EFFECT OF DNA G-QUADRUPLEX STRUCTURES (G4) ON PROKARYOTIC GENE EXPRESSION

EROKAY GÜRZ

Thesis for the Master's Program in Bioengineering

Graduate School

Izmir University of Economics

Izmir

2023

EFFECT OF DNA G-QUADRUPLEX STRUCTURES (G4) ON PROKARYOTIC GENE EXPRESSION

EROKAY GÜRZ

THESIS ADVISOR: ASSOC. PROF. DR. OSMAN DOLUCA

A Master's Thesis

Submitted to

the Graduate School of Izmir University of Economics

the Department of Bioengineering

Izmir

2023

ETHICAL DECLARATION

I hereby declare that I am the sole author of this thesis and that I have conducted my work in accordance with academic rules and ethical behaviour at every stage from the planning of the thesis to its defence. I confirm that I have cited all ideas, information and findings that are not specific to my study, as required by the code of ethical behaviour, and that all statements not cited are my own.

Name, Surname: Erokay Gürz

Date: 20/07/2023

ABSTRACT

EFFECT OF DNA G-QUADRUPLEX STRUCTURES (G4) ON PROKARYOTIC GENE EXPRESSION

Gürz**,** Erokay

Master's Program in Bioengineering

Advisor: Assoc. Prof. Dr. Osman Doluca

July, 2023

DNA G-quadruplexes (G4s) are 4-stranded secondary DNA structures that has a motif containing four consecutive tracts of guanine spaced by several nucleotides and form four stacks of guanine tetrads bond by Hoogsteen base pairing. They are found in many organisms at significant genomic regions which also have been investigated extensively for their molecular structure and functional role in cell. One G4 motif was identified by our group in *E. coli* genome to be capable of forming both G4 and hairpin structures, able to switch between those, and hypothesized to have regulatory role on gene expression. We aimed to investigate the role of that G4 on gene regulation since the functional influences of gene expression may be too variable for specific G4 structures. We have inserted variants of the G4 into promoter region of GFP reporter gene construct and transformed the vector into *E. coli* and measure gene expression through florescence spectroscopy. We also have treated the cells with ligands LiCl and TMPyP4 known to stabilize the G4 structure to investigate the effects on gene expression more comprehensively. We found that G4 structure is more persistent than

hairpin in the regulatory switch mechanism affecting on gene expression. LiCl treatment have not change the overall expression pattern and TMPyP4 treatment yielded an inconsistent pattern among variant groups that could be related to its destabilizing property or interaction with sequence specific unpredicted DNA secondary structure. These results are giving insight about regulation of gene expression mechanism in *E. coli* by these switchable G4s.

Keywords: G quadruplex, Hairpin, Gene Expression, Green Fluorescent Protein, Reporter Gene, *E. coli*.

ÖZET

DNA G DÖRTLÜLERİNİN (G4) PROKARYOTİK GEN IFADESİ ÜZERİNE ETKİSİ

Gürz**,** Erokay

Biyomühendislik Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Osman Doluca

Temmuz, 2023

DNA G-dörtlüleri (G4s), birkaç nükleotid aralıklı dört ardışık guanin tract'i içeren bir motife sahip olan ve Hoogsteen bağı ile bağlı üst üste dört guanin tetrad oluşturan 4 iplikli ikincil DNA yapılarıdır. G-dörtlüleri, birçok organizmada önemli genomik bölgelerde bulunurlar ve moleküler yapıları ve hücredeki işlevsel rolleri açısından kapsamlı olarak araştırılmışlardır. Grubumuz tarafından *E. coli* genomunda hem G4 hem de hairpin yapıları oluşturabilen, bunlar arasında geçiş yapabilen ve gen ifadesi üzerinde düzenleyici rolü olduğu varsayılan bir G4 motifi tanımlanmıştır. Gen ifadesinin işlevsel etkileri belirli G4 yapıları için çok değişken olabileceğinden, bu G4'ün gen düzenlemesi üzerindeki rolünü araştırmayı amaçladık. GFP raportör gen yapısının promotör bölgesine G4'ün varyantlarını ekledik, vektörü *E. coli*'ye transforme ettik ve floresan spektroskopisi ile gen ifadesini ölçtük. Ayrıca, gen ekspresyonu üzerindeki etkilerini daha kapsamlı bir şekilde araştırmak için hücreleri G4 yapısını stabilize ettiği bilinen ligandlar olan LiCl ve TMPyP4 ile muamele ettik. Gen ekspresyonunu etkileyen değişebilen anahtar mekanizmasında G4 yapısının

hairpinden daha etkili olduğunu bulduk. Varyant grupları arasında, LiCl uygulaması genel ekspresyon modelini değiştirmedi ve TMPyP4 uygulaması, zayıflatma özelliği veya diziye özgü öngörülemeyen DNA ikincil yapısı etkileşimi ile ilgili olabilecek tutarsız bir model verdi. Bu sonuçlar, bu değişebilen G4'ler tarafından *E. coli*'deki gen ekspresyon mekanizmasının düzenlenmesinin anlaşılmasına olanak tanımaktadır.

Anahtar Kelimeler: G-dörtlüsü, Hairpin, Gen Ekspresyonu, Yeşil Floresan Protein, Haberci Gen, *E. coli*

ACKNOWLEDGEMENTS

I'd like to give my regards and appreciation to my advisor Assoc. Prof. Osman Doluca for his indulgent approach to my enthusiastic attitude and his guidance to channel this enthusiasm into a more efficient work which have resulted with this study.

I would also like to give my regards to Prof. Dr. Eda Tahir Turanlı and Asst. Prof. Dr. Ayşegül Kaymak Özdemir for honoring me by agreeing to be jury members for the defence of my thesis and Assoc. Prof. Zeynep Fırtına Karagonlar for accepting my thesis and Assoc. Prof. Mine Güngörmüşler for guiding me through many obstacles throughout my postgraduate studies.

In addition, I would like to thank my lab mate Res. Asst. Hüseyin Saygın Portakal for his support and biggest contributions to this study and as a valuable friend I gained along this journey.

I also thank my family for their financial and moral support that allowed me to get to this point in my career since I was a little boy interested in science.

And last but not least, I want to thank me for believing in me, doing all this hard work, never quitting and being me at all times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALS: Amyotrophic lateral sclerosis

ANOVA: Analysis of Variance

FTD: Frontotemporal Degeneration

CRISPR-Cas9: Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR-associated protein 9

DNA: Deoxyribonucleic Acid

FIS: Factor for Inversion Stimulation

G4: G quadruplex

GFP: Green Fluorescent Protein

GQFS: G-quadruplex-forming Sequences

HDAC : Histone Deacetylase

hTERT: Human Telomerase Reverse Transcriptase

i-motif: Intercalated Motif

Lrp: Leucine-responsive Regulatory Protein

NHaA: Na(+):H(+) antiporter A

NHaB: Na(+):H(+) antiporter B

NM23-H2: Nucleoside diphosphate kinase B

NMM: N-methyl mesoporphyrin IX

ORF: Open Reading Frame

p21: cyclin-dependent kinase inhibitor 1

pBAD: Arabinose Inducible Promoter

PCR: Polymerase Chain Reaction

PNK: Polynucleotide Kinase

RE: Restriction Enzyme

REST: RE1 Silencing Transcription Factor

RNA: Ribonucleic Acid

RNAP RNA polymerase

SD: Shine-Dalgarno

sfGFP: Super Folder Green Fluorescent Protein

TMPyP4: Tetra-(N-methyl-4-pyridyl)porphine)

TNG Buffer: Tris NaCl Glycerol Buffer

TRF: Telomere Binding Protein

TSS: Transcription Start Sites

σ70: RNA Polymerase Sigma Factor 70

CHAPTER 1: INTRODUCTION

1.1. DNA G4 Structures

Guanine rich sequences on DNA can fold into four stranded structures called G quadruplex (G4) composed of multiple G-quartets (also known as G-tetrads), formed by *Hoogsteen hydrogen bonds* with a square-planar geometry between guanine bases (Zimmerman, Cohen and Davies, 1975; Sengupta, Roy and Chowdhury, 2021; Yadav et al., 2021). Those guanine rich sequences have to be in a specific repetitive motif to form G4 structures, called G-tracts. General motif of G-tracts can be described as "$G_{\geq 3}N_xG_{\geq 3}N_yG_{\geq 3}N_zG_{\geq 3}$", where there should be at least four repetition of sequential guanines makes the four corners of the square-planar geometry of G-quartets, and number of guanines in each repetition or G-tract determines the number of G-quartets. There should be at least 3 guanines in a G-tract in order for it to be stable. G-tracts forms stem and the filler nucleotides between each tract forms the loops overhanging from the G-quartets (Figure 1). Size of the loops determines the stability of G4 structure, as smaller loops 1-7 nucleotides in size forms more stable G quadruplexes. Another factor that determines the stability in addition to the size of the loops and the G-tracts is the monovalent cation concentration in the environment which compensates the electrostatic repulsion force of the oxygen atoms of the guanines in G-quartets. Stability of the G4 is also affected from strand stoichiometry and alignment (Williamson, Raghuraman and Cech, 1989; Hardin, Perry and White, 2001; Bochman, Paeschke and Zakian, 2012; Yadav et al., 2021).

G quadruplexes are formed from a motif containing 4 sequential G-tracts forming stems with loop sequences in between. As those motifs are on DNA, complementary strand contains 4 sequential C-tracts consisting of consecutive cytosine nucleotides with complementary of loop sequences in between. That complementary motif to the G4 forms another motif with different topology called i-motif also called as intercalated motif. DNA i-motif structures formed by intercalated Hoogsteen base pairs of hemi-protonated cytosines stacked onto each other (Brown and Kendrick, 2021).

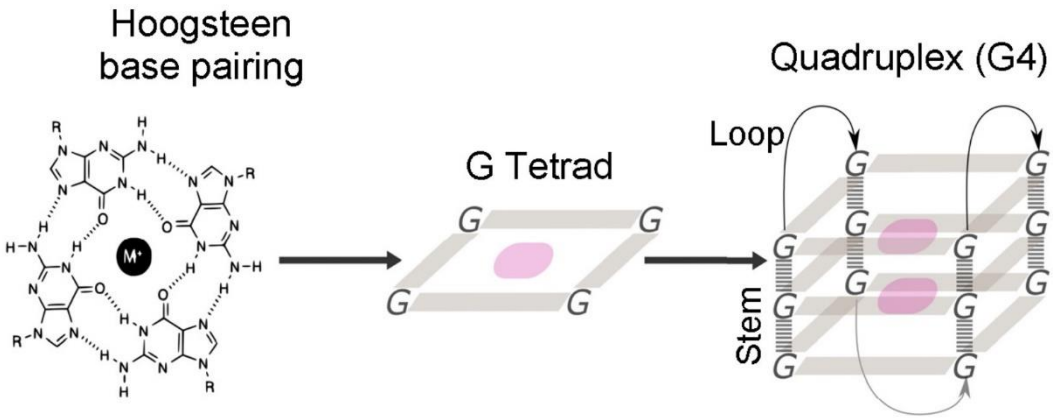

Figure 1. DNA G4 (G-quadruplex) structure. Molecular structure of a G-tetrad demonstrating Hoogsteen bonds between Guanine bases and a common DNA G-quadruplex formed by stacking of 3 G-tetrads. G-tracts formed by consecutive guanines in a motif forms the backbone of G4 structures called stem, and any nucleotide between G-tracts forms loops (Source: Sengupta, Roy and Chowdhury, 2021).

1.2. Types of G4

G4s can be classified according to G-tract size, loop size, and number, orientation and type of nucleic acid strands that joins the structure. G-tract size determines the number of stacks of G-tetrads in the G-quadruplex and for 2, 3 and 4 consecutive guanines in G-tract, G-quadruplex is named as G2, G3 and G4 respectively. Number of nucleic acid strands forming the G4 determines whether the G4 is intramolecular (formed from single strand) or intermolecular (formed with incorporation of multiple strands). Orientation of those strands according to each other determines whether its parallel in which all the strands are in same orientation, anti-parallel conformation which 2 of the strands are reverse orientation to the other 2, or hybrid conformation which 1 of the strands is reverse orientation to the other 3 (Figure 2) (Burge et al., 2006; Biver, 2022). Finally, G-quadruplexes can be categorized by type of nucleic acid strand as it could be RNA, DNA or hybrid of RNA-DNA (Bochman, Paeschke and Zakian, 2012; Varshney et al., 2020; Yadav et al., 2021).

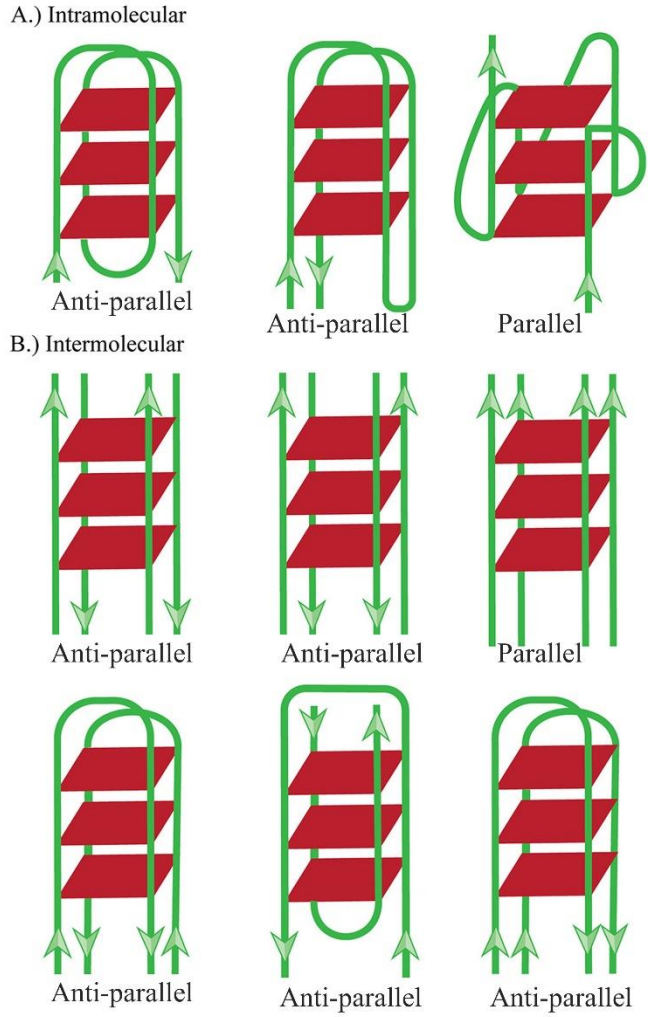

Figure 2. G4 (G-quadruplex) Types. G-quadruplexes formed in different topologies. Intramolecular G4s are formed by a single DNA molecule (A) while intermolecular G4s are formed by 2 or more DNA molecules (B). The orientation of the DNA strands determines whether the G4 is parallel (all strands in same direction), anti-parallel (2 of the strands are opposite to the other 2) or hybrid (1 strand is in reverse direction to the other 3). Strand direction indicated by green arrows. G-tetrads indicated by red planes. (Source: V. Yadav et al., 2017).

1.3. Genomic Location and Biological Roles

One of the common methods to analyse G4 DNA is by *in-silico* algorithms like G4Catchall (Doluca, 2019), Quadparser (Huppert and Balasubramanian, 2005), GRS mapper (D'Antonio and Bagga, 2004), G4 Hunter (Bedrat, Lacroix and Mergny, 2016), PQS-finder (Hon et al., 2017), and QuadBase (Dhapola and Chowdhury, 2016; Sengupta, Roy and Chowdhury, 2021). DNA motifs that can form DNA G-quadruplex

structures can be detected by such algorithms are called G-quadruplex-forming sequences (GQFS). GQFS can be found in genomes of wide variety of species in domains of eukarya and prokarya. G4 motifs are conserved evolutionary and the distribution of GQFS genomes of those species were discovered to be non-random and were enriched in certain genomic regions that have important function in cellular machinery such as promoters, transcription start sites (TSS), open reading frames (ORF), telomeres and mRNAs which indicates the biological role of G4s related to those regions, especially on gene regulation (Bochman, Paeschke and Zakian, 2012; Yadav et al., 2021).

An *In-silico* analysis study on several mammalian genomes revealed that G4 motifs were associated with transcription factor binding sites (TFBS) and this association was conserved. 9 transcription factor out of 220 was significantly related within 100bp flanking G4 motifs and 7 TF of those 9 TF were zinc-finger binding proteins which could provide an insight into the mechanism of G4s on gene expression (Kumar et al., 2011). This G4-TF association was also observed in *E. coli* and the promoters that G4 motifs present were of genes that are expressed during exponential growth phase (Sengupta, Roy and Chowdhury, 2021). In eukaryotes GQFS are found in promoter regions of protooncogenes such as *Bcl-2, c-myc, c-myb, c-kit, VEGF, KRAS* and *HIF-1*, indicating G4s role in carcinogenesis (Balasubramanian, Hurley and Neidle, 2011). In plant genomes such as *Arabidopsis thaliana, Zea mays, Oryza japonicum,* and *Oryza sativa* GQFS were identified upstream and downstream of TSS (transcription start site) indicating G quadruplexes' regulatory role on gene regulation (Mullen et al., 2010; Andorf et al., 2014; Yadav et al., 2017). In a study in human cell lines, in-silico analysis showed that presence of G4 motifs are negatively associated with CpG DNA methylation, in-vitro analysis on >600,000 CpGs from 18 individuals was supporting the in-silico finding. This inverse correlation between G4s and CpG methylation is indicating role of G4s in CpG DNA methylation (Halder et al., 2010). Other studies revealed that this association is through DNA methyltransferase enzymes, DNA methyltransferase DNMT1 was found to have high affinity to G4 DNA in vitro, and binding of G4 DNA to DNMT1 is decreasing enzymes methylation activity causing hypomethylated CpG islands (Cree et al., 2016; Mao et al., 2018). CpG methylation on G4 motifs were associated with neurodegenerative disorders like ALS and FTD (Zamiri et al., 2015; Sengupta, Roy and Chowdhury, 2021). These studies

cumulatively emphasizing the role of G4s in epigenetic gene regulation.

In prokaryotes, *in-silico* analyses have uncovered similar outcomes within eukaryotes as GQFS were found upstream of coding regions and in promoter regions of genes that have role in transcription and signal transduction. GQFS have been identified in the promoter regions of regulatory genes of *Escherichia coli* that are essential for transcription like RNA polymerase sigma factor 70 (σ^{70}), FIS and Lrp (leucine-responsive regulatory protein) upon analysing for binding sites of 55 DNA-binding protein (Rawal et al., 2006). In *Deinococcus radiodurans,* GQFS were found in promoter regions of genes involving DNA repair pathway: *recA, recF, recO, recR* and *andrecQ,* thus suggesting their contribution to the radiation resistance of the bacteria (Beaume et al., 2013). GQFS have also been identified in *Paracoccus denitrificans,* at the promoter region of nas (assimilatory nitrate/nitrite reductase system) genes by analysing with Quadparser algorithm (Waller et al., 2016). In *Mycobacterium tuberculosis*, promoter regions of genes *zwf1* (glucose-6-phosphate dehydrogenase 1), *clpx* (ATP-dependent Clp protease), *mosR* (oxidation-sensing regulator transcription factor), *andndhA* (membrane NADH dehydrogenase) have also been detected for GQFS by an in-house computational tool (Perrone et al., 2017).

With accumulated knowledge on G quadruplexes putative role on gene regulation, functional studies have been performed by mostly utilizing G4 binding ligands that stabilizes G4s such as TMPyP4 (Tetra-(N-methyl-4-pyridyl)porphine) (Wheelhouse et al., 1998; Arora and Maiti, 2008) and NMM (Yett et al., 2019) to unravel the epigenetic mechanisms. Studies have been found that gene regulation by G4s occurring at transcriptional and translational level. At transcriptional level, gene regulation was determined to be dependent on DNA orientation and strand (Holder and Hartig, 2014). In eukaryotes, DNA strand orientation and position of G4 motif with respect to functional DNA elements was observed to have differential regulatory effects on transcription and translation (Biffi et al., 2013). However, due to major differences between eukaryotic and prokaryotic transcriptional and translational machinery, outcomes of orientation and position cannot be expected in the same manner in prokaryotes as it is in eukaryotes. In a study in *E. coli* where GQFS were inserted into promoter, 5', and 3' UTR regions of a reporter gene; it was found out that DNA G4 structures on promoter and ribosomal binding regions were caused inhibition of the

reporter gene expression (Holder and Hartig, 2014). In bacteria, gene expression is reduced by downregulation of transcription due to G-rich sequences present on the antisense strand of promoter. Sigma 70 factor (σ70) joins to the RNA polymerase (RNAP) holoenzyme in E. coli and has a function of promoter recognition and promoter melting. σ70 binds -10 to -35 region of promoter and initiate transcription to produce a 10 nt long transcript, then it leaves the enzyme complex. Presence of a G4 in the antisense strand of promoter region prevents σ70 from binding to DNA and subsequently inhibits transcription. On the contrary, G4 in the downstream of promoter region has the potential to increase transcription by stabilizing the antisense strand in single strand conformation after RNAP separates the DNA strands (Feklistov, 2013). G-quadruplexes can play role in stress conditions such as osmotic shock and general stress responses, since in these conditions intracellular K^+ levels increase which may induce G4 formation. In a consistent manner, GQFS have been identified in 5' UTR regions of stress genes like *oxyR, relA, rseA* (Huo, Rosenthal and Gralla, 2008).

G4 DNA at the coding strand of Shine-Dalgarno (SD) regions of ORF putatively inhibits translation initiation by preventing ribosome from recognizing and binding to mRNA transcripts. In a study in *E. coli*, GQFS were inserted adjacent to SD region of a reporter gene and inhibition of protein expression was observed and level of inhibition is related to the thermodynamic stability of G4 (Wieland and Hartig, 2007; Endoh, Kawasaki and Sugimoto, 2013). In another reporter gene study in *E. coli*, GQFS inserted nearby of slippery sequence in ORF caused -1 ribosomal frameshift and efficiency of frameshift was increased with addition of G4-binding molecule berberine (Endoh and Sugimoto, 2013).

1.4. Prior Study

Prior to investigating the effects of DNA G4 structures on promoter region of a reporter gene construct, our research group has discovered a wild-type G4 structures that was identified to be formed in a regulatory switch mechanism between duplex, hairpin, and G4 DNA in *E. coli* genome (Kaplan et al., 2016). Then, we performed a study investigating the functional effects of those wild-type G4 structures. We initially have attempted to knock out the wild-type G4 motifs by site directed mutagenesis with CRISPR-Cas9 system to observe the changes in gene expression and overall physiological conditions however we could not achieve the knockouts, signifying the

need for developing CRISPR-Cas9 system for G4 motif knock-out. Then we implemented a reporter gene construct by inserting G4 motifs to several positions and orientations around pBAD promoter regulating the expression of a GFP gene. As a result, we have discovered significant reduction in the gene expression, with the most occurring in the plasmid having a G4 inserted downstream of the promoter on sense strand.

In this study, same reporter gene construct, plasmid coding for GFP gene under regulation of arabinose promoter araBAD was used. This promoter requires arabinose in the environment for the gene downstream of it to be expressed, allowing its inducible expression (Khlebnikov, Skaug and Keasling, 2002). Since the largest change observed in the G4 motif inserted downstream of the promoter on sense strand in prior study, 3 more G-quadruplex variants have been designed to analyse the effects of different features in the G-quadruplex structure and establish the relationship between *in-silico* motif sequence and the function of *in-vivo* secondary structure of DNA G-quadruplex on prokaryotic gene expression.

CHAPTER 2: METHODS

2.1. Insertion of G4 Sequences into Plasmid sfGFP-PBAD

Site directed mutagenesis by PCR performed to insert the designed 5 G4 motifs downstream of the araBAD promoter region of GFP coding plasmid sfGFP-pBAD (Addgene plasmid # 54519) (Pédelacq et al., 2005), at nucleotide position 1742 on plasmid. Site directed mutagenesis was resulted with 5 groups: WT, G4(+), hp(-), G4(-), G4(++).

WT group has sfGFP-pBAD plasmid with no insert thus bearing the native form of the plasmid. G4(+) group has sfGFP-pBAD plasmid inserted with a motif that can form either a G3 quadruplex or a cruciform hairpin structure which could reversibly switch between each other (Figure 5(A)). hp(-) group has sfGFP-pBAD plasmid inserted with G3 quadruplex motif with no cruciform hairpin forming motif (Figure 5(D)). G4(-) group has sfGFP-pBAD plasmid inserted with a motif of no G4 but cruciform hairpin forming motif only (Figure 5(C)). G4(++) group has sfGFP-pBAD plasmid inserted with potentially more stable G4 quadruplex motif that can also form a potentially more stable cruciform hairpin which could reversibly switch between each other (Figure 5(B)). Insertion of G4 motifs were performed via site directed mutagenesis by PCR resulting linearized plasmids which were recircularized by ligation following RE digestion and phosphorylation.

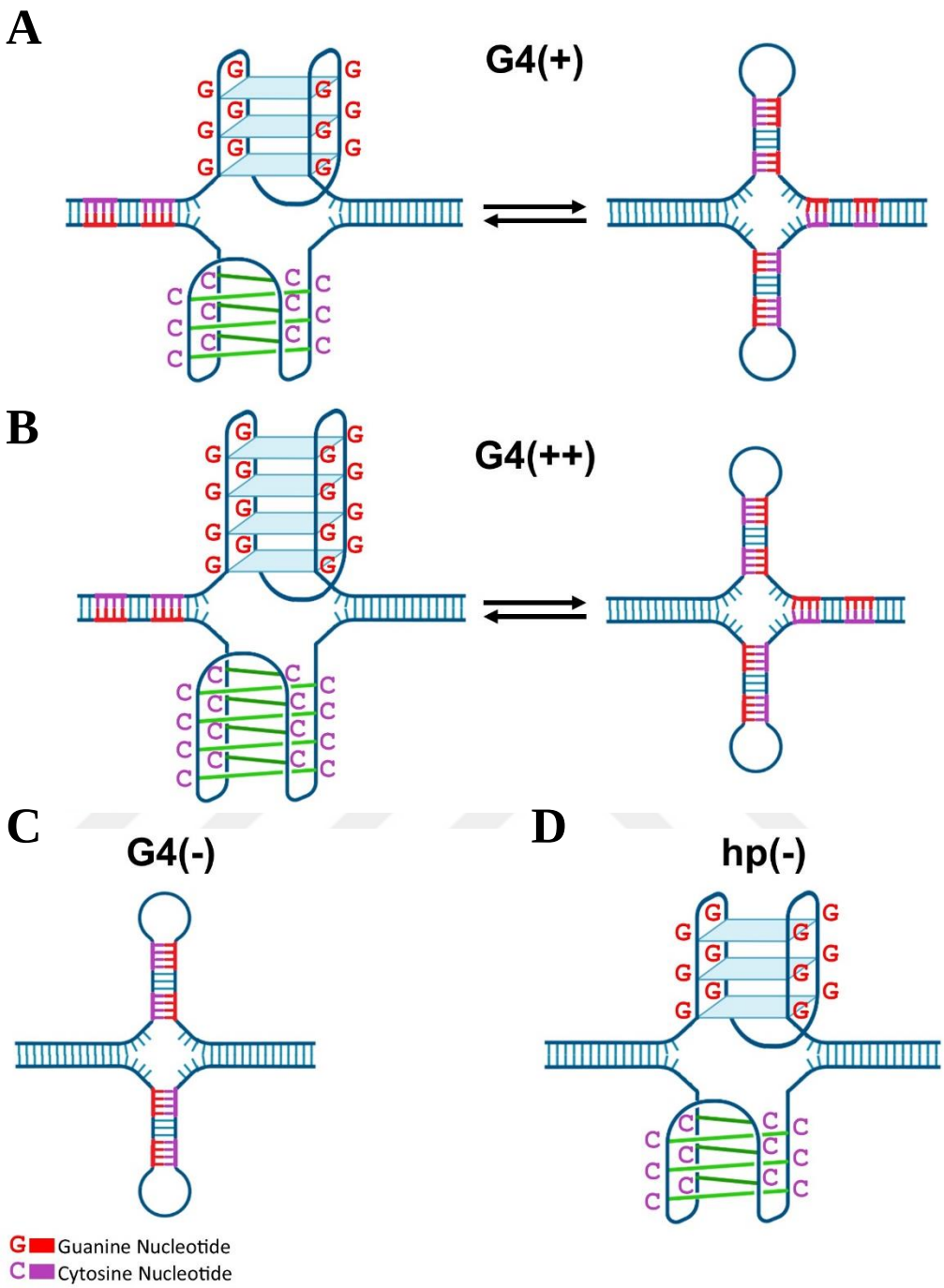

Figure 3. Estimated DNA Structures of Designed G4 Motifs. Structure of the DNA G4s that could be formed from the G4 motifs used in this study. Red nucleotides represent guanines constitutes G4; purple nucleotides represent cytosines constitutes i-motif which are complementary to the guanines forming G4; and blue nucleotides represent any nucleotide. (A) Variant G4(+): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif. (B) Variant G4(++): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif. (C) G4(-): *E. coli* transformed with plasmid inserted with only cruciform hairpin

forming motif. (D) hp(-): *E. coli* transformed with plasmid inserted with only G4 forming motif.

Site directed mutagenesis PCR primers with 5' polynucleotide tails containing G4 motif variants were designed (see primer sequences in Table 1) and synthesized (Sentebiolab, Ankara, Turkey). G4 motifs were included in reverse primers and a common forward primer was used in all site directed mutagenesis applications. Primers consist of 4 regions: a priming region complementary to the template sequence of insertion site on the plasmid, G4 motif variant sequence to be inserted, SpeI restriction site and additional nucleotides to maintain the stability of oligonucleotides. Last three regions are not hybridizing with the template sequence of plasmid as they are on 5' tail of primer. Nucleotide position of the priming regions on the plasmid are 1742-1720 for reverse primers containing motifs, and 1743-1775 for the forward primer (corresponding nucleotide of 5' end and 3' end of primer respectively).

Table 1. Primers used in site directed mutagenesis for insertion of G4 motif variants in plasmid sfGFP-pBAD. Green-coloured sequences indicate priming regions, orange-coloured sequences indicate restriction enzyme digestion sites by SpeI, blue-coloured sequences contain G4 motif variants with G-tracts underlined and black-coloured sequences are added for stability of the oligonucleotides. Sequences given in 5' to 3' direction.

Primers	**Sequence (5'→3')**
G4(+) (Reverse)	CTGAACTAGTCCCTCACCCTAACCCTCTCCCCAAAGGGGCGAGGGGACCCAAAAAAACGGGTATGGAGAA
Hp(-) (Reverse)	CTGAACTAGTCCCTCACCCTAACCCTCTCCCCAAAACCCAAAAAAACGGGTATGGAGAA
G4(-) (Reverse)	CTGAACTAGTCACTCACACTAACACTCTCACCAAAGGGGCGAGGGGACCCAAAAAAACGGGTATGGAGAA
G4(++) (Reverse)	CTGAACTAGTCCCCTCACCCCTAACCCCTCTCCCCAAAGGGGCGAGGGGACCCAAAAAAACGGGTATGGAGAA
Forward	GGTAACTAGTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG

Site directed mutagenesis PCR were setup for each 5 variant of G4 motif with Phusion High-Fidelity PCR Kit in 0.2 mL PCR tubes (Thermo Fisher Scientific, Waltham, MA, USA), master mix containing: Buffer 1X, dNTP (each) 0.2 mM, plasmid sfGFP-pBAD (Addgene, Watertown, MA, USA) 1.5 ng/µL, primers specific to variant group (each) 1 µM, DNA Polymerase 0.02 U/µL. 20 µL of reaction mixes were transferred in 0.2 mL PCR tubes and placed in thermal cycler TurboCycler 2 (Blue-ray Biotech, New Taipei City, Taiwan) and PCR was run with following conditions: initial denaturation at 98 ℃ for 30 seconds, 30 cycles of: denaturation at 98 ℃ for 9 seconds, annealing at 60 ℃ for 29 seconds, extension at 72 ℃ for 90 seconds and final extension at 72 ℃ for 10 minutes. After the PCR, amplified plasmids with inserted G4 motifs were confirmed by 1% agarose gel electrophoresis by Owl™ EasyCast™ B1A (Thermo Fisher Scientific, Waltham, MA, USA) and EV3020 (Consort, Turnhout, Belgium). Agarose gel was visualised with UV imager (Vilber, Collégien, France).

2.2. Recircularization of Linearized Plasmids

Linearized plasmids as a result of PCR amplification were recircularized by restriction enzyme (RE) digestion, phosphorylation of 5' ends and ligation. SpeI restriction digestion was made by FastDigest BcuI (Thermo Fisher Scientific, Waltham, MA, USA) to digest the sites on primers inserted. Restriction digestion reaction was set up following manufacturer's instructions with slight modifications: reaction mix containing 2 µL of Fast Digest Green Buffer, 1 µL of Fast Digest Enzyme and 17 µL of plasmid DNA were incubated at 37℃ for 5 min with Thermo Shaker Incubator MSC-100 (Allsheng, Hangzhou, Zhejiang, China). 1% agarose gel electrophoresis followed by gel digestion by Gel Extraction Kit (Omega Biotek, Norcross, GA, USA) was performed to purify the linearized RE digested plasmids.

Then, the plasmid DNA was phosphorylated for proper ligation at the 5' overhangs result of restriction digestion. Phosphorylation reaction was set up following manufacturer's instructions (New England Biolabs, Ipswich, MA, USA) with slight modifications: 5 µL of T4 PNK reaction buffer, 1 µL of T4 PNK 200 U/mL, 5 µL of ATP and 39 µL of plasmid DNA eluted from gel digestion. Reactions were incubated at 37℃ for 30 min for reaction and then 65℃ for 20 min for heat-inactivation of the polynucleotide kinase (PNK) enzyme with Thermo Shaker Incubator MSC-100 (Allsheng, Hangzhou, Zhejiang, China).

Finally, linearized RE digested and phosphorylated plasmid DNA were recircularized by T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA) following manufacturer's instructions with slight modifications: reaction mix containing 2 µL of T4 DNA Ligase Buffer, 1 µL of T4 DNA Ligase and 17 µL of plasmid DNA was incubated at room temperature overnight for ligation and at 65℃ for 20 min for heat-inactivation of T4 ligase with Thermo Shaker Incubator MSC-100 (Allsheng, Hangzhou, Zhejiang, China).

2.3. Transformation of E. coli DH5α strain with Plasmids

Plasmids with various G4 motif inserts were transformed by heat shock into competent *E. coli DH5α* strain cells to ensure the stability of the plasmids while increasing their number. Isolated transformant plasmids were verified for their integrity with sanger sequencing before transforming into *E. coli BL21* strain cells to perform expression study.

1 mL aliquot of *E. coli DH5α* stock cells were thawed and inoculated into sterile autoclaved (ST-50G (Jeio Tech, Billerica, MA, USA)) 100 mL LB broth (Neogen, Lansing, MI, USA) and incubated at 38 ℃ with 220 rpm shaking for 16 hours by BD56 (Binder, Tuttlingen, Germany) for cells to reach stationary phase to have maximum density by plotted cell curve at these conditions (data not shown). Passage of 2 mL of cell culture in stationary phase was performed into fresh 100 mL LB broth (1:50 (v/v) inoculation ratio) and incubated at 37 ℃ with 220 rpm shaking for 3 hours with incubator BD56 (Binder, Tuttlingen, Germany) for cells to reach exponential phase to have maximum viability for competent preparation by plotted cell curve at these conditions (data not shown).

E. coli DH5α chemically competent cells were prepared by $CaCl_2$ method. First, cell culture at exponential phase was aliquoted into 10 mL volumes and incubated at 0 ℃ on ice for 20 minutes to keep the cells cold for whole competency method, then centrifuged at 4 ℃, 4000 rpm for 10 minutes to pellet the cells and supernatant was discarded. Cell pellet was resuspended with 10 mL of sterile cold 0.1 M CaCl2 and incubated at 0 ℃ on ice for 30 minutes and cell pellet obtained by centrifugating at 4 ℃, 4000 rpm for 10 minutes and discarding the supernatant. 10 mL of sterile cold 0.1

M CaCl2 was placed on to the cell pellet and incubated at 0 ℃ on ice for 30 minutes, cell pellet obtained by cold centrifugation as described above. Finally, Cell pellet was resuspended with 5 mL of 0.1 M $CaCl_2$ + 15% Glycerol solution and aliquoted as 50 µL volumes.

Heat shock transformation of chemically competent *E. coli* DH5α cells with G4 inserted plasmids were performed: Aliquoted competent cells were thawed on ice for 30 minutes. 5 µL of each G4 variant plasmid were added on different aliquots of competent cells. Transformation mixtures were exposed to heat shock by incubating in water bath at 42 ℃ for 45 seconds and then transferred back on ice and incubated for 2 minutes. After heat shock, cells were suspended in 1 mL of SOC medium and incubated at 37 ℃ for 1 hour with 220 rpm shaking with Thermo Shaker Incubator MSC-100 (Allsheng, Hangzhou, Zhejiang, China).

Transformed *DH5α* cells were selected on LB agar media with ampicillin. Ampicillin stock solution (100 mg/mL) prepared by dissolving Ampicillin powder (Cayman Chemical, Ann Arbor, MI, USA) in dH_2O and filtrated by 0.22 µm PTFE syringe filters (ISOLAB, Eschau, Germany). Ampicillin stock solution was added into the media as the final concentration will be 100 µg/mL after heating the LB agar powder (AppliChem, Darmstadt, Germany) while it was still liquid but cooled down to approximately 60℃ to prevent degradation of ampicillin while ensuring its dissolution. 300 µL of transformed cells were spread onto LB agar media poured in petri dishes and incubated at 37 ℃ overnight with incubator BD56 (Binder, Tuttlingen, Germany). Single colonies were selected from LB agar plates and inoculated into fresh sterile LB broth with ampicillin (100 mg/mL) and incubated at 37 ℃ overnight with 220 rpm shaking with incubator BD56 (Binder, Tuttlingen, Germany) for cells to reach sufficient density.

2.4. Insert Sequence Confirmation by Sanger Sequencing

Transformed G4 variant plasmids were isolated with Nucleo Spin® Plasmid Isolation Kit (Macherey-Nagel, Dueren, Germany) following the manufacturer's instructions and insert sites were confirmed by sanger sequencing. Sequencing PCR to amplify the site to be sequenced was setup for each variant of G4 motif, master mix containing: Master mix containing dNTP, DNA Polymerase and buffer 1X (Ampliqon, Odense,

Denmark), sequencing primers common in all variant groups (Table 2) (each) 1.0 µM, , plasmids with G4 variants 3.0 ng/µL. 20 µL of reaction mixes were transferred in 0.2 mL PCR tubes and placed in thermal cycler TurboCycler 2 (Blue-ray Biotech, New Taipei City, Taiwan) and PCR was run with following conditions: initial denaturation at 98 ℃ for 5 minutes, 35 cycles of: denaturation at 94 ℃ for 20 seconds, annealing at 55 ℃ for 20 seconds, extension at 72 ℃ for 30 seconds and final extension at 72 ℃ for 10 minutes.

Table 2. Primers Used to Amplify Sequencing Region. Sequences given in 5' to 3' direction. Nucleotide position on plasmid sfGFP-pBAD were given as the corresponding nucleotide of 5' end and 3' end respectively.

Primers	**Sequence (5'→3')**	**Position on Plasmid**
Forward	CACTTTGCTATGCCATAGCATTTTTATCC	1641-1669
Reverse	CGTCGCCGTCCAGCTCGACCAG	1955-1934

Sequencing amplicons were confirmed by 1% agarose gel electrophoresis by Owl™ EasyCast™ B1A (Thermo Fisher Scientific, Waltham, MA, USA) and EV3020 (Consort, Turnhout, Belgium). Agarose gel was visualised with UV imager (Vilber, Collégien, France). Amplified DNA was extracted from the gel with Gel Extraction Kit (Omega Biotek, Norcross, GA, USA) following manufacturer's instructions. Isolated plasmids from the gel were analysed with sanger sequencing (Sentebiolab, Ankara, Turkey).

2.5. Transformation of E. coli BL21 strain with Plasmids

Plasmids with G4 variants were transformed into *E. coli BL21* strain cells and prepared aliquots of glycerol stocks for long-term storage. Competent cell preparation, transformation, and selection of *BL21* strain were prepared with the same method as described above for *DH5α* strain. *BL21* cells transformed with plasmids containing G4 variants were passaged into fresh LB broth with ampicillin (100 µg/mL) and incubated at 37 ℃ for 3h with 220 rpm shaking with incubator BD56 (Binder, Tuttlingen, Germany) for cells to reach exponential phase to have maximum viability for glycerol storage conditions. Glycerol stocks were prepared for *BL21* cell clones containing each G4 variant plasmids by mixing 500 µL of cell culture and 500 µL of 50% glycerol

solution. Bacterial stock solutions were placed in -80 ℃ for long-term storage.

2.6. Fluorescence Spectroscopy Measurement of GFP Expression

Fluorescence levels of *E. coli BL21* cells transformed with GFP encoding plasmids with G4 variants were measured after inducing GFP expression by growing in arabinose containing media to observe the effects of G4 structure variants on gene expression in *E. coli*. Effects of presence of Li^+ and TMPyP4 in the environment were also measured.

E. coli BL21 cells transformed with G4 variants inserted in araBAD promoter site were grown in 3 different media at 37 ℃ for 3 hours for GFP expression under araBAD promote in replicates of 5: L-arabinose (0.02% (w/v)) and ampicillin (100 µg/mL) containing media, LiCl (50mM), L-arabinose (0.02% (w/v)) and ampicillin (100 µg/mL) containing media and TMPyP4 (200 nM), L-arabinose (0.02% (w/v)) and ampicillin (100 µg/mL) containing media. After the incubation cell cultures were aliquoted as 1 mL. Cells were pelleted by centrifugation at 3,000 ×G, 5 ℃ for 15 minutes and supernatant was discarded. Cell pellets were suspended in 110 µL of TNG buffer (Tris HCl 100 mM, NaCl 150 mM, Glycerol 10%, pH 7.5). Cell suspensions in TNG buffer were diluted 1:40 in TNG buffer. 1 mL of diluted cell suspensions were transferred into measuring cuvette and placed in Spectrofluorometer FP-8300 (Jasco, Tokyo, Japan) for fluorescence measurement with excitation at 488 nm and emission wavelengths between 495-600nm.

2.7. Data Analysis

Primary analysis of the fluorescence spectra data was performed by Microsoft Excel software (version 2305) to sort and find the maxima and their corresponding wavelength value.

Statistical analysis of the fluorescence intensity data was performed by GraphPad Prism software (version 9.0.2). Mean of the maxima values of 5 replicates were plotted with their respective standard deviations. Ordinary one-way ANOVA test followed by Tukey's multiple comparison test were used to analyse the statistical difference of fluorescence intensities at maxima between G4 variant groups in each treatment group.

CHAPTER 3: RESULTS

3.1. Confirmation of Insertion by Sanger Sequencing

Insertion of designed G4 variant motifs into downstream of promoter was confirmed by sanger sequencing. Agarose gel electrophoresis of sequencing amplicons was resulted clear, single bands at around 350 bp, as the smallest band observed in WT group while other groups have approximately same size, indistinguishable from each other with the resolution of the agarose gel used in this setup (Figure 4).

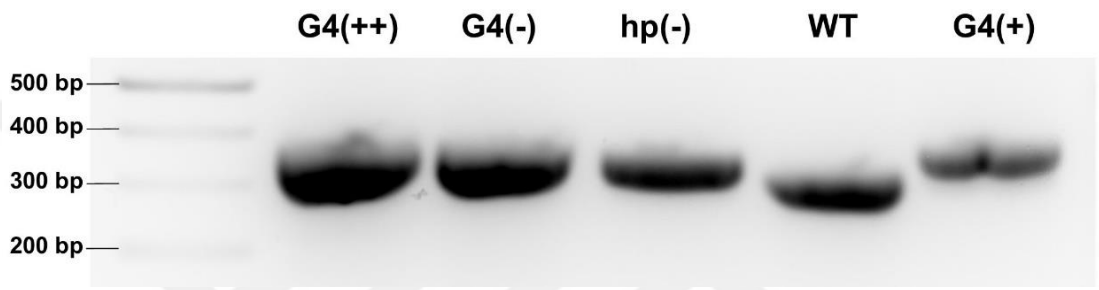

Figure 4. Agarose Gel Electrophoresis of Sequencing Amplicons. Sequencing PCR products were run in 1% Agarose gel prior to sanger sequencing. Leftmost well containing 1kb DNA marker. G4(++): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif; G4(-): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; hp(-): *E. coli* transformed with plasmid inserted with only G4 forming motif; WT: *E. coli* transformed with native plasmid; G4(+): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif.

Sequencing amplicons extracted from agarose gel were analysed by sanger sequencing and DNA sequence information of the insert region was obtained (Figure 5).

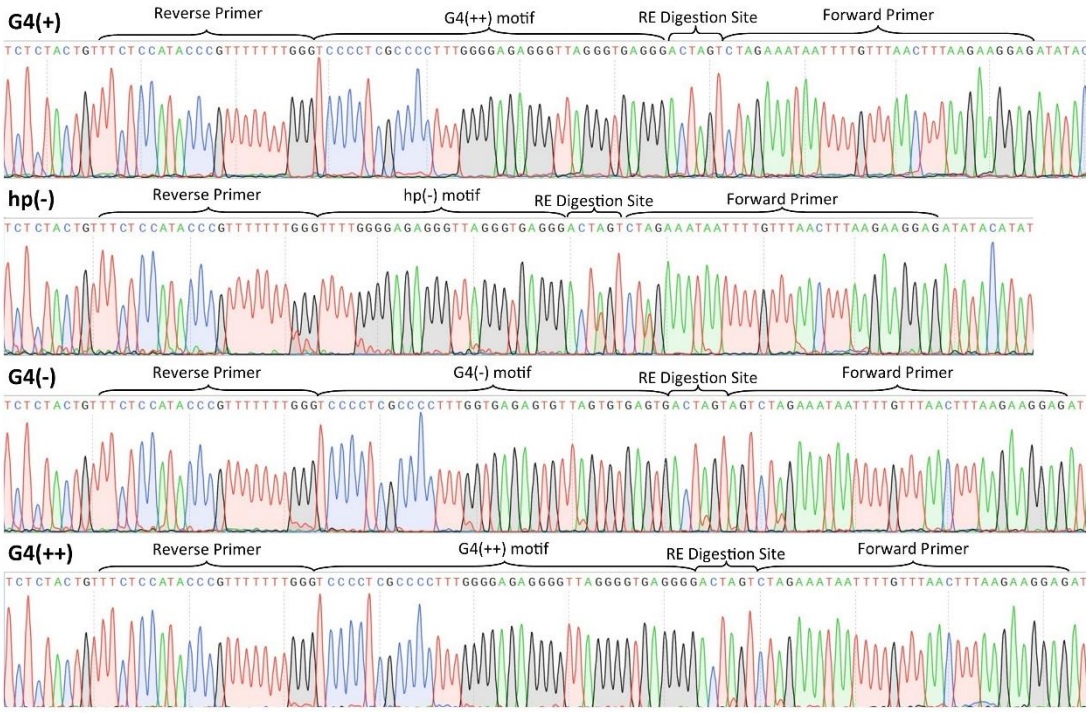

Figure 5. Electropherograms and Corresponding DNA Sequence of the G4 Variant Inserts. Sanger sequencing readings of the plasmid region where site directed mutagenesis were performed. Primer sites, restriction enzyme digestion sites and inserted G4 motifs were indicated above the electropherograms. Coloured peaks indicate nucleotide readings (red: thymine, blue: cytosine, green: adenosine, black: guanine). G4(+): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif; hp(-): *E. coli* transformed with plasmid inserted with only G4 forming motif; G4(-): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; G4(++): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif.

3.2. Fluorescence Spectra of GFP Expression

Fluorescence measurements of GFP expressed in *E. coli BL21* cells by sfGFP-pBAD plasmid with G4 variants on the promoter region were resulted in spectra with a peak value between 512 nm in all G4 variant groups in all treatment groups (Figure 6).

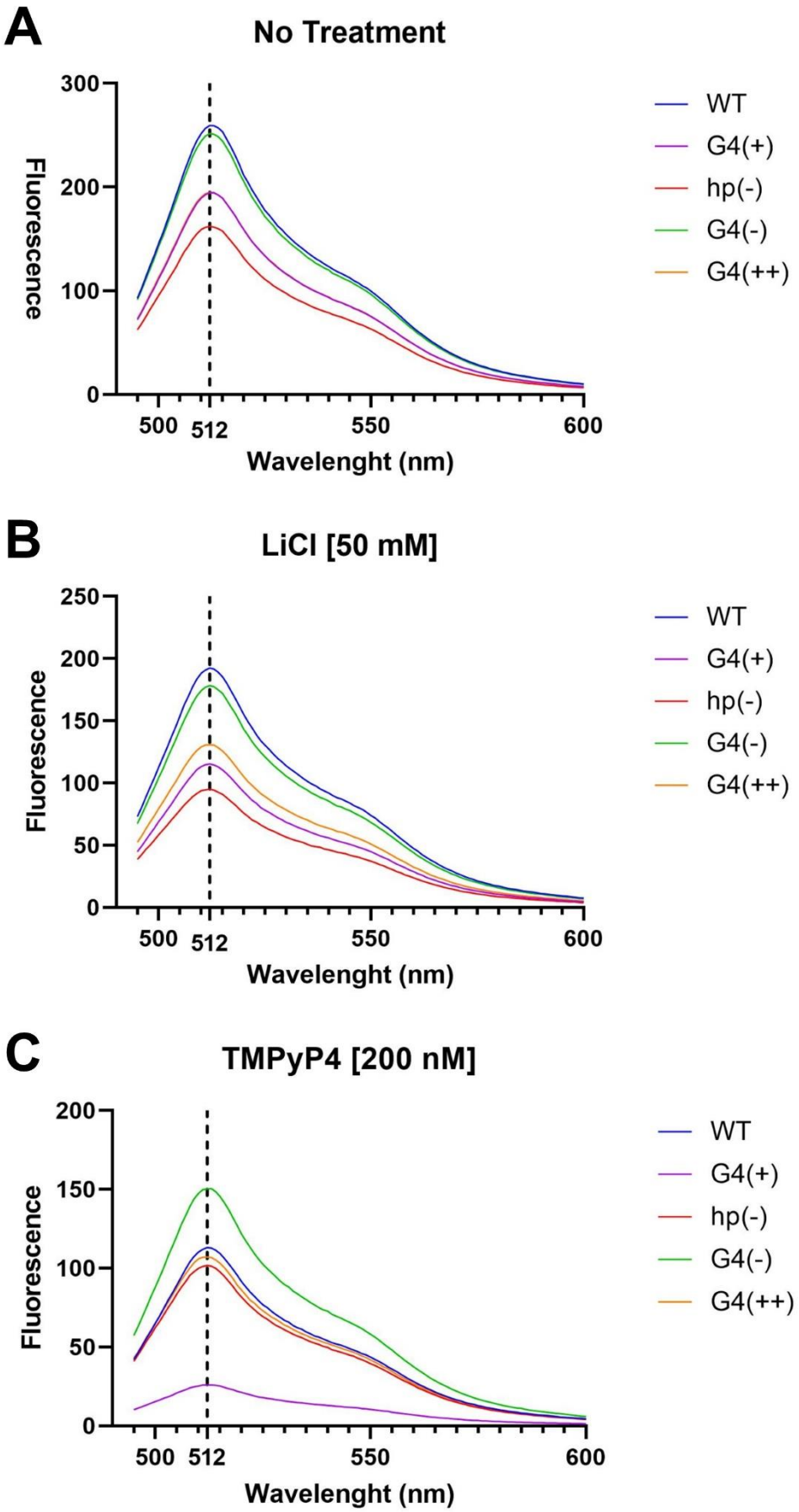
A
No Treatment
Fluorescence
300
200
100
0
500
512
550
600
Wavelenght (nm)
WT
G4(+)
hp(-)
G4(-)
G4(++)
B
LiCl [50 mM]
Fluorescence
250
200
150
100
50
0
500
512
550
600
Wavelenght (nm)
WT
G4(+)
hp(-)
G4(-)
G4(++)
C
TMPyP4 [200 nM]
Fluorescence
200
150
100
50
0
500
512
550
600
Wavelenght (nm)
WT
G4(+)
hp(-)
G4(-)
G4(++)

Figure 6. Fluorescence Spectra of GFP expressing *E. coli* with G4 Variants. Mean of the replicates of fluorescence intensity spectra of GFP expression between 495 and 600 nm wavelengths from *E. coli BL21* transformed with sfGFP-pBAD plasmid with G4 motif variants inserted downstream of pBAD promoter. Peak of spectra at 512 nm was indicated. Error bars represent standard deviation of 5 replicates. (A) *E. coli* grown in media with no treatment. (B) *E. coli* grown in media with 50 mM LiCl. (C) *E. coli* grown in media with 200 nM TMPyP4. WT (blue): *E. coli* transformed with native plasmid; G4(+) (purple): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif; hp(-) (red): *E. coli* transformed with plasmid inserted with only G4 forming motif; G4(-) (green): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; G4(++) (orange): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif.

Statistical comparison between mean of the peak values of GFP fluorescence intensities at 512 nm of G4 variant groups were resulted with significant differences between groups.

3.2.1. Peak Fluorescence Intensities of GFP Expression with No Treatment

GFP expressing *E. coli* with each G4 variant grown in media with no additional treatment other than arabinose and ampicillin has highest fluorescence intensities in WT group (259.3±19.74) followed by G4(-) (251.6±37.10), G4(++) (194.9±16.56), G4(+) (194.7±12.47) and hp(-) (162.0±12.56) with the lowest fluorescence intensity. WT group is significantly higher than hp(-) ($p<0.0001$), G4(+) ($p=0.0011$) and G4(++) ($p=0.0012$) groups while there is no significant difference with group G4(-) (0.9789). G4(+) group is significantly lower than G4(-) group ($p=0.0041$) and no significant difference between groups hp(-) ($p=0.1603$) and G4(++) ($p >0.9999$) (Figure 7).

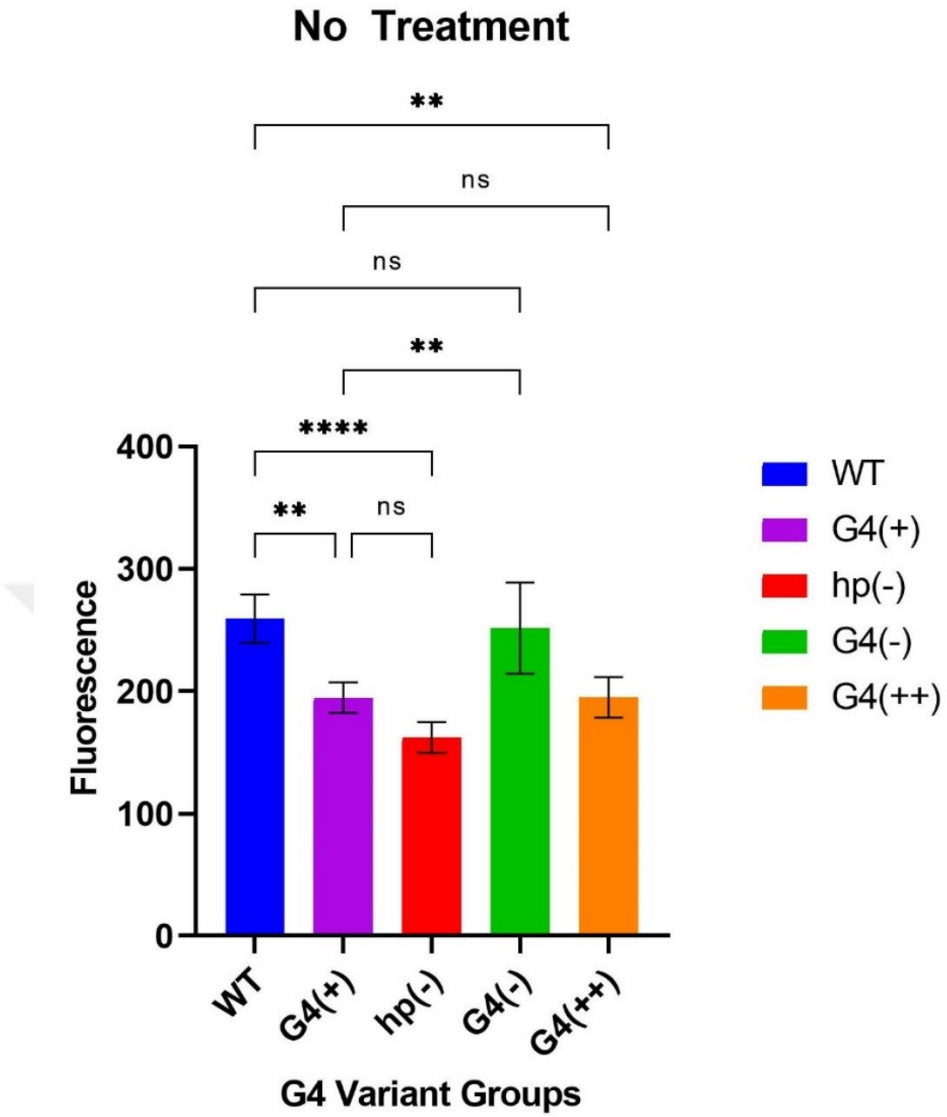

Figure 7. Peak Fluorescence Intensities of G4 Variant Groups with No Treatment. Mean fluorescence intensities at 512 nm of *E. coli* transformed with plasmid expressing GFP under regulation of promoter with G4 variants. Error bars represent standard deviation of 5 replicates. Statistical significance was shown by asterisks: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$. WT (blue): *E. coli* transformed with native plasmid; G4(+) (purple): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif; hp(-) (red): *E. coli* transformed with plasmid inserted with only G4 forming motif; G4(-) (green): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; G4(++) (orange): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif.

3.2.2. Peak Fluorescence Intensities of GFP Expression with LiCl Treatment

GFP expressing *E. coli* with each G4 variant grown in media with 50 mM LiCl has highest fluorescence intensities in WT group (192.4 ±32.20) followed by G4(-) (178.3 ±29.86), G4(++) (130.6 ±13.11), G4(+) (115.4 ±8.837) and hp(-) (95.03 ±8.359) with the lowest fluorescence intensity. WT group is significantly higher than hp(-) ($p<0.0001$), G4(+) ($p=0.0001$) and G4(++) ($p=0.0014$) groups while there is no significant difference with group G4(-) ($p=0.8268$). G4(+) group is significantly lower than G4(-) group ($p=0.0012$) and no significant difference between groups hp(-) ($p=0.5641$) and G4(++) ($p= 0.7845$) (Figure 8).

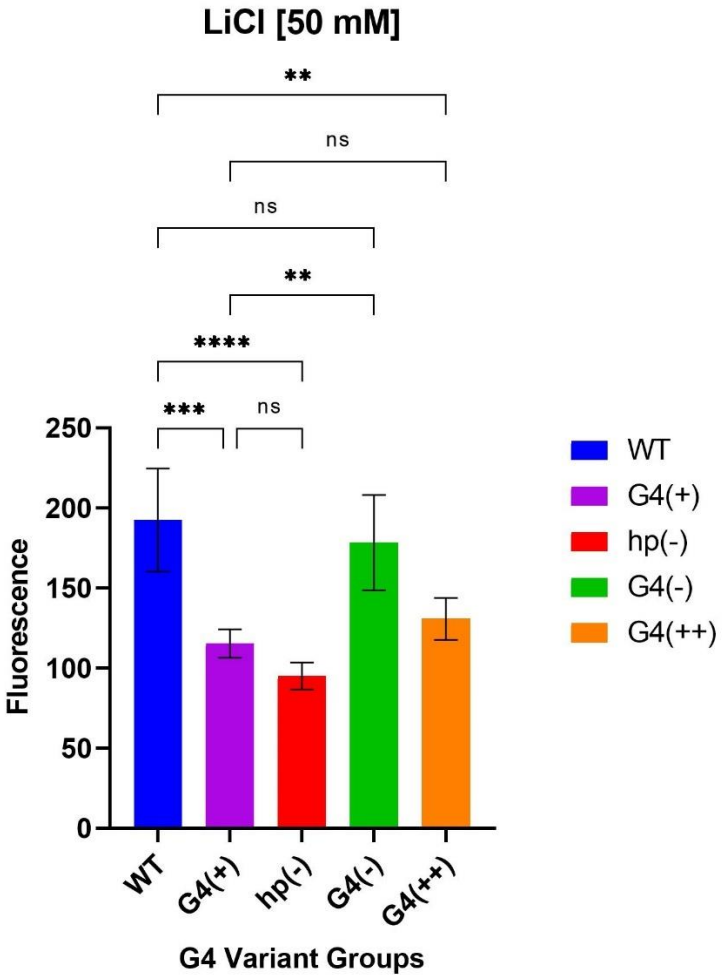

Figure 8. Peak Fluorescence Intensities of G4 Variant Groups with LiCl Treatment. Mean fluorescence intensities at 512 nm of *E. coli* treated with 50 mM LiCl transformed with plasmid expressing GFP under regulation of promoter with G4 variants. Error bars represent standard deviation of 5 replicates. Statistical significance was shown by asterisks: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** p

≤0.0001. WT (blue): *E. coli* transformed with native plasmid; G4(+) (purple): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif; hp(-) (red): *E. coli* transformed with plasmid inserted with only G4 forming motif; G4(-) (green): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; G4(++) (orange): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif.

3.2.3. Peak Fluorescence Intensities of GFP Expression with TMPyP4 Treatment

GFP expressing *E. coli* with each G4 variant grown in media with 200 nM TMPyP4 treatment has highest fluorescence intensities in G4(-) group (150.2±22.37) followed by WT (113.0±15.12), G4(++) (107.0±16.68), hp(-) (101.8±5.695) and G4(+) (25.90±4.025). WT group is significantly higher than G4(+) group ($p<0.0001$) and significantly lower than G4(-) group ($p=0.0051$) and no significant difference with hp(-) ($p=0.7388$) and G4(++) ($p=0.9635$) groups. G4(+) group is significantly lower than hp(-) ($p<0.0001$), G4(-) ($p<0.0001$) and G4(++) groups ($p<0.0001$) (Figure 9).

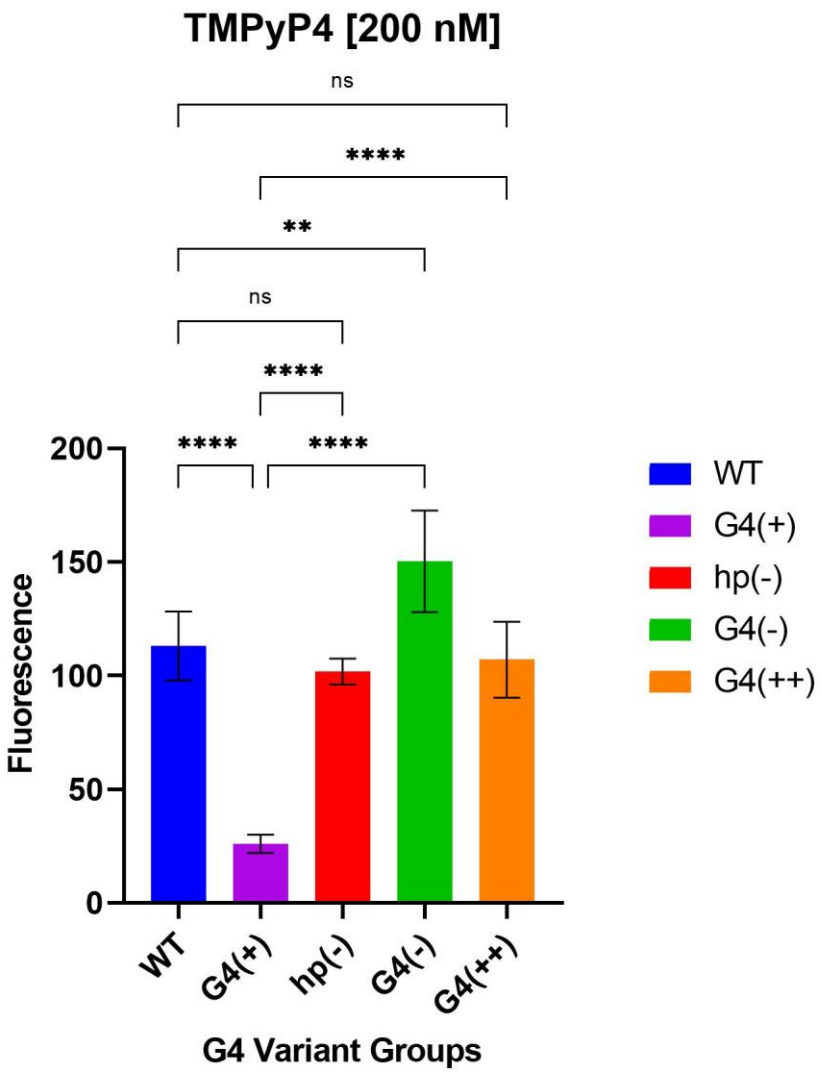

Figure 9. Peak Fluorescence Intensities of G4 Variant Groups with TMPyP4 Treatment. Mean fluorescence intensities at 512 nm of *E. coli* treated with 200 nM TMPyP4 transformed with plasmid expressing GFP under regulation of promoter with G4 variants. Error bars represent standard deviation of 5 replicates. Statistical significance was shown by asterisks: ns $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$. WT (blue): *E. coli* transformed with native plasmid; G4(+) (purple): *E. coli* transformed with plasmid inserted with cruciform hairpin forming G4 motif; hp(-) (red): *E. coli* transformed with plasmid inserted with only G4 forming motif; G4(-) (green): *E. coli* transformed with plasmid inserted with only cruciform hairpin forming motif; G4(++) (orange): *E. coli* transformed with plasmid inserted with potentially more stable cruciform hairpin forming G4 motif.

CHAPTER 4: DISCUSSION

DNA G4 structures has been shown to alter gene expression in many organisms, commonly silencing the gene when they are present in the promoter region (Beaume et al., 2013; Holder and Hartig, 2014; Waller et al., 2016; Yadav et al., 2021). The mechanism of this gene regulation is thought to be involving transcription machinery such as transcription factors and RNA Polymerase in mammalian cell lines and *E. coli* (Sengupta, Roy and Chowdhury, 2021; Shitikov et al., 2022). In mammalian cells, HeLa S3 and A549, researchers have been identified complex formations of transcription factors NM23-H2 and TRF with HDAC (histone deacetylase), REST (RE1 silencing transcription factor), co-REST and LSD1 (lysine-specific-histone-demethylase-1) to directly interact with DNA G quadruplexes on promoter region of genes that have roles in cancer metabolism to regulate their expression through histone demethylation: hTERT (human telomerase reverse transcriptase), p21 (cyclin-dependent kinase inhibitor 1) and c-myc (Saha et al., 2017; Mukherjee et al., 2019; Sengupta, Roy and Chowdhury, 2021). In *E. coli*, G quadruplex structures on promoter regions were proposed to counteract RNA polymerase sigma factor 70 (σ^{70}) which is essential for transcription initiation (Yadav et al., 2017). Although the effect of G4s on gene expression studied considerably in the literature, there are many types and topologies of G4s that could have different effects (Burge et al., 2006; Huppert, 2010; Biver, 2022). In this study, to investigate the effects of 4 specific variants of G4-hairpin switching motifs on gene expression upon insertion downstream to the promoter region were investigated by florescence spectroscopy of GFP as a reporter gene.

Sequencing results of the insertions confirms that the motifs were inserted into the targeted site as expected. These insertions cause increments in size of the plasmid relative to size of the inserted motifs as it can be seen by the gel electrophoresis image (Figure 4). Expected specific size of the inserts can be calculated by addition of size of the motif and size of a single restriction enzyme digestion size, since the restriction enzyme cuts a palindromic motif and in ligation step the two halves are merged. Sizes of the G4 variant motifs G4(+), hp(-), G4(-) and G4(++) are 43 bp, 32 bp, 43 bp and 46 bp respectively. Gel electrophoresis image shows sequencing PCR amplicons, having a size of 315 bp in WT group (no insert native plasmid). For the other groups,

sizes were expected to be the sum of the motif size and sequencing amplicon size of native plasmid, which are 358 bp, 347 bp, 358 bp and 361 bp for G4(+), hp(-), G4(-) and G4(++) groups respectively, which is consistent with the gel image (Figure 4).

Fluorescence spectra of the GFP expressing *E. coli BL21* cells under regulation of promoter with G4 variants has a peak emission at 512 nm which is close to the maximum emission wavelength of sfGFP (super-folder Green Fluorescent Protein) used in this study (Addgene plasmid # 54519) (Pédelacq et al., 2005). The level of fluorescence intensity at the peak wavelength of 512 nm could indicate the expression level of GFP gene.

In the cells grown in medium with no treatment, WT group has significantly higher peak fluorescence intensity than G4(+) and G4(++) groups, which indicates that the G4 structure downstream of promoter has reduced gene expression, possibly by interfering with the transcription mechanism (Holder and Hartig, 2014; Shitikov et al., 2022). WT group has also significantly higher fluorescence intensity and thus higher GFP expression than Hp(-) group and has no significant difference than G4(-) group, presenting strong evidence that the secondary structure reducing gene expression is the G quadruplex rather than hairpin and presence of cruciform hairpin motif does not affect gene expression since G4(+) group can form both G quadruplex and cruciform hairpin whereas hp(-) group can only form G quadruplex and G4(-) group can only form cruciform hairpin. It can also be observed that G4(+) group has significantly lower gene expression than G4(-), supporting the same outcome that G quadruplex is the effective factor on gene expression. Likewise, statistically non-significant difference of gene expression between G4(+) and hp(-) groups suggesting that the cruciform hairpin has no significant effect on gene expression, or it does not form stable enough in the reversible switching mechanism between cruciform hairpin structure and G quadruplex in G4(+). In addition, no significant difference of gene expression between groups G4(+) and G4(++) is showing that the estimated enhanced stability of G4 quadruplex in G4(++) group over G3 quadruplex in G4(+) has no additional functional effect on gene expression, might be caused by the additional guanine nucleotides in G-tracts to form stacks of 4 G-tetrads may also contribute to the stability of hairpin structure as 2 of the G tracts are included in the hairpin motif as well, eventually neglecting the potential dominance of G4 over hairpin.

Cations, especially monovalent cations are known to be stabilize the G quadruplex structures with the order of stabilisation of: Sr^{+2}>K^{+}> Ca^{+2}> NH_4^{+}, Na^{+}, Rb^{+}>Mg^{+2}>Li^{+}≥Cs^{+} (Largy, Mergny and Gabelica, 2016). In the cells grown in medium with 50 mM LiCl treatment, similar pattern of statistical comparison with no treatment group is observed. The greater likelihood of gene expression of WT higher than G4(+) in LiCl treatment group (p=0.0001) comparing to no treatment group (p=0.0011) could because of Li^{+} ion function to stabilize G4 structure as a monovalent cation by intercalating between the Guanine bases (Largy, Mergny and Gabelica, 2016; Yadav et al., 2017), and thus could be stabilizing the G quadruplex competing with cruciform hairpin even more in the reversible switch mechanism. No other additional significant effect of LiCl added to the medium compared to no treatment group was observed as it could be at insufficient concentration to create an observable impact and it is the second least stabilizing agent among other cations that are known to stabilize G4s. However cellular uptake of Li^{+} ions by *E. coli* cells should also be considered as the intracellular ion concentration in media would not expect to be the same with the extracellular environment. Cellular uptake and release by Na^{+} (Li^{+})/H^{+} antiporters NhaA and NhaB as well as the toxicity should be considered as it was shown that lethal dose for *E. coli* strain W3133-2 is 700 mM (INABA et al., 1994). To observe significant change, the cell culture media could be treated at higher concentrations of LiCl or a NhaA and/or NhaB knock out cells could be employed with non-lethal doses of Li^{+} concentration.

TMPyP4 is a commonly studied ligand that is known to stabilize G quadruplexes (Izbicka et al., 1999). In the cells grown in medium with 200 nM TMPyP4, gene expression levels of groups changed more drastically. Higher gene expression of WT group than G4(++) in TMPyP4 treatment group (p<0.0001) can be observed with greater likelihood compared to no treatment group (p=0.0011) this could be explained by TMPyP4 stabilizing the G4 structure and making it even more persistent in the reversible switch mechanism. Conversely, the statistically significant difference between WT and G4(++) groups in no treatment group cannot be observed in TMPyP4 treated group although it bears the same property of reversible switch. Interestingly, there are no significant difference of gene expression between WT, hp(-) and G4(++) groups in TMPyP4 treated group, unlike no treatment group. More interestingly, in the

TMPyP4 treatment group gene expression of G4(-) group is significantly higher than WT group, unlike no significant difference in no treatment group. Also, G4(+) group is significantly lower than every other group: WT, hp(-), G4(-) and G4(++). These unexpected results might be caused by TMPyP4 treatment stabilizing an unpredicted G quadruplex structure which otherwise will not occur due to its unstable characteristics such as low G repeats on G-tracts and longer loop size or another secondary DNA structure that TMPyP4 can bind, which is competing with the anticipated structures since it is at the close site of the inserted motifs but has no functional effect on GFP gene expression. Therefore, such structure could decrease the permanence of G quadruplex and its functional effect of lowering gene expression. G4(-) group having a higher gene expression than WT group could be explained by the same reasoning as its specific motif sequence could not permit such structure to be formed or TMPyP4 demonstrating a specific effect to the hairpin structure formed by motif of G4(-). Cellular toxicity of TMPyP4 should be considered like LiCl as it was shown in *E. coli* K-12 strain (Huang et al., 2018). On the contrary that TMPyP4 stabilizing DNA G4 structures, an interesting study employing molecular dynamics suggest a destabilizing property of TMPyP4 on RNA G quadruplexes (Haldar et al., 2022), and another study shows that a mitochondrial DNA G quadruplex, mt9438 is destabilized upon interaction with TMPyP4 (Zhou et al., 2021). Altogether, these considerations could help explaining the mechanism that give rise to these distinct expression pattern amongst G4 variant groups.

To uncover the functional properties of unpredicted G quadruplex structures such as the one yielding the results of TMPyP4 treatment, the sequence of the insertion region in the plasmid could be analysed more in depth by utilizing in-silico G quadruplex searching tools such as G4Catchall (Doluca, 2019), Quadparser, GRS mapper, G4 Hunter, PQS-finder, Quadbase (D'Antonio and Bagga, 2004; Huppert and Balasubramanian, 2005; Bedrat, Lacroix and Mergny, 2016; Dhapola and Chowdhury, 2016; Hon et al., 2017). In addition, molecular dynamics simulations could be performed to analyse the interaction between TMPyP4 and possible G4 structures or other DNA secondary structures more in-depth. Moreover, additional insertions at different sites close to the one in this study (promoter region), or into different expression vectors which would yield different sequence compositions upon insertion of G4 variant motifs could be performed to analyse the effect of unpredicted G4

structures.

CHAPTER 5: CONCLUSION

DNA G quadruplex structures were studied extensively on their molecular structure, topology, genomic locations, and functions in the literature. They hold significance as they are also having important roles in gene expression of complex phenotypes, which could be manipulated via specific ligands that binds to them. In current study we have investigated the functional roles of specific G4-hairpin switch variants on regulation of gene expression in a reporter construct in *E. coli* as a model organism. Main outcomes of the findings of this study are that the G quadruplex is the prevalent structure over cruciform hairpin in the switch conformation that has effect on downregulating gene expression. In addition, our design for constructing a more stable G quadruplex structure seems to be not effective as hypothesized since it rendered competing hairpin structure more stable as well. LiCl treatment was also observed to be not effective as expected, which could be related to the lower stabilizing capacity of Li^+ over other monovalent ions, or cellular metabolism of the bacteria for ion transport. Another cell line mutant of NhaA and/or NhaB could be used to observe considerable effects. TMPyP4 treatment results can be evaluated by the ligand stabilizing a putative unpredicted DNA structure which also competes with the intended motifs of G4 and presence of it depends on sequence at the insertion site, which could be evaluated in-depth by utilizing *in-silico* tools. TMPyP4 should also be considered for its toxicity and reported destabilizing properties in literature for our G4 variants which could be evaluated by molecular dynamics simulations. Finally, different vector-insert constructs could be studied to analyse the functional effects of these specific G4 variants more comprehensively. We anticipate that outcomes of this study and the following studies will contribute to our understanding about G quadruplexes, their functions, and potential biotechnological applications in the future.

REFERENCES

Andorf, C.M., Kopylov, M., Dobbs, D., Koch, K.E., Stroupe, M.E., Lawrence, C.J. and Bass, H.W. (2014) *G-Quadruplex (G4) Motifs in the Maize (Zea mays L.) Genome Are Enriched at Specific Locations in Thousands of Genes Coupled to Energy Status, Hypoxia, Low Sugar, and Nutrient Deprivation*. Journal of Genetics and Genomics, Vol. 41(12), pp. 627–647.

Arora, A. and Maiti, S. (2008) *Effect of Loop Orientation on Quadruplex−TMPyP4 Interaction*. The Journal of Physical Chemistry B, Vol. 112(27), pp. 8151–8159.

Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) *Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?* Nature Reviews Drug Discovery, Vol. 10(4), pp. 261–275.

Beaume, N., Pathak, R., Yadav, V.K., Kota, S., Misra, H.S., Gautam, H.K. and Chowdhury, S. (2013) *Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation*. Nucleic Acids Research, Vol. 41(1), pp. 76–89.

Bedrat, A., Lacroix, L. and Mergny, J.L. (2016) *Re-evaluation of G-quadruplex propensity with G4Hunter*. Nucleic acids research, Vol. 44(4), pp. 1746–1759.

Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) *Quantitative visualization of DNA G-quadruplex structures in human cells*. Nature Chemistry, Vol. 5(3), pp. 182–186.

Biver, T. (2022) *Discriminating between Parallel, Anti-Parallel and Hybrid G-Quadruplexes: Mechanistic Details on Their Binding to Small Molecules*. Molecules, Vol. 27(13), p. 4165.

Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) *DNA secondary structures: stability and function of G-quadruplex structures*. Nature Reviews Genetics, Vol. 13(November).

Brown, S.L. and Kendrick, S. (2021) *The i-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure*. Pharmaceuticals 2021, Vol. 14(2), p. 96.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) *Quadruplex DNA: sequence, topology and structure*. Nucleic Acids Research, Vol. 34(19), p. 5402.

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 Letters, Vol. 590(17), pp. 2870–2883.

D'Antonio, L. and Bagga, P. (2004) *Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences*. Proceedings - 2004 IEEE Computational Systems Bioinformatics Conference, CSB 2004, pp. 590–591.

Dhapola, P. and Chowdhury, S. (2016) *QuadBase2: web server for multiplexed guanine quadruplex mining and visualization*. Nucleic Acids Research, Vol. 44(Web Server issue), p. W277.

Doluca, O. (2019) *G4Catchall: A G-quadruplex prediction approach considering atypical features*. Journal of Theoretical Biology, Vol. 463, pp. 92–98.

Endoh, T., Kawasaki, Y. and Sugimoto, N. (2013) *Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability*. Angewandte Chemie International Edition, Vol. 52(21), pp. 5522–5526.

Endoh, T. and Sugimoto, N. (2013) *Unusual −1 Ribosomal Frameshift Caused by Stable RNA G-Quadruplex in Open Reading Frame*. Analytical Chemistry, Vol. 85(23), pp. 11435–11439.

Feklistov, A. (2013) *RNA polymerase: in search of promoters*. Annals of the New York Academy of Sciences, Vol. 1293(1), pp. 25–32.

Haldar, S., Zhang, Y., Xia, Y., Islam, B., Liu, S., Gervasio, F.L., Mulholland, A.J., Waller, Z.A.E., Wei, D. and Haider, S. (2022) *Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex*. Journal of the American Chemical Society, Vol. 144(2), pp. 935–950.

Halder, R., Halder, K., Sharma, P., Garg, G., Sengupta, S. and Chowdhury, S. (2010) *Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide*. Molecular BioSystems, Vol. 6(12), p. 2439.

Hardin, C.C., Perry, A.G. and White, K. (2001) *Thermodynamic and Kinetic Characterization of the Dissociation and Assembly of Quadruplex Nucleic Acids*. Biopolymers (Nucleic Acid Sciences), Vol. 56, pp. 147–194.

Holder, I.T. and Hartig, J.S. (2014) *A Matter of Location: Influence of G-Quadruplexes on Escherichia coli Gene Expression*. Chemistry & Biology, Vol. 21(11), pp. 1511–1521.

Hon, J., Martínek, T., Zendulka, J. and Lexa, M. (2017) *pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R*. Bioinformatics (Oxford, England), Vol. 33(21), pp. 3373–3379.

Huang, L., El-Hussein, A., Xuan, W. and Hamblin, M.R. (2018) *Potentiation by*

potassium iodide reveals that the anionic porphyrin TPPS4 is a surprisingly effective photosensitizer for antimicrobial photodynamic inactivation. Journal of Photochemistry and Photobiology B: Biology, Vol. 178, pp. 277–286.

Huo, Y.-X., Rosenthal, A.Z. and Gralla, J.D. (2008) *General stress response signalling: unwrapping transcription complexes by DNA relaxation via the sigma38 C-terminal domain*. Molecular Microbiology, Vol. 70(2), pp. 369–378.

Huppert, J.L. (2010) *Structure, location and interactions of G-quadruplexes*. The FEBS Journal, Vol. 277(17), pp. 3452–3458.

Huppert, J.L. and Balasubramanian, S. (2005) *Prevalence of quadruplexes in the human genome*. Nucleic Acids Research, Vol. 33(9), p. 2908.

INABA, K., KURODA, T., SHIMAMOTO, T., KAYAHARA, T., TSUDA, M. and TSUCHIYA, T. (1994) *Lithium Toxicity and Na+(Li+)/H+ Antiporter in Escherichia coli*. Biological and Pharmaceutical Bulletin, Vol. 17(3), pp. 395–398.

Izbicka, E., Wheelhouse, R.T., Raymond, E., Davidson, K.K., Lawrence, R.A., Sun, D., Windle, B.E., Hurley, L.H. and Von Hoff, D.D. (1999) *Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells.* Cancer research, Vol. 59(3), pp. 639–44.

Kaplan, O.I., Berber, B., Hekim, N. and Doluca, O. (2016) *G-quadruplex prediction in* E. coli *genome reveals a conserved putative G-quadruplex-Hairpin-Duplex switch*. Nucleic Acids Research, p. gkw769.

Khlebnikov, A., Skaug, T. and Keasling, J.D. (2002) *Modulation of gene expression from the arabinose-inducible araBAD promoter*. Journal of Industrial Microbiology and Biotechnology, Vol. 29(1), pp. 34–37.

Kumar, Pankaj, Yadav, V.K., Baral, A., Kumar, Parveen, Saha, D. and Chowdhury, S. (2011) *Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide*. Nucleic Acids Research, Vol. 39(18), pp. 8005–8016.

Largy, E., Mergny, J.-L. and Gabelica, V. (2016) *Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability*. in Metal Ions in Life Sciences, pp. 203–258.

Mao, S.-Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D. and Balasubramanian, S. (2018) *DNA G-quadruplex structures mold the DNA methylome*. Nature Structural & Molecular Biology, Vol. 25(10), pp. 951–957.

Mukherjee, A.K., Sharma, S., Bagri, S., Kutum, R., Kumar, P., Hussain, A., Singh, P., Saha, D., Kar, A., Dash, D. and Chowdhury, S. (2019) *Telomere repeat–binding factor 2 binds extensively to extra-telomeric G-quadruplexes and regulates the epigenetic status of several gene promoters*. Journal of Biological Chemistry, Vol. 294(47), pp. 17709–17722.

Mullen, M.A., Olson, K.J., Dallaire, P., Major, F., Assmann, S.M. and Bevilacqua, P.C. (2010) *RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles*. Nucleic Acids Research, Vol. 38(22), pp. 8149–8163.

Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2005) *Engineering and characterization of a superfolder green fluorescent protein*. Nature Biotechnology 2005 24:1, Vol. 24(1), pp. 79–88.

Perrone, R., Lavezzo, E., Riello, E., Manganelli, R., Palù, G., Toppo, S., Provvedi, R. and Richter, S.N. (2017) *Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions*. Scientific Reports, Vol. 7(1), p. 5743.

Rawal, P., Kummarasetti, V.B.R., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K. and Chowdhury, S. (2006) *Genome-wide prediction of G4 DNA as regulatory motifs: Role in* Escherichia coli *global regulation*. Genome Research, Vol. 16(5), pp. 644–655.

Saha, D., Singh, A., Hussain, T., Srivastava, V., Sengupta, S., Kar, A., Dhapola, P., Dhople, V., Ummanni, R. and Chowdhury, S. (2017) *Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex–dependent fashion*. Journal of Biological Chemistry, Vol. 292(37), pp. 15205–15215.

Sengupta, A., Roy, S.S. and Chowdhury, S. (2021) *Non-duplex G-Quadruplex DNA Structure: A Developing Story from Predicted Sequences to DNA Structure-Dependent Epigenetics and beyond*. Accounts of Chemical Research, Vol. 54(1), pp. 46–56.

Shitikov, E.A., Bespiatykh, D.A., Bodoev, I.N. and Zaychikova, M. V. (2022) *G-Quadruplex Structures in Bacteria: Functional Properties and Prospects for Use as Biotargets*. Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry, Vol. 16(4), pp. 292–304.

Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. and Balasubramanian, S. (2020) *The regulation and functions of DNA and RNA G-quadruplexes*. Nature Reviews

Molecular Cell Biology 2020 21:8, Vol. 21(8), pp. 459–474.

Waller, Z.A.E., Pinchbeck, B.J., Buguth, B.S., Meadows, T.G., Richardson, D.J. and Gates, A.J. (2016) *Control of bacterial nitrate assimilation by stabilization of G-quadruplex DNA*. Chemical Communications, Vol. 52(92), pp. 13511–13514.

Wheelhouse, R.T., Sun, D., Han, H., Han, F.X. and Hurley, L.H. (1998) *Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(* N *-methyl-4-pyridyl)porphine with Quadruplex DNA*. Journal of the American Chemical Society, Vol. 120(13), pp. 3261–3262.

Wieland, M. and Hartig, J.S. (2007) *RNA Quadruplex-Based Modulation of Gene Expression*. Chemistry & Biology, Vol. 14(7), pp. 757–763.

Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) *Monovalent cation-induced structure of telomeric DNA: The G-quartet model*. Cell, Vol. 59(5), pp. 871–880.

Yadav, P., Kim, N., Kumari, M., Verma, S., Sharma, T.K., Yadav, V. and Kumar, A. (2021) *G-quadruplex structures in bacteria: Biological relevance and potential as an antimicrobial target*. Journal of Bacteriology, Vol. 203(13).

Yadav, V., Hemansi, Kim, N., Tuteja, N. and Yadav, P. (2017) *G quadruplex in plants: A ubiquitous regulatory element and its biological relevance*. Frontiers in Plant Science, Vol. 8.

Yett, A., Lin, L.Y., Beseiso, D., Miao, J. and Yatsunyk, L.A. (2019) *N-methyl mesoporphyrin IX as a highly selective light-up probe for G-quadruplex DNA*. Journal of Porphyrins and Phthalocyanines, Vol. 23(11n12), pp. 1195–1215.

Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R.B. and Pearson, C.E. (2015) *Quadruplex formation by both G-rich and C-rich DNA strands of the* C9orf72 *(GGGGCC)8•(GGCCCC)8 repeat: effect of CpG methylation*. Nucleic Acids Research, p. gkv1008.

Zhou, W., Cheng, Y., Song, B., Hao, J., Miao, W., Jia, G. and Li, C. (2021) *Cationic Porphyrin-Mediated G-Quadruplex DNA Oxidative Damage: Regulated by the Initial Interplay between DNA and TMPyP4*. Biochemistry, Vol. 60(48), pp. 3707–3713.

Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) *X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid*. Journal of Molecular Biology, Vol. 92(2), pp. 181–192.