimethyl sulfate footprinting for characterization of G-quadruplexes and G-quadruplex-protein complexes. *Methods Mol Biol* 2035, 201-222 (2019) (*see abstracts*)
9. Clement Lin, **Guanhui Wu**, Kai-Bo Wang, Buket Onel, Saburo Sakai, Yong Shao and Danzhou Yang, Molecular recognition of the hybrid-2 human telomeric G-quadruplex by epiberberine: insights into conversion of telomeric G-quadruplex structures. *Angew Chem Int Ed Engl* 57, 10888-10893 (2018) (*see abstracts*)
10. Robert V. Brown, Ting Wang, Venkateshwar Reddy Chappeta, **Guanhui Wu**, Buket Onel, Reena Chawla, Hector Quijada, Sara M. Camp Camp, Eddie T. Chiang, Quinea R. Lassiter, Carmen Lee, Shivani Phanse, Megan A. Turnidge, Ping Zhao, Joe G. N. Garcia, Vijay Gokhale, Danzhou Yang and Laurence H. Hurley, The consequences of overlapping G-quadruplexes and i-Motifs in the platelet-derived growth factor receptor beta core promoter

nuclease hypersensitive element can explain the unexpected effects of mutations and provide opportunities for selective targeting of both structures by small molecules to downregulate gene expression. *J Am Chem Soc* 139, 7456-7475 (2017) (*see abstracts*)

11. Buket Onel, Megan Carver, **Guanhui Wu**, Daria, Timonina, Salil Kalarn, Marti Larriva and Danzhou Yang, A new G-quadruplex with hairpin loop immediately upstream of the human *BCL2* P1 promoter modulates transcription. *J Am Chem Soc* 138, 2563-70 (2016) (*see abstracts*)

## PAPER ABSTRACTS

### **Custom DNA microarrays reveal diverse binding preferences of proteins and small molecules to thousands of G-quadruplexes**

Sreejana Ray, Desiree Tillo, Robert E. Boer, Nima Assad, Mira Barshai, **Guanhui Wu**, Yaron Orenstein, Danzhou Yang, John S. Schneekloth Jr., Charles Vinson,  
ACS Chem. Biol. 15, 4, 925-935 (2020)

Single-stranded DNA (ssDNA) containing four guanine repeats can form G-quadruplex (G4) structures. While cellular proteins and small molecules can bind G4s, it has been difficult to broadly assess their DNA-binding specificity. Here, we use custom DNA microarrays to examine the binding specificities of proteins, small molecules, and antibodies across ~15,000 potential G4 structures. Molecules used include fluorescently labeled pyridostatin (Cy5-PDS, a small molecule), BG4 (Cy5-BG4, a G4-specific antibody), and eight proteins (GST-tagged nucleolin, IGF2, CNBP, FANCI, PIF1, BLM, DHX36, and WRN). Cy5-PDS and Cy5-BG4 selectively bind sequences known to form G4s, confirming their formation on the microarrays. Cy5-PDS binding decreased when G4 formation was inhibited using lithium or when ssDNA features on the microarray were made double-stranded. Similar conditions inhibited the binding of all other molecules except for CNBP and PIF1. We report that proteins have different G4-binding preferences suggesting unique cellular functions. Finally, competition experiments are used to assess the binding specificity of an unlabeled small molecule, revealing the structural features in the G4 required to achieve selectivity. These data demonstrate that the microarray platform can be used to assess the binding preferences of molecules to G4s on a broad scale, helping to understand the properties that govern molecular recognition.

**PDGFR- $\beta$  promoter forms a vacancy G-quadruplex that can be filled in by dGMP: solution structure and molecular recognition of guanine metabolites and drugs**

Kai-Bo Wang, Jonathan Dickerhoff, **Guanhui Wu**, Danzhou Yang

*J. Am. Chem. Soc.* 142, 11, 5204-5211 (2020)

Aberrant expression of PDGFR- $\beta$  is associated with a number of diseases. The G-quadruplexes (G4s) formed in PDGFR- $\beta$  gene promoter are transcriptional modulators and amenable to small molecule targeting. The major G4 formed in the PDGFR- $\beta$  gene promoter was previously shown to have a broken G-strand. Herein, we report that the PDGFR- $\beta$  gene promoter sequence forms a vacancy G-quadruplex (vG4) which can be filled in and stabilized by physiologically relevant guanine metabolites, such as dGMP, GMP, and cGMP, as well as guanine-derivative drugs. We determined the NMR structure of the dGMP-fill-in PDGFR- $\beta$  vG4 in K<sup>+</sup> solution. This is the first structure of a guanine-metabolite-fill-in vG4 based on a human gene promoter sequence. Our structure and systematic analysis elucidate the contributions of Hoogsten hydrogen bonds, sugar, and phosphate moieties to the specific G-vacancy fill-in. Intriguingly, an equilibrium of 3'- and 5'-end vG4s is present in the PDGFR- $\beta$  promoter sequence, and dGMP favors the 5'-end fill-in. Guanine metabolites and drugs were tested and showed a conserved selectivity for the 5'-vacancy, except for cGMP. cGMP binds both the 3'- and 5'-end vG4s and forms two fill-in G4s with similar population. Significantly, guanine metabolites are involved in many physiological and pathological processes in human cells; thus, our results provide a structural basis to understand their potential regulatory functions by interaction with promoter vG4s. Moreover, the NMR structure can guide rational design of ligands that target the PDGFR- $\beta$  vG4.

## **Structures of 1:1 and 2:1 complexes of BMVC and *MYC* promoter G-quadruplex reveal a mechanism of ligand conformation adjustment for G4-recognition**

Wenting Liu, Clement Lin, **Guanhui Wu**, Jixun Dai, Ta-Chau Chang, Danzhou Yang

Nucleic Acids Res. 47, 11931-11942 (2019)

BMVC is the first fluorescent probe designed to detect G-quadruplexes (G4s) *in vivo*. The *MYC* oncogene promoter forms a G4 (MycG4) which acts as a transcription silencer. Here, we report the high-affinity and specific binding of BMVC to MycG4 with unusual slow-exchange rates on the NMR timescale. We also show that BMVC represses *MYC* in cancer cells. We determined the solution structures of the 1:1 and 2:1 BMVC - MycG4 complexes. BMVC first binds the 5'-end of MycG4 to form a 1:1 complex with a well-defined structure. At higher ratio, BMVC also binds the 3'-end to form a second complex. In both complexes, the crescent-shaped BMVC recruits a flanking DNA residue to form a BMVC-base plane stacking over the external G-tetrad. Remarkably, BMVC adjusts its conformation to a contracted form to match the G-tetrad for an optimal stacking interaction. This is the first structural example showing the importance of ligand conformational adjustment in G4 recognition. BMVC binds the more accessible 5'-end with higher affinity, whereas sequence specificity is present at the weaker-binding 3'-site. Our structures provide insights into specific recognition of MycG4 by BMVC and useful information for design of G4-targeted anticancer drugs and fluorescent probes.

## **Electrophoretic mobility shift assay and dimethyl sulfate footprinting for characterization of G-quadruplexes and G-quadruplex-protein complexes**

Buket Onel, **Guanhui Wu**, Daekyu Sun, Clement Lin, Danzhou Yang

*Methods Mol Biol* 2035, 201-222 (2019)

DNA G-quadruplexes are globular nucleic acid secondary structures which occur throughout the human genome under physiological conditions. There is accumulating evidence supporting G-quadruplex involvement in a number of important aspects of genome functions, including transcription, replication, and genomic stability, and that protein and enzyme recognition of G-quadruplexes may represent a key event to regulate physiological or pathological pathways. Two important techniques to study G-quadruplexes and their protein interactions are the electrophoretic mobility shift assay (EMSA) and dimethyl sulfate (DMS) footprinting assay. EMSA, one of the most sensitive and robust methods for studying the DNA-protein interactions, can be used to determine the binding parameters and relative affinities of a protein for the G-quadruplex. DMS footprinting is a powerful assay for the initial characterization of G-quadruplexes, which can be used to deduce the guanine bases involved in the formation of G-tetrads under physiological salt conditions. DMS footprinting can also reveal important information in G-quadruplex-protein complexes on protein contacts and regional changes in DNA G-quadruplex upon protein binding. In this paper, we will provide a detailed protocol for the EMSA and DMS footprinting assays for characterization of G-quadruplexes and G-quadruplex-protein complexes. Expected outcomes and references to extensions of the method will be further discussed.

**Molecular recognition of the hybrid-2 human telomeric G-quadruplex by epiberberine: insights into conversion of telomeric G-quadruplex structures**

Clement Lin, **Guanhui Wu**, Kai-Bo Wang, Buket Onel, Saburo Sakai, Yong Shao, Danzhou Yang  
*Angew Chem Int Ed Engl* 57, 10888-10893 (2018)

Human telomeres can form DNA G-quadruplex (G4), an attractive target for anticancer drugs. Human telomeric G4s bear inherent structure polymorphism, challenging for understanding specific recognition by ligands or proteins. Protoberberines are medicinal natural-products known to stabilize telomeric G4s and inhibit telomerase. Here we report epiberberine (EPI) specifically recognizes the hybrid-2 telomeric G4 predominant in physiologically relevant  $K^+$  solution and converts other telomeric G4 forms to hybrid-2, the first such example reported. Our NMR structure in  $K^+$  solution shows EPI binding induces extensive rearrangement of the previously disordered 5'-flanking and loop segments to form an unprecedented four-layer binding pocket specific to the hybrid-2 telomeric G4; EPI recruits the (-1) adenine to form a "quasi-triad" intercalated between the external tetrad and a T:T:A triad, capped by a T:T base pair. Our study provides structural basis for small-molecule drug design targeting the human telomeric G4.

**The consequences of overlapping G-quadruplexes and i-Motifs in the platelet-derived growth factor receptor beta core promoter nuclease hypersensitive element can explain the unexpected effects of mutations and provide opportunities for selective targeting of both structures by small molecules to downregulate gene expression**

Robert V. Brown, Ting Wang, Venkateshwar Reddy Chappeta, **Guanhui Wu**, Buket Onel, Reena Chawla, Hector Quijada, Sara M. Camp Camp, Eddie T. Chiang, Quinea R. Lassiter, Carmen Lee, Shivani Phanse, Megan A. Turnidge, Ping Zhao, Joe G. N. Garcia, Vijay Gokhale, Danzhou Yang, Laurence H. Hurley

J Am Chem Soc 139, 7456-7475 (2017)

The platelet-derived growth factor receptor  $\beta$  (*PDGFR- $\beta$* ) signaling pathway is a validated and important target for the treatment of certain malignant and nonmalignant pathologies. We previously identified a G-quadruplex-forming nuclease hypersensitive element (NHE) in the human *PDGFR- $\beta$*  promoter that putatively forms four overlapping G-quadruplexes. Therefore, we further investigated the structures and biological roles of the G-quadruplexes and i-motifs in the *PDGFR- $\beta$*  NHE with the ultimate goal of demonstrating an alternate and effective strategy for molecularly targeting the *PDGFR- $\beta$*  pathway. Significantly, we show that the primary G-quadruplex receptor for repression of *PDGFR- $\beta$*  is the 3'-end G-quadruplex, which has a GGA sequence at the 3'-end. Mutation studies using luciferase reporter plasmids highlight a novel set of G-quadruplex point mutations, some of which seem to provide conflicting results on effects on gene expression, prompting further investigation into the effect of these mutations on the i-motif-forming strand. Herein we characterize the formation of an equilibrium between at least two different i-motifs from the cytosine-rich (C-rich) sequence of the *PDGFR- $\beta$*  NHE. The apparently conflicting mutation results can be rationalized if we take into account the single base point mutation made in a critical cytosine run in the *PDGFR- $\beta$*  NHE that dramatically affects the equilibrium of i-motifs formed from this sequence. We identified a group of ellipticines that targets the G-quadruplexes in the *PDGFR- $\beta$*  promoter, and from this series of compounds, we selected the ellipticine analog GSA1129, which selectively targets the 3'-end G-quadruplex, to shift the dynamic equilibrium in the full-length sequence to favor this structure. We also identified a benzothiophene-2-carboxamide (NSC309874) as a *PDGFR- $\beta$*  i-motif-interactive compound. In vitro, GSA1129 and NSC309874 downregulate *PDGFR- $\beta$*  promoter activity and transcript in the

neuroblastoma cell line SK-N-SH at subcytotoxic cell concentrations. GSA1129 also inhibits *PDGFR- $\beta$* -driven cell proliferation and migration. With an established preclinical murine model of acute lung injury, we demonstrate that GSA1129 attenuates endotoxin-mediated acute lung inflammation. Our studies underscore the importance of considering the effects of point mutations on structure formation from the G- and C-rich sequences and provide further evidence for the involvement of both strands and associated structures in the control of gene expression.

## **A new G-quadruplex with hairpin loop immediately upstream of the human *BCL2* P1 promoter modulates transcription**

Buket Onel, Megan Carver, **Guanhui Wu**, Daria, Timonina, Salil Kalarn, Marti Larriva, Danzhou Yang

J Am Chem Soc 138, 2563-70 (2016)

The abnormal overexpression of the *BCL2* gene is associated with many human tumors. We found a new 28-mer G-quadruplex-forming sequence, P1G4, immediately upstream of the human *BCL2* gene P1 promoter. The P1G4 is shown to be a transcription repressor using a promoter-driven luciferase assay; its inhibitory effect can be markedly enhanced by the G-quadruplex-interactive compound TMPyP4. G-quadruplex can readily form in the P1G4 sequence under physiological salt condition as shown by DMS footprinting. P1G4 and previously identified Pu39 G-quadruplexes appear to form independently in adjacent regions in the *BCL2* P1 promoter. In the extended *BCL2* P1 promoter region containing both Pu39 and P1G4, P1G4 appears to play a more dominant role in repressing the transcriptional activity. Using NMR spectroscopy, the P1G4 G-quadruplex appears to be a novel dynamic equilibrium of two parallel structures, one regular with two 1-nt loops and a 12-nt middle loop and another broken-strand with three 1-nt loops and a 11-nt middle loop; both structures adopt a novel hairpin (stem-loop duplex) conformation in the long loop. The dynamic equilibrium of two closely related structures and the unique hairpin loop conformation are specific to the P1G4 sequence and distinguish the P1G4 quadruplex from other parallel structures. The presence of P1G4 and Pu39 in adjacent regions of the *BCL2* P1 promoter suggests a mechanism for precise regulation of *BCL2* gene transcription. The unique P1G4 G-quadruplex may provide a specific target for small molecules to modulate *BCL2* gene transcription.