INVESTIGATION THE FUNCTIONS OF NOVEL G-QUADRUPLEX STRUCTURES IN *E. coli* GENOME

HÜSEYİN SAYGIN PORTAKAL

Master's Thesis

Graduate School

Izmir University of Economics

İzmir

2021

INVESTIGATION THE FUNCTIONS OF NOVEL G-QUADRUPLEX STRUCTURES IN *E. coli* GENOME

HÜSEYİN SAYGIN PORTAKAL

A Thesis Submitted to

The Graduate School of Izmir University of Economics

Master of Science Program in Bioengineering

Izmir

2021

ABSTRACT

INVESTIGATION THE FUNCTIONS OF NOVEL G-QUADRUPLEX STRUCTURES IN *E. coli* GENOME

Portakal, Hüseyin Saygın

Master of Science Program in Bioengineering

Advisor: Assoc. Prof. Dr. Osman Doluca

August, 2021

G-quadruplexes are noncanonical topologies with special tertiary structures of nucleic acid molecules. During many years biological roles of G-quadruplexes have been investigated and revealed in various organisms. In particular, the regulatory functions of G-quadruplexes by foundation on regulatory regions of human genome demonstrates their significance on cellular metabolism. *Escherichia coli* (*E. coli*) is one of the most popular bacterial organisms by being model organisms in laboratory studies since its genomic and proteomic structures have been revealed with details. So far 52 G-quadruplex forming sequences closer to various genes on *E. coli* genome had been analyzed with computational tools. However, their biological functions have not been discovered yet. As such, functions of those G-quadruplex structures were investigated in this study by characterizing G-quadruplex binding protein with MALDI-TOF-TOF technique and looking for effect of G-quadruplex knock-outs on expression levels of adjacent genes with CRISPR/Cas9 technique. MALDI-TOF-TOF results have demonstrated that two distinct proteins which are acetaldehyde alcohol dehydrogenase and DNA dependent RNA polymerase beta subunit have great

possibility to being G-quadruplex binding protein considering the appropriate molecular weights and matching scores. In this scope, it's predicted that the G-quadruplexes could act a role on alcohol fermentation process and transcription of adjacent genes. However, unluckily CRISPR/Cas9 approach which is developed in this study have to be improved since that G-quadruplexes have not been edited due to their higher stability. The revealing of biological functions of G-quadruplexes will keep light to their significance on bacterial metabolisms and will be a milestone for further studies.

Keywords: G-quadruplex, Tertiary structures, DNA topologies, Biological function, *E. coli*, MALDI-TOF-TOF, CRISPR/Cas9

ÖZET

E. coli GENOMUNDA BULUNAN YENİ G-QUADRUPLEX YAPILARININ FONKSİYONLARININ İNCELENMESİ

Portakal, Hüseyin Saygın

Biyomühendislik Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Osman Doluca

Ağustos, 2021

G-dörtlüleri nükleik asit moleküllerinin özel üçüncül yapılardan biri olup kanonik olmayan topolojilerdir. Yıllar boyunca birçok organizmadaki G-dörtlülerinin biyolojik rolleri incelenmiş ve ortaya çıkarılmıştır. Özellikle insan genomunun düzenleyici bölgelerinde bulunması G-dörtlülerinin düzenleyici rollerini ve hücre metabolizmasındaki önemini göstermektedir. *Escherichia coli* (*E. coli*) proteomik ve genomik yapılarının detaylıca ortaya çıkarılması ile laboratuvar uygulamalarında model organizma haline gelerek günümüzde en popüler bakteriyel organizmalardan birisidir. Bu zamana kadar *E. coli* genomunda çeşitli genlerin yakınında bulunan 52 G-dörtlüsü oluşturan sekans hesaplama araçları ile analiz edilmiştir. Ancak biyolojik fonksiyonları henüz keşfedilmemiştir. Bu bağlamda bu çalışmada G-dörtlülerinin biyolojik rolleri G-dörtlüsüne bağlanan proteinlerin MALDI-TOF-TOF tekniği ve CRISPR/Cas9 tekniğiyle G-dörtlüsü nakavtının civarındaki genlerin ekspresyon seviyesine olan etkisinin araştırılmasıyla incelenmiştir. MALDI-TOF-TOF sonuçları uyumlu moleküler ağırlıkları ve eşleşme skorları göz önünde bulundurularak asetaldehit alkol dehidrogenaz (adh) ve DNA bağımlı RNA polimeraz beta altbirimi

olarak iki farklı proteinin G-dörtlüsüne bağlanan protein olmak üzere yüksek olasılığa sahip olduğunu göstermiştir. Bu doğrultuda G-dörtlülerinin alkol fermentasyonu işleminde ve yakın genlerin transkripsiyonunda rol oynadığı tahmin edilmektedir. Ancak maalesef ki bu çalışma doğrultusunda geliştirilen ve G-dörtlülerinin yüksek stabilite sahip olmasından dolayı G-dörtlülerini düzenleyemeyen CRISPR/Cas9 yaklaşımının iyileştirilmesi gerekmektedir. G-dörtlülerinin biyolojik fonksiyonlarının ortaya çıkarılması bakteri metabolizmasındaki önemine ışık tutacak ve gelecek çalışmalar için bir mihenk taşı olacaktır.

Anahtar Kelimeler: G-dörtlüleri, Üçüncül yapılar, DNA topolojileri, Biyolojik fonksiyon, *E. coli*, MALDI-TOF-TOF, CRISPR/Cas9

Dedicated to *all seekers*…

ACKNOWLEDGEMENTS

While I'm completing the Master of Science program, I would like to thank my advisor Assoc. Prof. Dr. Osman DOLUCA now that he gave me the chance for the related project which I evaluate as big opportunity. Additionally, I'm considering myself as so lucky for the relation beyond lecturer-student's and his high patience to me. I appreciate such those facts that could belong only few students.

Besides, I would like to thank my other lecturers who are found also in my jury members; Asst. Prof. Dr. Cihangir YANDIM, Asst. Prof. Dr. Hümeyra TAŞKENT SEZGİN for their huge contributions to my academic career. In addition, Assoc. Prof. Dr. Zeynep FIRTINA KARAGONLAR, Asst. Prof. Dr. Mine GÜNGÖRMÜŞLER, and Asst. Prof. Dr. Fatma Pınar GÖRDESLİ DUATEPE share the same contributions. I would like to declare that their effects are quite huge on my participation to family of Genetics and Bioengineering Department as a Research Assistant.

In addition, I would like to thank my valuable lab-mates and class mates who are creating great harmony in our laboratory. I'm also so lucky to have such those friends in laboratory life.

Except those valuable people, I would like to thank my parents Zafer PORTAKAL, Meral AYSAL and my sister Simge YALÇIN for their great supports, believes, and being a driving force for me.

And finally, I would like to declare my most prominent motivation source who is Mawlana Jalaluddin Rumi now that I could find my way in darkness with the light sourced from his quotes. When I lost my power in every breaking points he had shouted at me from inside a book "Troubles show the direction, and what you seek is seeking you".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1. G-quadruplex Structures and Their Roles in Mammalians

It's known that DNA molecules which are responsible for genetically hereditary have double stranded helix structures. However many biochemistry based studies that had been conducted last years demonstrated that these molecules could be found in various forms with special folding between their nucleotide residues. The most important one of these forms are G-quadruplex structures (Neidle, 2016). G-tetrad structures are formed through the folding of Guanine residues subsequently aligned in a single strand and creating hydrogen bonds to one another via Hoogsten edges (Figure 1A). G-quadruplexes are created by the accumulation of these planar structured G-tetrads subsequently (Sun et al., 2019; Zhang et al., 2009). Nucleotide sequences of these DNA molecules possessing the G-quadruplexes have been revealed as $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ basically (Hänsel-hertsch, Antonio, and Balasubramanian, 2017). G-quadruplex formation requires the accumulation of G-tetrads around many monovalent cations such as Na^{+} and K^{+}(Figure 1B) (Winnerdy, and Phan, 2020).

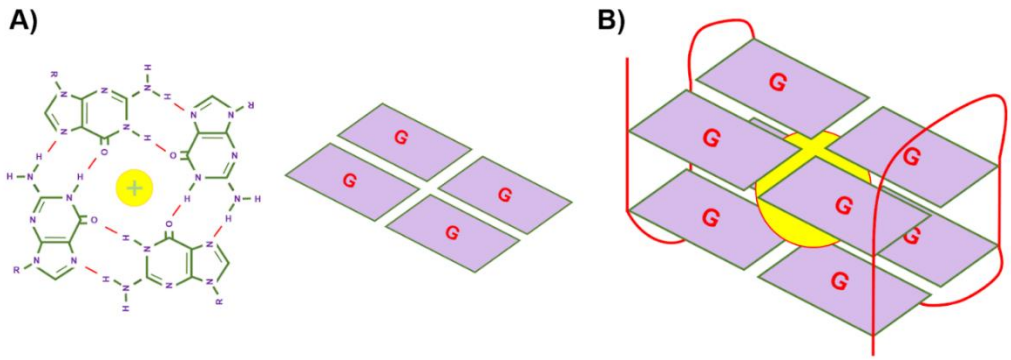

Figure 1. G-quadruplex forming process A) G-tetrad structure created by folding of four guanine residues, B) G-quadruplex structure formed by stacking of G-tetrads.

These non-canonical structures might be observed as both intra- and intermolecular types (Figure 2) (Kim et al., 2003). Besides, several studies have revealed that these structures have higher thermal stability compared to double helix structures (Fujimoto et al., 2009). Although G-quadruplexes had been discovered *in vitro* (Kikin, D'Antonio, and Bagga, 2006), it's also discovered that these structures are found on the genome of various living organisms from microorganisms to human

(Hänsel-hertsch, Antonio, and Balasubramanian, 2017; Harris, and Merrick, 2015). In line with researches in progress, it's indicated that G-quadruplexes could have significant functions by considering their particular location in the genome of organisms and interactions between their ligands (Rhodes, and Lipps, 2015).

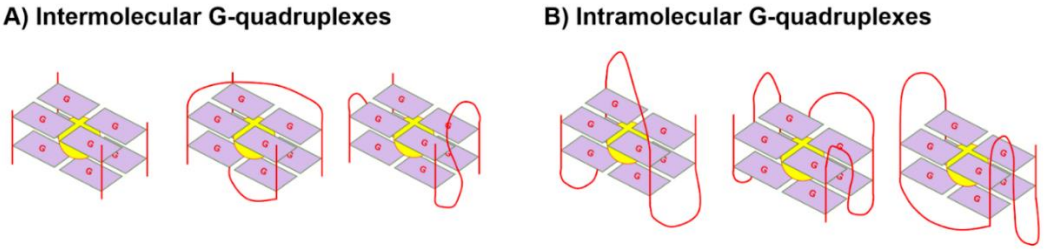

Figure 2. G-quadruplex variants A) intermolecular G-quadruplex structures, B) Intramolecular G-quadruplex structures.

Several computational analyses have revealed that over 300,000 sequences have G-quadruplex forming potential in the human genome (Ding, Fleming, and Burrows, 2018). Furthermore, that these structures do not exhibit random distribution and they are clustered on functional regions of the genome have also been determined. G-quadruplexes are most frequently found in telomere regions of the human genome (Structures, Lin, and Yang, 2017). It's revealed that the G-quadruplexes found in these regions are responsible for inhibiting the telomerase activity (Wang et al., 2011). Apart from them, G-quadruplexes might be formed at the promoter regions of various genetic segments (Balasubramanian, Hurley, and Neidle, 2011), the borders of intron-exon sequences (Weldon et al., 2018), the gene regions responsible for immunoglobulin recombination etc (Almeida et al., 2018). Moreover, the creation of these structures at promoters of significant oncogenes such as c-MYC (You et al., 2014), KRAS (Cogoi, and Xodo, 2006) etc. demonstrates its relations with cancer progression. Strikingly 90% of the human genome's DNA replication origins carry G-quadruplex structures (Rhodes, and Lipps, 2015). In addition, many studies have shown that G-quadruplexes can act a significant role on gene expression by interacting with helicases that are functional on transcription such as XPB and XPD (Sauer, and Paeschke, 2017). Therewithal, it's discovered that the mRNA molecules expressed from approximately 300 gene regions of the human genome have G-quadruplexes at 5'-UTR sequences

and process translation regulation via these structures (Song et al., 2016). Additionally, G-quadruplexes have also been observed on Guanine rich TERRA mRNA products expressed from telomere regions (Wang et al., 2015).

Now that the DNA molecules are found as double helix chromatin structures under standard conditions, it's expected that G-quadruplexes are formed during various processes requiring the disrupting of these double helix structures such as gene transcription, DNA replication, DNA repair (Rhodes, and Lipps, 2015) etc. In such processes Watson-Crick base pairs between the DNA strands are breaking and possible Hoogsten pairings are causing the DNA molecules exhibiting distinct topologies such as G-quadruplexes (Moon et al., 2015). In light of this information that G-quadruplex structures have quite importance on cell metabolism by to be formed during many processes is emphasized. The creation and degradation of these structures through foldings are dependent on many kinetics (Liu et al., 2012). For instance, they are demonstrated in literature that the chaperon proteins (Rajendran et al., 2013) and helicase enzymes (Mendoza et al., 2016) act on the forming and loosening of G-quadruplexes respectively.

Thus it's introduced that G-quadruplexes possess quite importance on cancer progression since they are found on promoter regions of significant oncogenes and they regulate telomerase activity. As such, many researches struggling to develop drug molecules recognizing G-quadruplexes in order to treat cancer disease are frequently produced in literature (Onel, Lin, and Yang, 2014). Furthermore, it's discovered that various diseases could be sourced by mutations on the gene regions coding helicases, similar proteins, and other ligands interacting with G-quadruplex structures (Kharel et al., 2020). In addition, that G-quadruplexes regulate several processes such as transcription, translation, and replication is another parameter showing their importance on metabolic activities. In this scope, further studies that explore G-quadruplex structures, their properties, folding and unfolding kinetics, locations over genome, functions will set light to robust developments in the genetic research area.

1.2. The Roles of G-quadruplexes in Microorganisms

Although the quantity of the studies investigating the importance of G-quadruplexes on mammalian cell metabolism increases incessantly, their function on microorganisms is not analyzed plainly. Nonetheless the evidence is proving that these

structures play a role in metabolism and pathogenicity of microorganisms. The most significant strategy of microorganisms during infection is that they could provide antigenic variation which is processed by recombination of specific genes (Woude, and Ba, 2004). That the observation of G-quadruplexes over several such recombination regions demonstrates the importance of them on antigenic variation (Walia, and Chaconas, 2013). Besides, it's discovered that Human Immunodeficiency Virus (HIV) causing wide range deaths contains long terminal repeats between gene regions and on the terminal of its genome and these repeats are forming G-quadruplex structures. As such, HIV performs many complex processes through G-quadruplexes (Krafčíková et al., 2017). Furthermore these structures might be observed on long terminal repeats of other lentivirus' genomes (Perrone et al., 2017). Correlatively, G-quadruplexes have been discovered on promoter regions of genes which are responsible for virulence factors of Human Herpesvirus (HHV) carrying double stranded DNA as genetic molecule (Biswas et al., 2016). Similarly, it's reported that Human Papilloma Virus (HPV) causing cervical cancer regulates expression of many genes via G-quadruplexes found on their promoters. Viral latency is a mechanism performed by viral organisms after infection in order to protect their existence within the host and to escape from immune response. This mechanism is based on expression of the genes which are providing heterochromatization of the viral genome and inhibiting viral replication (Grinde, 2013). So far, it's revealed that the genes of different HHV strains which are responsible for viral latency are regulated by G-quadruplex structures (Wood, and Royle, 2017). In another case, it's clearly reported that Epstein-Barr virus (EBV) could escape from immune response by suppressing EBNA1 protein. It's recently demonstrated that mRNA of EBNA1 is containing G-quadruplex forming sequences and recognized by nucleolin protein. Through this pathway EBV could cause various types of cancer (Lista et al., 2017). It's also discovered that Adeno-associated viruses (AAV) contain G-quadruplexes on their viral genome and their reproduction is regulated by binding of nucleophosmin (NPM1) to those genomic G-quadruplexes (Satkunanathan, Thorpe, and Zhao, 2017). These findings demonstrate that G-quadruplexes enable ease to viruses during infection and on protecting themselves after infection.

Although the functions of G-quadruplexes on bacteria and yeasts have not been enlightened clearly, the existence frequency of these structures on the genomes of

those organisms shows their importance. So far the data have demonstrated that G-quadruplexes are accumulated on promoters of genes responsible for carbohydrate, amino acids, and nucleotide metabolisms of those organisms more frequently (Saranathan, and Vivekanandan, 2018). For instance, it's discovered that *Deinococcus radiodurans* strain's radiation resistance gene performs activity with G-quadruplexes on its promoter regions (Kota et al., 2015). Surprisingly, the findings reveal that thermophilic bacterial strains' genes' providing resistance to high temperature promoters are rich for G-quadruplexes (Bartas et al., 2019). Ding and his colleagues have investigated G-quadruplex structures of *Thermales* and *Deinococcales* bacterial strains on the study completed in 2018. While obtained results have shown that Guanine rich sequences of genome of *Thermales* strain do not form G-quadruplexes, these structures are observed on the gene regulation regions of the genome of stress resistant *Deinococcales* strain of bacteria (Ding, Fleming, and Burrows, 2018). In the light of this information, particularly, it's found out that the strains of bacteria living in certain environmental conditions could exhibit such properties with G-quadruplexes. Alike viruses, G-quadruplexes are encountered on the genome of pathogenic bacterial strains as well. E.g. Jain, et al. have carried out a research analyzing G-quadruplexes of *Salmonella enterica* strain of bacteria in 2018. In this bioinformatics based study, it's revealed that these strains effective on food poisoning carry G-quadruplex structures over regulation regions of genes providing Mg^{+2}, Fe^{+3} ions, and maltose transportation. Therefore the bacteria could survive under the existence of reactive nitrogen/oxygen species produced by macrophages, and consume the primary sugar source of the human gastrointestinal tract (Jain et al., 2020). In another study Mishra and his colleagues have investigated the G-quadruplexes in the genome of 160 strains of *Mycobacterium tuberculosis*. Results have demonstrated that three essential genes which are *espB*, *espK*, and *cyp51* genes contain G-quadruplex forming sequences at the genomic level. While the genes are functional on survival of bacteria within the host cell, it highlights the importance of G-quadruplexes on virulence status of bacteria (Mishra et al., 2019). While the G-quadruplexes could be investigated via molecular biology based techniques, many bioinformatics tools, too, are used in order to detect the location of G-quadruplexes. For instance, Bartas, et al. have utilized G4Hunter for analyzing the G-quadruplexes among complete bacterial genomes exported from NCBI. According to results, G-quadruplexes are found in all species of bacterial organisms, yet maximum frequency is observed on subgroups of

Deinococcus-Thermus, and minimum frequency is observed on Thermotogae. The scientists have declared that considering non-random localizations of G-quadruplexes onto bacterial genomes, the evolutionary relations between the bacterial species are shown with this study (Bartas et al., 2019).

Eschericia coli (*E. coli*) is one of the types of bacteria that are living in mammalian gastrointestinal tract (Jang et al., 2017). In the first instance Theodor Escherich who is German scientist working on pediatrics and bacteriology has discovered *E. coli* on the fecula of a baby (Shulman, Friedmann, and Sims, 2007). It's predicted that between 100 billion and 10 trillion *E. coli* are traveling daily throughout the human body (Katouli, 2010). While the bacteria composing fecula are commonly aerobic, the ratio of selective anaerobic *E. coli* according to other bacteria species is approximately 0.1%. *E. coli* belongs to intestinal flora, it is also found in the enteric bacteria family as biological classification. These rod shaped bacteria have 1-2 μm length and 0.1-0.5 μm diameter (Reshes et al., 2008). *E. coli* is a type of gram-negative bacterial species, that is its cell wall is formed by peptidoglycan layer and membrane, and it does not generate endospor. As a protection mechanism, the peptidoglycan layer which is found outside of the cell wall protects bacteria inhibits penetration of various antibiotics such as penicillin (Allocati et al., 2013). Besides E. coli is a motile bacterial species now that it possesses flagella. Since that it is adapted to grow in mammalian animal intestines, it could be reproduced on body temperature most efficiently (Nakamura, and Minamino, 2019). Due to natural mutations, quite a number of *E. coli* strains have been discovered during years. Those strains exhibit differentiations in physiology and life cycles because of their molecular level differences. For instance a strain could obtain pathogenicity or antibiotic resistance sourced from the differentiation (Okeke et al., 2000). Furthermore, many serotypes of *E. coli* have been classified by considering their antigens (O: surface antigen, H: flagella antigen). The most famous *E. coli* serotype is O157:H7 now that it could exhibit fatality (Saranathan, and Vivekanandan, 2018).

Although that *E. coli* does not form significant disorders over host organisms due to their coherent interactions, it could cause mild diseases such as urinary tract infection only if it travels to other regions of the host body. While some E. coli strains do not harm animals that they are found into, they could create several diseases over the human body if they transmit. The diseases sourced by E. coli could be listed as

basically; *i)* diarrheal diseases, *ii)* urinary tract infection, *iii)* meningitis, *iv)* peritonitis, *v)* mastitis, *vi)* septicemia, *vii)* gram-negative pneumonia (Vila et al., 2016). Furthermore there are many strains of *E. coli* bacteria classified according to genetic variations. These variations could be encountered as both little mutations on genomic regions and existence or lacking of a whole gene (Aasen, 1998; Souza et al., 1999). These genes are found on bacteriophages, transposons, and plasmids that are transferred from other bacterial species. Even various virulence factors such as adhesion forces or toxins might be sourced from variations of those genetic segments. These parameters bring mild pathogenicity in many strains of *E. coli* bacteria (Galardin et al., 2017; Rousset et al., 2021).

From its first discovery huge numbers of researches have been carried out over E. coli bacteria. Almost all structural features of *E. coli* such as genomic sequences, its proteomic properties, metabolic processes, and aspects have been analyzed with details. In this scope during many years *E. coli* has become a model organism utilized in the microbiology research area. Insomuch as it has become a tool utilized in molecular biology research areas (Blount, 2015). Bacterial conjugation, genetic recombination, operons have been discovered in *E. coli*, and many metabolic activities such as DNA replication, RNA transcription, and protein synthesis have been enlightened by using *E.coli*. Standard *E. coli* strain used in laboratories is K12 (Idalia, and Bernardo, 2017). Whole genomic sequences of K12 and O157:H7 have been revealed and reported. While the K12 genome is composed of 4200 genes (Blattner et al., 1997), O157:H7's genome is 25% bigger than K12's genome (Lukjancenko, Wassenaar, and Ussery, 2010). This difference is sourced by virulent factors of O157:H7 that are not found on the genome of K12. In fact, K12 has lost its capsule forming feature and adapted to laboratory conditions (Zhang et al., 2007). In modern biological engineering applications, *E. coli* is manipulated and used such as a production factory of desired DNA or protein molecules. High yield and high volume production of these molecules is achieved by recombinant biotechnology applications on *E. coli* bacteria (Rosano, and Ceccarelli, 2014).

Although complete genome sequences of K12 strain of *E. coli* bacteria have been published in 1997, over 300 strains of *E. coli*'s genome are known at the present time (Verma, Qian, and Adhya, 2019). The genome of *E. coli* K12 strain is composed of 4.6 million base pair length, over 4200 protein expressing genes, over 80 tRNA

coding gene regions, and 7 rRNA expressing gene regions (Blattner et al., 1997). However, so far G-quadruplex structures of the *E. coli* genome were not enlightened with details. Considering the functions of G-quadruplexes on mammals, viral and bacterial organisms, investigation of their existence and roles in *E. coli* bacteria exhibit significance in the molecular biology research area. In the research which was conducted and published in the Nucleic Acid Research journal in 2016 by Kaplan and his colleagues, the scientists have performed a phylogenetic analysis about G-quadruplexes on *E. coli* genome. They have realized that there are two conserved G-quadruplex forming motifs (HPGQ1 and HPGQ2) and they have discovered 52 G-quadruplex forming sequences with their developed algorithm. Furthermore the locations of these sequences are introducing the regulatory roles of them since they are localized on either upstream or downstream of various genes such as HrpB, TauD, adK, alx, cstA, pstA, yihS, yhhP etc (Kaplan et al., 2016). As such, in this study the biological roles of these conserved sequences have been investigated by revealing their protein ligands interacting with them, knockin-out of them from *E. coli* genome by CRISPR/Cas9 application, and analysing the alteration on expression level of genes possibly regulated by G-quadruplexes by treating cells G-quadruplex stabilizing ligands.

1.3. The Protein Ligands Interacting with G-quadruplexes

One of the most significant evidence revealing the roles of G-quadruplexes is interactions with observed those non-canonical structures and their protein ligands (Ruggiero, and Richter, 2018). Taking into account that G-quadruplexes are frequently found into regulatory regions of genetic segments, dynamics of G-quadruplex formation exhibit variation with the range of cell types, it's predicted that G-quadruplex structure could be regulated by various types of proteins (Kosiol et al., 2021). Accordingly quite a number of proteins have been reported as interacting with G-quadruplexes in literature (Vlasenok et al., 2018). For instance, it's previously highlighted that telomeric regions of genomes consist of G-quadruplex forming sequences. While these telomeric regions are expected to be 50-100 nucleotides shortened after each replication, many proteins are known as responsible for the capping process which is protecting the genome by binding to DNA ends (Muraki et al., 2012). Furthermore several G-quadruplex binding antibodies and ligands targeting G-quadruplexes on telomeric regions have been developed during the years (Neidle,

2010). Besides, it's also proven that many other cellular telomere-associated proteins could bind and regulate G-quadruplexes on telomeric regions of the genome. For instance, Shelterin is a protein complex constructed with six different proteins which are POT1, TRF1, TRF2, TPP1, TIN2, and RAP1. While the main objective of Shelterin complex is protecting telomeric ends against to be recognized as DNA break, it's discovered that Shelterin complex could bind G-quadruplexes on telomeres and G-quadruplexes are regulated via interactions with POT1-TPP1 complex (Masai, and Tanaka, 2020). In addition various helicase variants have been reported as interacting and unwinding G-quadruplexes such as DHX36, Blooms (BLM), Werner (Wrn), XPB, and XPD (McRae et al., 2017).

In addition, since G-quadruplexes could be formed in promoter regions of several genes frequently, many transcription factors or other proteins are predicted to bind those structures. In particular, the foundation of G-quadruplexes on some significant oncogenes renders those structures to serve as binding sites for various factors in cancer progression (Nakanishi, and Seimiya, 2020). For instance, p53 is one of the most studied tumor suppressor genes since it is found on at least 50% human tumors. Inactivation of p53 through mutations plays an important role in cancer progression. Contrary to, few mutations could cause a special p53 variant and introduce a novel function. Although the role of this novel p53 variant is not clearly explained on cancer progression, it is revealed that it could stabilize G-quadruplex structure by binding to G-C rich DNA sequences (Adámik et al., 2016). In addition it's discovered that nonmetastatic NM23-H2 factor could bind and unwind the G-quadruplex found on c-MYC promoter (Thakur et al., 2009). Furthermore nucleolin is a kind of phosphoprotein overexpressed in overproliferating cells and functional on transcription, chromatin remodelling, biogenesis of ribosome, and apoptosis (Chen, and Xu, 2016). As an interesting discovery, nucleolin protein could bind both *in vitro* and *in vivo* with high affinity and selectivity to G-quadruplex of c-MYC promoter according to other G-quadruplex structures (González et al., 2009). Similarly it's also revealed that poly(ADP-ribose) polymerase 1 (PARP1) and Myc-associated zinc finger (MAZ) could bind to the G-quadruplex structure found upstream of KRAS which is another oncogene (Cogoi et al., 2010).

Apart from them, many chromatin remodelling enzymes have been reported as binding to G-quadruplexes. For instance ATR-X is one of these chromatin remodelling

proteins, it's ability to bind C-G rich sequences forming G-quadruplexes has been proven (Wang et al., 2019). Moreover, the binding of DNA methyltransferase enzymes (DNMTs) which are catalyzing 5-methylcytosine formation of CpG islands of mammalian cells to G-quadruplexes has been demonstrated *in vitro*. For instance DNMT1 is able to bind G-quadruplexes with high affinity and loses its enzymatic activity (Mao et al., 2019). Besides, G-quadruplex formations could be triggered by DNA methylation and many G-quadruplex binding proteins could be recruited to the related site. For instance, that methylation of CpG islands onto the promoter of hTERT gene is inducing G-quadruplex formation and recruiting CCCTC-binding factor (CTCF) in order to upregulate expression (Tikhonova et al., 2021).

Furthermore several G-quadruplex binding proteins have been reported as binding to G-quadruplexes in RNA molecules. It's also revealed that G-quadruplexes in RNAs could be functional on many cellular processes such as alternative splicing, regulation of translation, telomerase activity, and termination of transcription (Lyu et al., 2021). As such various RNA G-quadruplex binding proteins showing the biological roles of G-quadruplexes have been demonstrated such as serine-arginine rich splicing factors (SRSF), U2AF splicing factor, TLS, FRM2, TRF2, and ribosomal protein (Brázda et al., 2014). It's also revealed that telomeric RNA naming as TERRA contains G-quadruplex structures and is recognized by TLS/FUS and TRF2 complex (Huang et al., 2018). Moreover, various hnRNP proteins playing a role in splicing, transporting, and packaging of pre-mRNA molecules have been proven as binding to RNA G-quadruplexes (Huang et al., 2017).

G-quadruplex binding proteins are encountered within a range of organisms from eukaryotes to prokaryotes. For instance in a study conducted by Niu and his colleagues a protein called BmLARK carrying RNA recognition motif has been discovered and demonstrated as binding to G-quadruplex found in the promoter region of *BmPOM2* transcription factor. It's also revealed that variants of LARK belonging to *Bombyx mori*, *Mus musculus*, *Drosophila melanogaster*, and *Homo sapiens* could bind G-quadruplexes of *BmPOM2* and other genes of *B. mori* and *H. sapiens*. The findings show that binding of BmLARK to G-quadruplexes is upregulating expression of related genes. In addition, Kang, et al. have discovered novel non-telomeric G-quadruplex binding proteins in *E. coli*, *Arabidopsis*, yeast, and human and they have

declared that purification, characterization, and identification of these proteins will be conducted on their further studies (Niu et al., 2019).

1.4. The Methods for Revealing G-quadruplex Binding Proteins

G-quadruplex binding proteins and ligands could be revealed and investigated via several bioinformatics analysis and computational tools (Brázda et al., 2018). Contrary to these *in silico* techniques many *in vitro* and *in vivo* approaches have been developed in order to identify G-quadruplex binding proteins during years. Pull-down assay is one of these approaches and it uses related DNA molecules as a bait for binding ligands (Sui, Chen, and Imamichi, 2020). In this technique the secondary structure forming DNA molecules has been designed synthetically and functionalized with biotin molecules. The cells of ligand protein containing organisms are lysed with chemical or physical methods. Once that incubation the biotinylated DNA with cell lysate, ligand-DNA complexes are collected with streptavidin conjugated materials. In the final step of this technique, DNA binding proteins are eluted with high purity from synthetic DNA molecules by dissolving in appropriate buffer conditions (Figure 3) (Meissner, 2019).

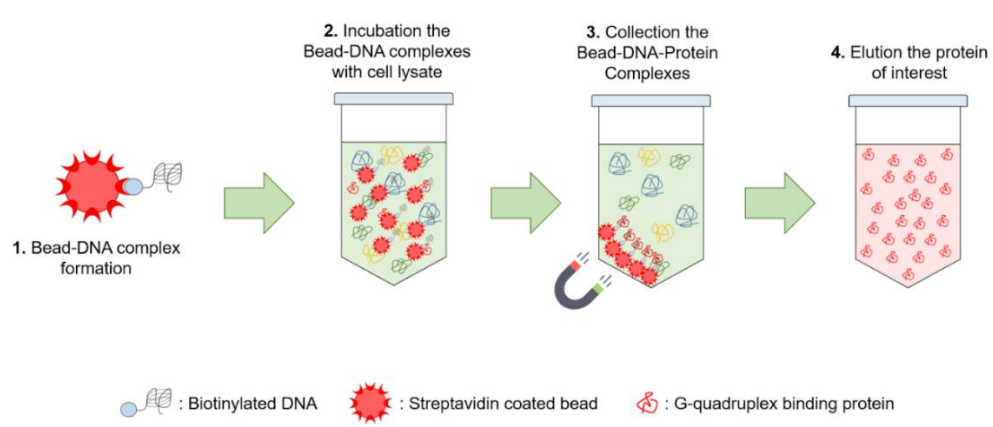

Figure 3. Pull-down assay process that is performed in order to isolate G-quadruplex binding proteins of *E. coli*.

Once that collection of DNA binding proteins, identification and characterization require application of several proteomic techniques based on mass spectrometry. For instance liquid chromatography, gas chromatography, X-ray crystallography, and nuclear magnetic resonance (NMR) are found under the protein identification and characterization techniques (Kaltashov, Bobst, and Abzalimov,

2013). However, one of the most important techniques is Matrix Assisted Laser Desorption/Ionization time-of-flight Mass spectrometry (MALDI-TOF MS) which was developed by German-Japonese scientist group in 1985 (Stanssens et al., 2004). This technique analyzes protein and peptide samples according to their fingerprints. In its working principle, a matrix which absorbs laser energy is used in order to ionise molecules without disrupting the samples. MALDI-TOF MS works with three step: *i)* fixing the samples within appropriate matrix, *ii)* pulsing laser over the samples in order that ablation or desorption of the matrix and samples, *iii)* ionisation of molecules by protonation or deprotonation in order to accelerate within mass spectrometer for analysis (Figure 4) (Graham, 2015). The basic parameter for analysis is mass of ions of the sample. While each ion could have its own mass, the flight time could be detected and the signal could be converted to the TOF mass spectrum. Comparing with databases published in literature, the fingerprints of the samples which are their peaks exhibiting a certain mass/time ratio with special intensities are analyzed and the molecules are identified (Hrabák, Chudác'ková, and Walková, 2013). In the biotechnological research area MALDI-TOF MS technique takes place in biochemistry, organic chemistry, polymer science, microbiology, and medicine in order to identify and detect related molecules (Croxatto, Prod'hom, and Greub, 2012).

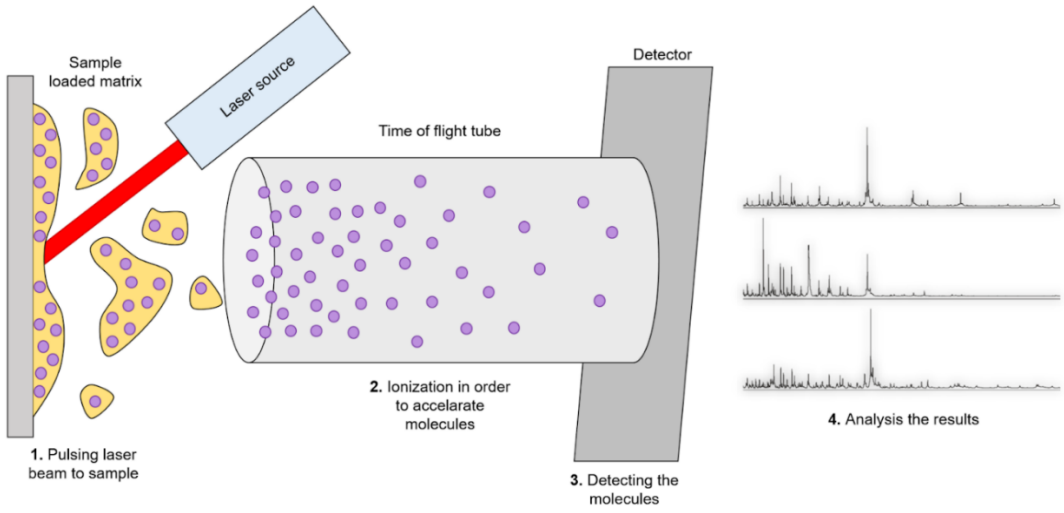

Figure 4. Working principle of MALDI-TOF MS technique which is carried out for characterization of protein of interest.

1.5. Knocking-Out of G-quadruplexes by CRISPR/Cas9 Technology

Since that G-quadruplexes are found closer to genes at either their upstream or downstream, it's predicted that these structures possess regulatory function over the adjacent genes. In another approach, the effect of removing G-quadruplex structures from genomes on products of adjacent genes indicates the biological roles of G-quadruplexes. In order to remove these structures from, a technique with high performance, efficiency and specificity is required. During several years various gene editing techniques have been developed under the biotechnological research area (Khalil, 2020). One of the most important of these techniques is the CRISPR/Cas system, which is a Nobel Prize awarded in 2020 in Chemistry (Westermann, Neubauer, and Köttgen, 2021). CRISPR/Cas systems have been discovered as an immune response of bacterial species against viral organisms. This response is basing on utilization of a special digestion enzyme naming Crispr associated (Cas) enzyme leaded by an RNA complex called guide RNA (sgRNA) composed of trans-activating Crispr RNA (tracrRNA) and Crispr RNA (crRNA). Due to its cleavage activity, gRNA targeted Cas enzyme recognizes and degrades viral genome with sequence specificity (Rath et al., 2015). During the years 11 CRISPR/Cas systems have been revealed as natural mechanisms (Lundgren, Charpentier, and Fineran, 2015). Once these systems had been revealed with rather details, manipulation of CRISPR/Cas systems in genome editing became one of the major efforts of biotechnology. In addition lots of efficient methods and tools were developed in the line of this intent (Anzalone, Koblan, and Liu, 2020).

Genome editing via CRISPR/Cas system is depending on double strand break (DSB) creation on specific sequence of genome of organism and repairing this DSB with natural repair mechanisms which are non-homologous end joining (NHEJ) and homology directed repair (HDR) mechanisms (Adli, 2018). While NHEJ repairs DSB by insertion or deletion (indels) of a few bases, HDR requires template oligonucleotides containing homology arms in order to repair. Through the manipulation of both these mechanisms, genome editing might be performed by creating mutations on desired sequence of the organism's genome (Yoshimi et al., 2021). In the genome editing approach with the CRISPR/Cas system an engineered RNA molecule naming single guide RNA (sgRNA) is designed to mimic gRNA. In addition, Cas9 enzyme belonging to *Streptococcus pyogenes* is frequently used to

create DSBs on the genome (Qin et al., 2020). Fundamentally genome editing with CRISPR/Cas system is performed by following the steps; *i)* designing sgRNAs with computational methods in order to recognize related sequence adjacent to a special sequence which is protospacer adjacent motif (PAM) -short nucleotide sequence such as NAG or NGG-, *ii)* delivery of sgRNA and Cas9 expressing vectors to organism, *iii)* loading of sgRNA into Cas9 enzyme within cytoplasm, *iv)* DSB creation by Cas9 enzyme guided with sgRNA, *v)* repairing of DSB via HDR or NHEJ mechanisms (Figure 5) (Rodríguez-Rodríguez et al., 2019). Various sgRNA designing tools such as Benchling, CHOP-CHOP, Elevation etc. and expression vectors for both sgRNA and Cas9 enzyme such as pCas9, pCRISPR, pREDCas9 etc. have been developed and utilized over several organisms from mammalians to microorganisms in our era (Cui et al., 2018; Fajrial et al., 2020). However it's highlighted in literature that, due to inadequate efficiency of their repair mechanisms, complications are observed in genome editing on microorganisms such as bacteria by CRISPR/Cas systems (Vento, Crook, and Beisel, 2019). Nonetheless, few studies using the tools for expression of desired systems in bacterial organisms have been published (Arroyo-Olarte, Bravo Rodríguez, and Morales-Ríos, 2021).

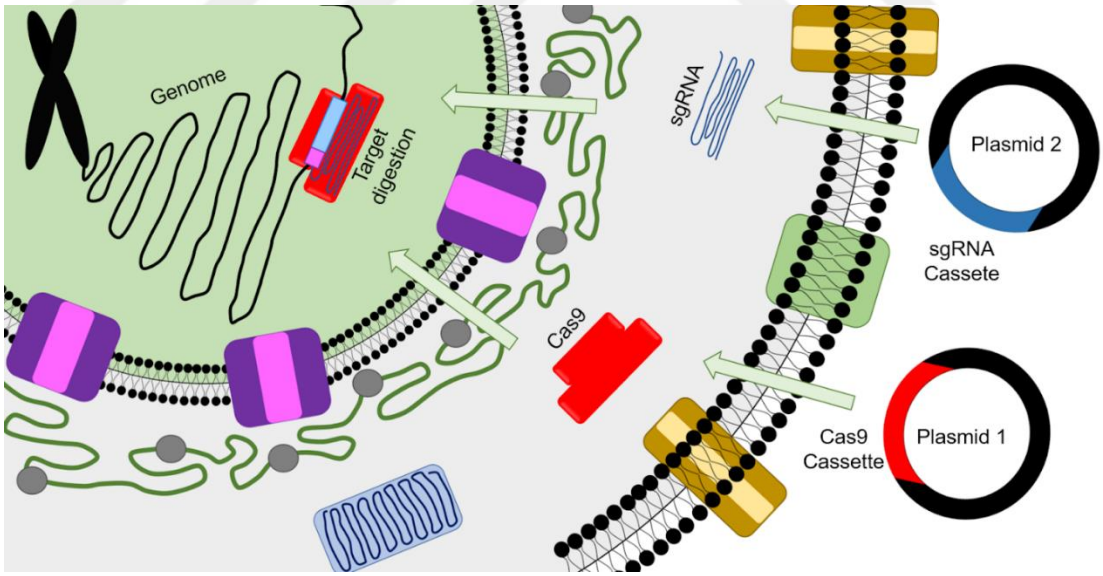

Figure 5. Working principles of CRISPR/Cas9 technique in genome editing.

In literature, Jiang and his colleagues have developed a convenient CRISPR/Cas system that might edit the genome of *S. pneumoniae* and *E. coli* with 100% and 65% efficiencies respectively. During *E. coli* trials, scientists have transformed the Mg1655 strain of *E. coli* bacteria with pCas9 plasmid initially. It's

reported that pCas9 expresses tracrRNA, Cas9 enzyme, and provides chloramphenicol resistance. In second step pCas9 containing Mg1655 bacteria are co-transformed with crRNA inserted pCRISPR plasmid which is providing crRNA leader regions, directed repeats providing BsaI restriction site in order to insert crRNA, and kanamycin resistance and single stranded mutation containing template oligonucleotide which will be used in repairing and introducing desired mutation into bacterial genome. However since that Mg1655 is not sufficiently recombineering, convincing success has not been obtained with this strain. As such, further applications have been conducted with adequate success with HME63 strain which is quite recombineering due to its functions of Gam, Exo and Beta obtained from ƛ-red phage (Jiang et al., 2013). In another study Li, et al. have developed a strategy succeeding iterative genome editing within *E. coli* bacteria. This attractive strategy is basing on delivery of pREDCas9 plasmid which is expressing Cas9 enzyme, recombineering ƛ-red genes, and spectinomycin resistance product. Similarly with Jiang's approach, sgRNA expressing plasmid and template oligonucleotides are delivered to pREDCas9 containing bacteria in order to edit its genome. Furthermore these plasmids are containing L-arabinose inducible plasmid curing systems, thus iterative genome editing might be completed over the same bacterial organisms. Through ƛ-red recombineering system elevated efficiency is obtained with their published CRISPR/Cas9 based genome editing system (Li et al., 2015). In light of these information, the combination of those strategies have been applied in order to remove G-quadruplex forming sequences from *E. coli* in this study. The efficiency of CRISPR/Cas9 applications have been analyzed with blue/white colony screening system, PCR technique depending amplification of the regions that G-quadruplex removed from, and sanger sequencing technique in order to investigate whether insertion or not of desired mutations into related regions of genome.

1.6. Effect of G-quadruplex Forming Agents on Their Biological Roles

It's highlighted that the regulatory role of G-quadruplexes on transcription is depending on its formation to appropriate structure and interactions with their ligand molecules. Furthermore, it's also highlighted that G-quadruplex formation is depending on folding of G-tetrads around a monovalent cation such as sodium (Na^+), potassium (K^+), and lithium (Li^+) (You et al., 2017). Furthermore many agents such as various monoclonal antibodies, BRACO-19, Phen-DC 3, TMPyP4, etc. have been

reported according to their ability to stabilize G-quadruplexes with strong selectivities in literature (Figure 6) (Tian et al., 2018). As such, it's known that the existence of these compounds and different concentrations of desired monovalent cations within cells of organisms might affect the G-quadruplex structures in the genome and their regulatory functions over cellular processes such as replication, transcription etc. Investigating the variations on the level of products of genes closer to G-quadruplexes under the existence of the G-quadruplex stabilizing agents and different concentrations of monovalent cations signifies the biological role of G-quadruplexes.

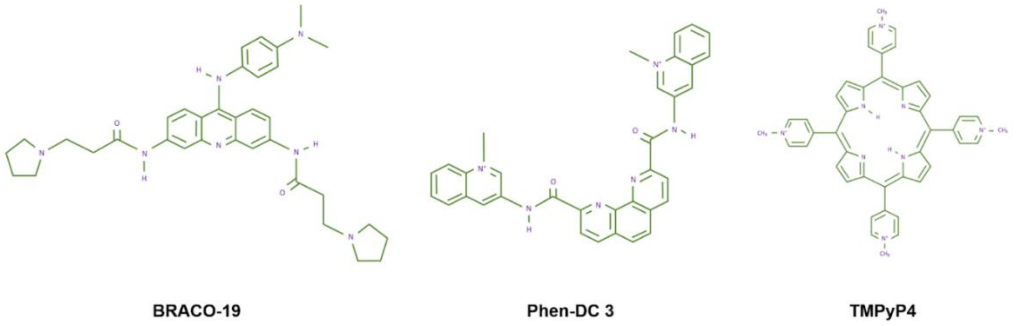

Figure 6. Chemical structures of various G-quadruplex binding molecules; BRACO-19, Phen-DC 3, TMPyP4.

1.7. The Aim of Study

In light of all aforementioned information, the biological roles of novel G-quadruplex forming sequences within the E. *coli* genome discovered in study published in the journal of Nucleic Acid Research in 2016 by Kaplan and his colleagues (Kaplan et al., 2016) are investigated in this research. The first step of this project is depending on characterization of G-quadruplex binding proteins of *E. coli* bacteria through pull-down assay and MALDI-TOF-TOF technique. Accordingly, G-quadruplex forming oligonucleotides have been designed, produced synthetically, and functionalized with biotin molecules. Once that incubation of synthetic oligonucleotide with E. coli lysate, the DNA-protein complexes have been collected by streptavidin coated magnetic beads, and G-quadruplex binding proteins have been eluted with proper chemical buffers. Subsequently, protein samples have been characterized and analyzed with MALDI-TOF-TOF technique and reported considering their scores indicating the possibilities. In the second approach in order to

reveal biological roles of G-quadruplexes in *E. coli*, these structures have been tried to be silenced with CRISPR/Cas9 technique. The G-quadruplex targeting sgRNAs and single stranded template oligonucleotides that will be utilized during repair of DSBs have been designed by Benchling online tool considering their off-target scores. As such, HrpB, tauD, adk, alx, cstA, pstA, yihS, and yhhP genes have been decided to be silenced considering study of Kaplan, et al. In addition, efficiency of CRISPR/Cas9 technique in genome editing within *E. coli* bacteria has been analyzed with blue/white colony screening through targeting lacZ gene. Furthermore, once that trials on Mg1655 strain have been completed, in order to enhance efficiency of CRISPR/Cas9 applications, further trials have been conducted on recombineering HME63 strain. In the last approach of the study, *E. coli* cells will be incubated with different concentrations of G-quadruplex forming monovalent cations and G-quadruplex stabilizing agents. Effect of constant formation of G-quadruplex on mRNA level of the genes shown in paper of Kaplan, et al. will be analyzed with qPCR technique. Therefore the possible functions of G-quadruplexes on transcription of related genes will be revealed. In this scope, the biological roles of putative G-quadruplex structures of the *E. coli* genome have been enlightened and will be proven with this research by applying three different approaches. While their roles might be revealed with few other techniques, the findings obtained from our research will open a new door into the significance of G-quadruplex structures and their roles.

CHAPTER 2: METHODS

2.1. Characterization of the G-quadruplex Binding Proteins

2.1.1. Bacterial Growth

Luria Broth (LB) media which is frequently used in bacterial growth was prepared according to the suggestion of supplier company (Sigma-aldrich). 200 µl stock Mg1655 strains of E. coli bacteria were inoculated into 8 ml LB, and starter culture was incubated at 37 ^{0}C, 200 rpm overnight. Once that sufficient turbidity had been observed, the culture was diluted with 1:100 ratio into fresh LB media. While fresh culture had been incubated at 37 ^{0}C, 200 rpm, variations on optical density (OD_{600}) value was analyzed with 30 minutes ranges by measuring absorbance value of culture on 600 nm ultraviolet wavelength. Incubation was holded when OD_{600} value had achieved to 0.6-0.8 interval now that the culture had reached exponential phase of the growth curve of bacterial growth. 2500 ml of the culture was separated to falcon tubes and collapsed by centrifugation with 4000 rpm for 20 minutes. The bacterial pellets were stored at -80 ^{0}C in order to be used for further steps of lysis and total cytoplasmic extraction. Besides, remaining bacterial culture on exponential phase was stored at -80 ^{0}C in 25% (v/v) glycerol.

2.1.2. The Buffers Used in Characterization the G-quadruplex Binding Protein

Bacterial lysis, total cytoplasmic protein extraction, and pull-down assay were carried out by following the method published by Jutras, et al. in 2012. As such two distinct chemical buffers naming THES (50 mM Tris HCl (pH 7.5), 10 mM EDTA, 20% (w/v) sucrose, 140 mM NaCl) and 5X BS (50 mM HEPES, 25 mM CaCl2, 250 mM KCl, 60% (v/v) glycerol) were prepared as stock solutions. BS/THES buffer were prepared as containing 44.3% THES buffer, 20% 5X BS buffer, and 35.7% nuclease free water. In addition Lysis Buffer which is used in chemically lysis of bacterial cells were prepared according to Jutras and his colleagues' recipes. As such, 10 ml fresh Lysis Buffer were prepared as containing 400 µl B-Per II reagent, 9.6 ml BS/THES Buffer, 0.1 g Lysozyme, 1 mini tablet protease inhibitor cocktail, and 1 mini tablet phosphatase inhibitor cocktail for each trial. Additionally 2X B/W Buffer (10 mM Tris HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl), and Elution Buffers with different concentrations (E1: 25 mM Tris HCl, 100 mM NaCl, E2: 25 mM Tris HCl, 200 mM

NaCl, E3: 25 mM Tris HCl, 300 mM NaCl, E4: 25 mM Tris HCl, 500 mM NaCl, E5: 25 mM Tris HCl, 750 mM NaCl, E6: 25 mM Tris HCl, 1 M NaCl) were prepared in order to be used in pull-down assays. Biotin functionalized DNA was dissolved in Tris-EDTA (TE) Buffer.

2.1.3. Total Cytoplasmic Protein Extraction from E. coli Cells

Bacterial cell pellets stored in -80 ^{0}C were used in extraction of total cytoplasmic proteins. Thawed bacterial cell pellets in each tube were resuspended by vortexing and incubated at room temperature and 100 rpm for 30 minutes in 1 ml Lysis Buffer which is optimized value according to subsequent trials. Cell lysates had been centrifuged at 17,000 g, 4 ^{0}C for 30 minutes, and the supernatant containing total cytoplasmic proteins were stored at -20 ^{0}C in order to be used in further experiments.

2.1.4. Pull-Down Assay

In order to collect protein bound biotin functionalized DNA molecules, commercially Dynabead M-280 Streptavidin beads were resuspended homogeneously in their solution. Once the beads resuspended in 200 µl solution had been collapsed under the magnetic area, the solution was discarded. Bead pellets were washed with 500 µl 2X B/W buffer three times, and washed beads were resuspended in 200 µl 2X B/W buffer for further applications. Afterwards 200 µl biotinylated DNA solution (200 ng/µl stock) were added and samples were incubated at room temperature 200 rpm for 20 minutes in order to create DNA-Bead complexes through interactions between biotin and streptavidin molecules. Once that beads had been collapsed by the magnetic area, the incubation step was repeated so that all beads in solution might be saturated with biotinylated DNAs. DNA-Bead complexes were collapsed, and washed with 500 µl TE buffer three times, with 500 µl BS/THES Buffer two times, and with 500 µl BS/THES Buffer containing 5 µg/ml calf thymus (CT) DNA once. The collapsed DNA-Bead complexes after the washings were resuspended in 200 µl BS/THES buffer, and incubated with 700 µl bacterial cell lysate as well as 5 µl CT DNA (5 µg/ml stock) in order to inhibit non-specific protein bindings at room temperature 500 rpm for 30 minutes. Incubation step had been repeated until DNA-Bead complexes were saturated with G-quadruplex binding proteins. Once the incubation Protein-DNA-Bead complexes were collapsed by the magnetic area and washed with 500 µl BS/THES buffer containing 5 µg/ml CT DNA five times, and with 500 µl BS/THES

buffer two times. The G-quadruplex binding proteins were collected from Protein-DNA-Bead complexes by incubating elution buffers containing different concentrations of NaCl. As such, the collapsed Protein-DNA-Bead complexes were resuspended in 120 µl E1 buffer containing 25 mM Tris, 100 mM NaCl, and incubated at room temperature, 500 rpm for 5 minutes. Since that appropriate conditions for eluting G-quadruplex binding proteins with high purity is unknown, elution steps were repeated with remaining elution buffers, and protein elutions were stored at -20 ^{0}C for characterization experiments. In addition, now that Dynabead M-280 Streptavidin beads are reusable, biotinylated DNA molecules were removed from beads by breaking interactions between biotin and streptavidin molecules through incubation DNA-Bead complexes resuspended in 35 µl nuclease free water at 70 ^{0}C for 10 minutes.

2.1.5. Circular Dichoism (CD) Spectrometry

In order to analyze G-quadruplex structures in synthetic biotinylated DNA molecules, they were dissolved as 5 µM concentration in the BS/THES buffer containing 150 mM HEPES, and denatured by heating at 90 ^{0}C for 20 minutes. Once the G-quadruplexes had been formed by cooling the DNA molecules at room temperature, samples were scanned with circular dichroism (CD) spectrometry in 2 mm quartz cuvettes at the range of 220-350 nm wavelengths.

2.1.6. SDS-Page Electrophoresis

Once that pull-down assay had been completed, eluted G-quadruplex binding proteins were analyzed with SDS-Page electrophoresis technique. As such, two layer polyacrylamide gel consists of resolving gel (10% Acrylamide/bisacrylamide, 400 mM Tris HCl, 1% (w/v) sodium dodecyl sulfate (SDS), 1% (w/v) ammonium persulfate (APS), 1:1000 (v/v) TEMED) -in order to separate protein samples according to their sizes- and stacking gel (5% Acrylamide/bisacrylamide, 400 mM Tris HCl, 1% (w/v) SDS, 1% (w/v) APS, 1:1000 (v/v) TEMED) -in order to create proper bands by stacking proteins- were prepared for protein analysis. Protein samples were dissolved in 5X SDS Loading Buffer (10 mM β-Mercaptoethanol, 1% (w/v) Bromophenol blue, 50% (v/v) glycerol, 8% (w/v) SDS, 250 mM Tris HCl) with 1:5 ratio, and denatured by incubation at 90 ^{0}C for 10 minutes. While the protein ladder had been loaded into the first well of gel by following the suggestion of supplier

company (Sigma-Aldrich), protein samples were loaded to remaining wells. Samples were runned in stacking gel by applying 60 V and in resolving gel with 120 V electrical current. Once that running had been completed, gel was stained with coomassie brilliant blue overnight. In order to observe bands as sharp, gel was incubated in distilled water for de-staining. Gel results were analyzed with UV visualization technique.

2.1.7. MALDI-TOF-TOF

The characterization of isolated bacterial protein was carried out with MALDI-TOF-TOF technique. As such, in order to increase concentration of protein, isolation steps were repeated constantly, gathered in a single tube and lyophilized. Protein pellets and salt crystals belonging to elution buffers were observed after lyophilization. MALDI-TOF-TOF technique was performed by procurement of services from Mass Spectrometry Analysis Laboratory of Prof. Dr. Bekir Salih in Hacettepe University. As such, proteins were exposed to tryptic digestion and investigated by loading into Bruker Rapiflex Maldi system. Protein characterization results were analyzed by using SwissProt and NCBInr databases.

2.2. Knocking-Out of G-quadruplexes by CRISPR/Cas9 Technology

2.2.1. pREDCas9, pCRISPR, and pCas9 Isolation

The bacterial strains including pREDCas9, pCRISPR, and pCas9 plasmids had been grown within 50 µg/ml spectinomycin, 50 µg/ml kanamycin, and 25 µg/ml chloramphenicol containing LB media respectively by incubating at 37 ^{0}C 270 rpm overnight. In order to store plasmid including bacteria, growth cultures were inoculated into fresh media and incubated until they reached the exponential phase. Afterwards they were resuspended in 25% glycerol and stored at -80 ^{0}C for long term. pREDCas9, pCRISPR, and pCas9 plasmids were isolated from overnight growth culture by using commercial MACHEREY-NAGEL NucleoSpin Plasmid DNA Purification kit. In order to visualize the plasmids, agarose gel electrophoresis technique were performed. As such, 1% agarose gel was prepared by dissolving 1 g agarose in 100 ml TAE Buffer, DNA ladder and plasmid samples were loaded into wells of gel, and runned under 100 A for 2 hours.

2.2.2. sgRNA and Template Oligonucleotide Design

The oligonucleotide sequences that will be inserted into pCRISPR plasmid and transformed as sgRNA as well as the template DNAs which will be used during homology directed repair after DSB creation were designed by Benchling online tool. As such G-quadruplex forming sequences found on either upstream or downstream of totally 8 gene (hrpB, TauD, yhhP, yihS, adK, alx, pstA, cstA) and lacZ gene now that it enables to display genome editing efficiency with blue/white colonies technique were targeted with at least two distinct regions. Once that oligonucleotide had been designed they were purchased from Sente BioLab company.

2.2.3. pREDCas9 transformation

While G-quadruplex editing on genome level were performed on HME63 strain of *E. coli* in order to enhance efficiency now that it's recombineering strain, lacZ knock-out were performed on genome of Mg1655 strain of *E. coli* since that lacZ gene is mutant in genome of HME63 strain. In this scope, pREDCas9 plasmid was transformed to both strains through heat shock transformation technique. As such, bacteria were grown at 37 ^{0}C 270 rpm overnight, inoculated into fresh media, and incubated at same conditions until they reached an exponential phase. Afterwards 10 ml bacterial cultures were collapsed with centrifugation at 4C and 4000 rpm for 20 minutes. Once the supernatant had been removed, cells were resuspended in 10 ml sterile ice col 0.1 M CaCl2, and incubated on ice for 30 minutes. Once that incubation had been completed cells were collapsed with centrifugation at 4C and 4000 rpm for 20 minutes, and cells were resuspended in 5 ml sterile ice cold 0.1 M CaCl2 15% glycerol solution. The prepared competent cells were aliquoted to fresh tubes with 50 µl volume and stored at -80 ^{0}C. During heat shock transformation 50 µl competent cells were incubated on ice for 30 minutes, 100 ng pREDCas9 were added into the tubes, and cells were incubated with pREDCas9 on ice for 30 minutes as well. Once that incubation had been completed, cultures were exposed to heat shock by incubating at 42 ^{0}C for 45 seconds and on ice for 2 minutes. Thus cells were transformed with pREDCas9, they were resuspended in rich S.O.C. media, incubated at 37 ^{0}C for 45 minutes, spreaded on 50 µg/ml spectinomycin containing LB-Agar plates, and single colonies indicating pREDCas9 Mg1655 and HME63 strains were observed after overnight incubation at 37 ^{0}C.

2.2.4. pCRISPR Digestion and Gel Purification

In order to insert sgRNA forming oligonucleotide, DR regions should have been removed from pCRSISPR through digestion with the BsaI restriction enzyme. As such isolated pCRISPR plasmids were digested with the BsaI restriction enzyme by following recommended protocol of supplier company (ThermoFisher Scientific). The digestion mix was prepared with 15 µl pCRISPR (0.5-1 µg), 2 µl Buffer G, 1 µl BsaI enzyme, and 3 µl nuclease free water. Afterwards digestion mixes were incubated at 37 ^{0}C and 350 rpm for 16 hours, and BsaI enzyme was inactivated by incubating at 65 ^{0}C and 350 rpm for 20 minutes. Once that digestion had been completed, digested pCRISPR plasmids were loaded into 1% agarose gel and runned under 100 A for 2 hours. Digested pCRISPR plasmids were visualized by UV visualization, extracted from 1% agarose gel by using commercial MACHEREY-NAGEL NucleoSpin Gel and PCR Clean-up kit, and stored at -20 ^{0}C for further steps.

2.2.5. sgRNA Phosphorylation and Annealing

Designed sgRNAs consist of distinct forward and reverse oligonucleotides. Those sgRNA forming oligonucleotides have to be phosphorylated and annealed in order to be inserted into digested pCRISPR plasmids. As such, sgRNA forming forward and reverse oligonucleotides were phosphorylated with T4 polynucleotide kinase in a single test tube by following the recommended protocol of supplier company (Applied Biological Materials). Phosphorylation mixes were prepared with 1 µl forward oligonucleotide (100 µM), 1 µl reverse oligonucleotide (100 µM), 2.5 µl 10X T4 Polynucleotide Kinase Reaction Buffer, 2.5 µl ATP solution (10 mM), 1 µl T4 Polynucleotide Kinase, and 17 µl nuclease free water. Phosphorylation mixes were incubated at 37 ^{0}C for 10 minutes and T4 polynucleotide kinases were inactivated by incubating at 65 ^{0}C for 20 minutes. Once that phosphorylation had been completed, 2.5 µl 1 M NaCl solution dissolved in nuclease free water was added into reaction mixes; the solutions were incubated at 95 ^{0}C for 5 minutes and at room temperature in order to annealing of oligonucleotides. Afterwards annealed oligonucleotides were diluted 10 times for ligation into digested pCRISPR plasmids.

2.2.6. Ligation of Digested pCRISPR and Annealed sgRNA Forming Oligonucleotides

Annealed sgRNA forming oligonucleotides were inserted into BsaI digested pCRISPR plasmids through ligation. As such, ligation was performed according to Jiang and his colleagues' protocol. The ligation mixes were prepared with 1 µl annealed sgRNA forming oligonucleotides, 1 µl BsaI digested pCRISPR plasmid, 2 µl 10X T4 Ligase Buffer (New England BioLabs), 1 µl T4 Ligase (New England BioLabs), and 15 µl nuclease free water. Reaction mixes were incubated at room temperature for 2 hours and delivered to competent DH5α cells by heat shock transformation. sgRNA:pCRISPR plasmids were isolated from overnight growth DH5α cultures with MACHEREY-NAGEL NucleoSpin Plasmid DNA Purification kit.

2.2.7. Transfer of sgRNA:pCRISPR and Template DNAs to pREDCas9 Containing Bacteria

sgRNA inserted pCRISPR plasmids (sgRNA:pCRISPR) and template oligonucleotides were co-transferred to bacterial cells through electroporation. While lacZ sgRNA:pCRISPR were transferred to pREDCas9 containing Mg1655 strain, remaining G-quadruplex targeting sgRNA:pCRISPRs were delivered to pREDCas9 containing HME63 strain of *E. coli*. As such, *E. coli* cultures which at exponential phase were incubated at 42 ^{0}C for 15 minutes in order to induce ƛ-red genes of pREDCas9 plasmids. Afterwards cells were incubated on ice for 20 minutes and collapsed with centrifugation at 4000 rpm for 20 minutes. Cell pellets were washed twice with ice cold sterile water and 20 fol concentrated by resuspending in lower volume of sterile water. Electroporation reaction mixes were prepared with 50 µl culture, 100 ng sgRNA:pCRISPR, and 100 ng template DNA. Reaction mixes were added into 2 mm electroporation cuvettes and reactions were conducted with pre-set protocol providing electrical pulses with 2500 V for 25 msec of Gene Pulser Xcell Electroporation System. Transformed bacterial cells were spreaded over LB-Agar plates containing 50 µg/ml spectinomycin and 50 µg/ml kanamycin. Additionally in order to reveal genome editing efficiency blue/white colonies technique was performed by spreading lacZ KO Mg1655 strains over the LB-Agar plates containing 0.1 mM IPTG, 40 µg/ml X-gal, 50 µg/ml spectinomycin, and 50 µg/ml kanamycin.

2.2.8. PCR Applications for Control of sgRNA Insertion

The efficiency of sgRNA insertion into BsaI digested pCRISPR was analyzed with Touch-down PCR technique. While reverse oligonucleotide of related sgRNA was used as reverse primer, pCRISPR recognizing forward primer were designed in order to amplify the region including sgRNA forming oligonucleotides within pCRISPR plasmid. PCR reaction mixes were prepared with 2 µl sgRNA:pCRISPR, 2 µl forward and reverse primer mixes (10 µM), 6 µl distilled water, and 10 µl 2X commercial PCR MasterMix. Meanwhile these reaction mixes were prepared using pCRISPR plasmid instead of sgRNA:pCRISPR as control. Now that optimum Tm values are not known, Touch-down PCR application was carried out and results were analyzed in 1% agarose gel.

2.2.9. PCR Applications for Control of Genome Editing

In order to analyze genome editing efficiency of CRISPR/Cas9 technique the colony PCR application was performed. As such, overnight growth colonies after genome modification were picked and inoculated into a 20 µl TE buffer. Colonies were incubated at 100 ^{0}C for 20 minutes and centrifuged at 13000 g for 2 minutes. Genomes containing supernatants were taken into fresh test tubes and used in PCR application. Theoretically, after genome editing with CRISPR/Cas9 technique approximately 80 bp containing G-quadruplexes will be removed from the bacterial genome. In this scope, sequencing primers recognizing related regions of the genome were designed. As such, PCR reaction mixes were prepared with 3 µl genome of colonies, 2 µl forward-reverse sequencing primer mixes (10 µM), 5 µl distilled water, and 10 µl 2X commercial PCR MasterMix. Meanwhile these reaction mixes were prepared with wild type HME63 genome instead of genome of colonies as control. Now that optimum Tm values are not known, Touch-down PCR application was carried out and results were analyzed in 2.5% agarose gel.

CHAPTER 3: RESULTS

3.1. Characterization of the G-quadruplex Binding Proteins

3.1.1. Bacterial Growth

Cultivating the large amount of bacterial culture had been required in order to extract total cytoplasmic proteins from the E. *coli* Mg1655 strain. In this scope, as in "method" section, stock bacterial cultures were inoculated to LB media as starter culture, incubated overnight, and diluted with 1:100 ratio to fresh LB media. Culture had been incubated until OD_{600} reached to range between 0.6-0.8 so that achieving to exponential phase that is described as most efficient phase of bacterial growth curve. The increasing on OD_{600} value and observed turbidity demonstrated the proper growth of bacterial culture. Once that culture had been obtained at desired phase, totally 2 L culture was separated to 50 ml falcon tubes and collapsed with centrifugation at 4000 rpm for 20 minutes in order to be used on further steps. Cultivated and centrifuged bacteria are demonstrated in Figure 7. These obtained bacterial pellets were stored at -80 ^{0}C for following applications.

A)

B)

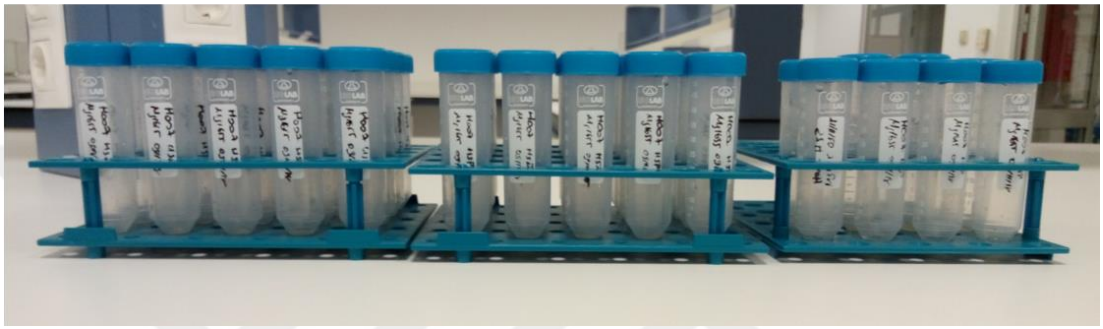

Figure 7. A) 2 L Bacterial culture on exponential phase which were grown to 0.6-0.8 of OD_{600} value, B) Bacterial pellets which are collapsed by centrifugation on 4000 rpm for 20 minutes.

3.1.2. Extraction of Total Bacterial Cytoplasmic Proteins

The bacterial pellets that had been collapsed by centrifugation and stored at -80 ^{0}C were thawed at room temperature in order to be used in total cytoplasmic protein extraction. 250 µl lysis buffer were added into each 50 ml falcon tube, and optimum amount of lysis buffer were identified by adding an extra 250 µl on each trial. In order to lyse the bacterial cells and extraction of total cytoplasmic proteins, 1 ml lysis buffer was identified as sufficient and optimal amount. As such, bacterial pellets were broken in 1 ml lysis buffer by vortexing, and resuspended cultures were incubated at room temperature at 100 rpm for 30 minutes. Thus all cytoplasmic proteins were extracted and solved in the solution. Samples were centrifuged at 4 ^{0}C at 17000 g for 30 minutes, and the phase separation between pellets formed by debrisis and supernatant containing cytoplasmic proteins were observed. Supernatant solutions from each sample were taken into fresh containers and stored at -20 ^{0}C for following applications. In each trial fresh protein extracts were used and the samples stored for long term were not used in pull-down assays.

3.1.3. Pull-down Assay

The conducted pull-down assay introduces significant results about whether there is a bacterial protein that could interact with synthetic oligonucleotide designed as G-quadruplex forming sequence or not. As such, G-quadruplex forming sequences repetitive in E. coli genome were designed, functionalized with biotin molecule at 5' terminal (5'/Biotin/ACG GAC GGT CCC CTC GCC CCT TTG GGG AGA GGG TTA GGG TGA GGG GAA 3') and obtained commercially. Lyophilized DNA samples were reconstructed in TE buffer as recommendation of supplier company so that to be 100 µM concentratioın. During pull-down assays, concentration of DNA samples were reduced to 200 ng/µl, and desired interactions between Dynabead M-280 Streptavidin beads and biotinylated DNA were achieved by incubating as highlighted in the "method" section.

Bead-DNA complexes were incubated with extracted cytoplasmic proteins within the BS/THES buffer at room temperature. Circular Dichroism (CD) spectrometry technique were carried out in order to analyze G-quadruplex structures formed on DNA sequence in these solution. As such, CD-sprectrometry were performed for G-quadruplex forming DNAs in both BS/THES buffer and 150 mM KCl 10 mM HEPES buffer which is known as G-quadruplex inducer (Figure 8). While CD spectrometry is a kind of secondary structure characterization technique for biomolecules, it's working principle is based on analysis the polarization of molecules under certain wavelength. The graph which is obtained from CD spectrometry shows the ellipticity (absorbance value with a factor of 32.98), and wavelength of light exposed in analysis. The peaks indicating G-quadruplex formation which are minimum at 240 nm and maximum at 260 nm were observed on both samples. This indicates the G-quadruplex structure was also obtained in BS/THES buffer.

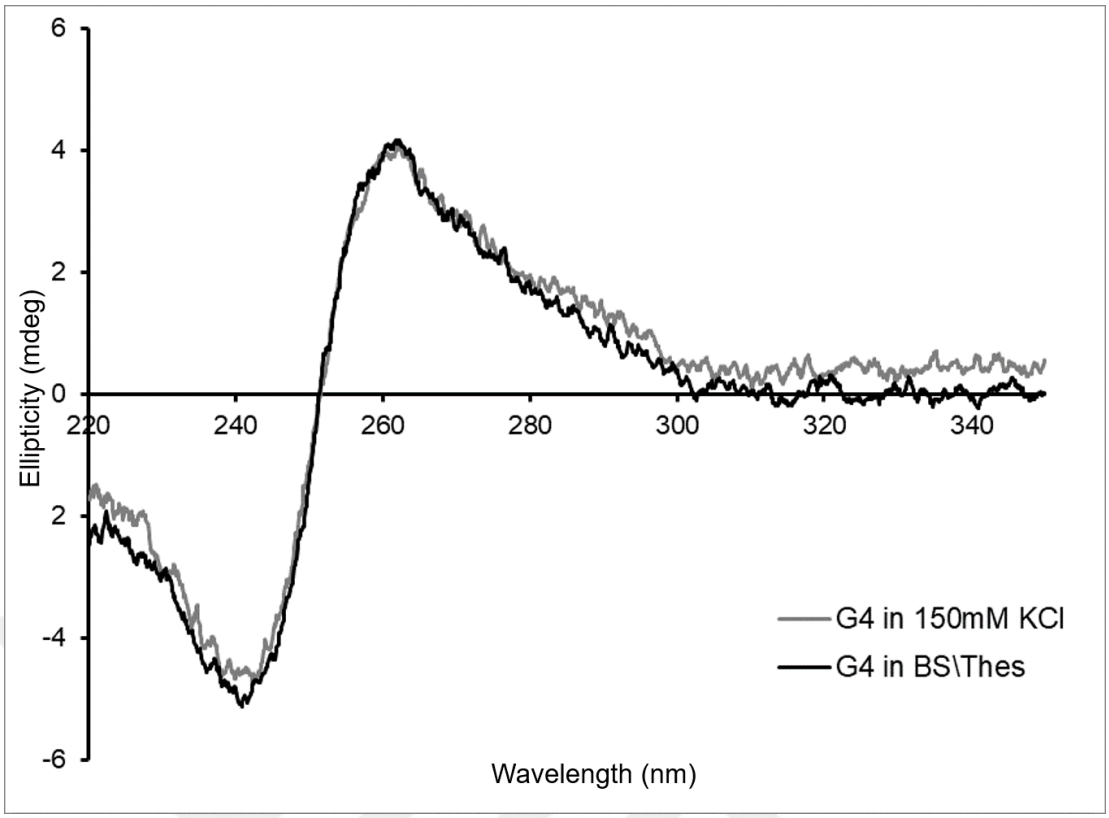

Figure 8. Circular dichroism spectrums of G-quadruplex forming biotinylated DNA samples in BS/THES and 150 mM KCl 10 mM HEPES buffers, respectively.

During pull-down assay CT-DNA were used in order to inhibit nonspecific interactions. Furthermore, for saturation of biotinylated DNA, they were incubated with large amounts of protein extracts. Optimization studies showed that maximum interactions had been observed on incubation with 10 ml protein extracts obtained from bacterial lysis. Once the necessary incubations, proteins were eluted from protein-DNA-Bead complexes with different elution buffers (E1: 25 mM Tris HCl, 100 mM NaCl, E2: 25 mM Tris HCl, 200 mM NaCl, E3: 25 mM Tris HCl, 300 mM NaCl, E4: 25 mM Tris HCl, 500 mM NaCl, E5: 25 mM Tris HCl, 750 mM NaCl, E6: 25 mM Tris HCl, 1 M NaCl) since that the optimum NaCl concentration was not known. While magnetic beads were removed in the meantime protein samples containing supernatants were collected after pull-down assay. The protein samples were stored at -20 ^{0}C for following applications. In addition, biotinylated DNAs were dissolved in nuclease free water after breaking interactions between biotin and streptavidin by heating the molecules at 70 ^{0}C. The following studies demonstrated that pre-used biotinylated DNAs and magnetic beads could be used on next experiments with less efficiency.

3.1.4. SDS-Page Results

SDS-Page technique were carried out in order to reveal whether there is a protein or not that could interact with G-quadruplex forming sequence whose function is to be investigated after the elution steps of pull-down assays. As such, polyacrylamide gel consists of stacking and separating gels which are containing distinct concentrations of acrylamide/bisacrylamide were prepared. The samples obtained from elutions were dissolved in 5X loading buffer with 1:4 ratio. Afterwards, proteins were denatured by heating the samples at 95 ^{0}C for 10 minutes. While commercial protein marker were loaded into the first well of gel, samples were loaded to remaining wells, and *E. coli* bacterial lysates containing total cytoplasmic proteins were loaded into the last well of polyacrylamide gel. Samples were run under 60 V and 120 V in stacking and separating gels respectively. The SDS-Page results obtained by UV visualization after coomassie brilliant blue staining is demonstrated in Figure 9. The results showed that G-quadruplex binding protein were obtained by elution with E3 buffer containing 300 mM NaCl. Molecular weight of obtained protein was identified as 150 kDA compared to commercial protein marker. Consequently, it is revealed that a G-quadruplex binding protein with great molecular weight were discovered, and demonstrated with SDS-Page technique.

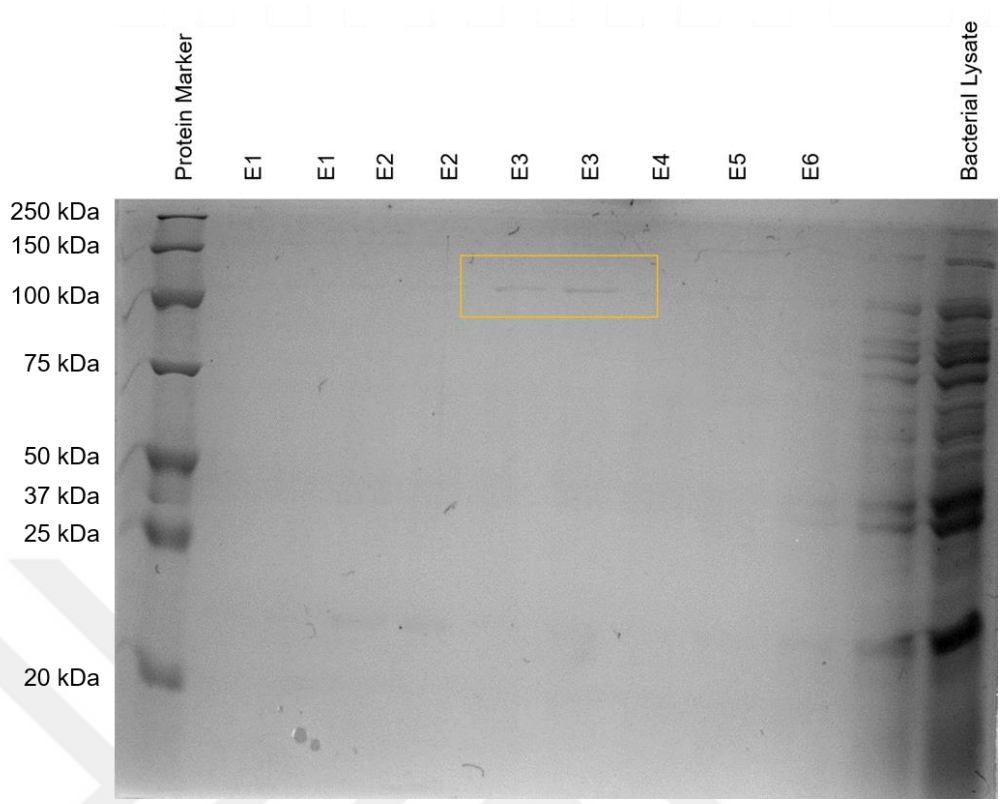

Figure 9. SDS-Page results. Protein ladder, G-quadruplex binding protein with eluted in an elution buffer containing 300 mM NaCl, and total cytoplasmic proteins from bacterial lysis.

Gel image was cropped and its background was removed with ImageJ software. The profiles of protein bands found in the E3 buffer were analyzed. The profiles show single bands belonging to related protein (Peterson T). The average of the values observed in profiles were computed in order to analyze their purity (Figure 10). Standard deviation (σ) of data were computed, it was identified that according to normal distribution 95% of values had been distributed between $\mp 2\sigma$ of mean, and in light of this results existence of proteins were proven. Furthermore, 100% purity were observed in both bands by comparing band profiles with outlier values on remaining regions of the gel image. This purity value was reduced when ignoring standard deviation of data to 95.5% and 96.5% for both bands.

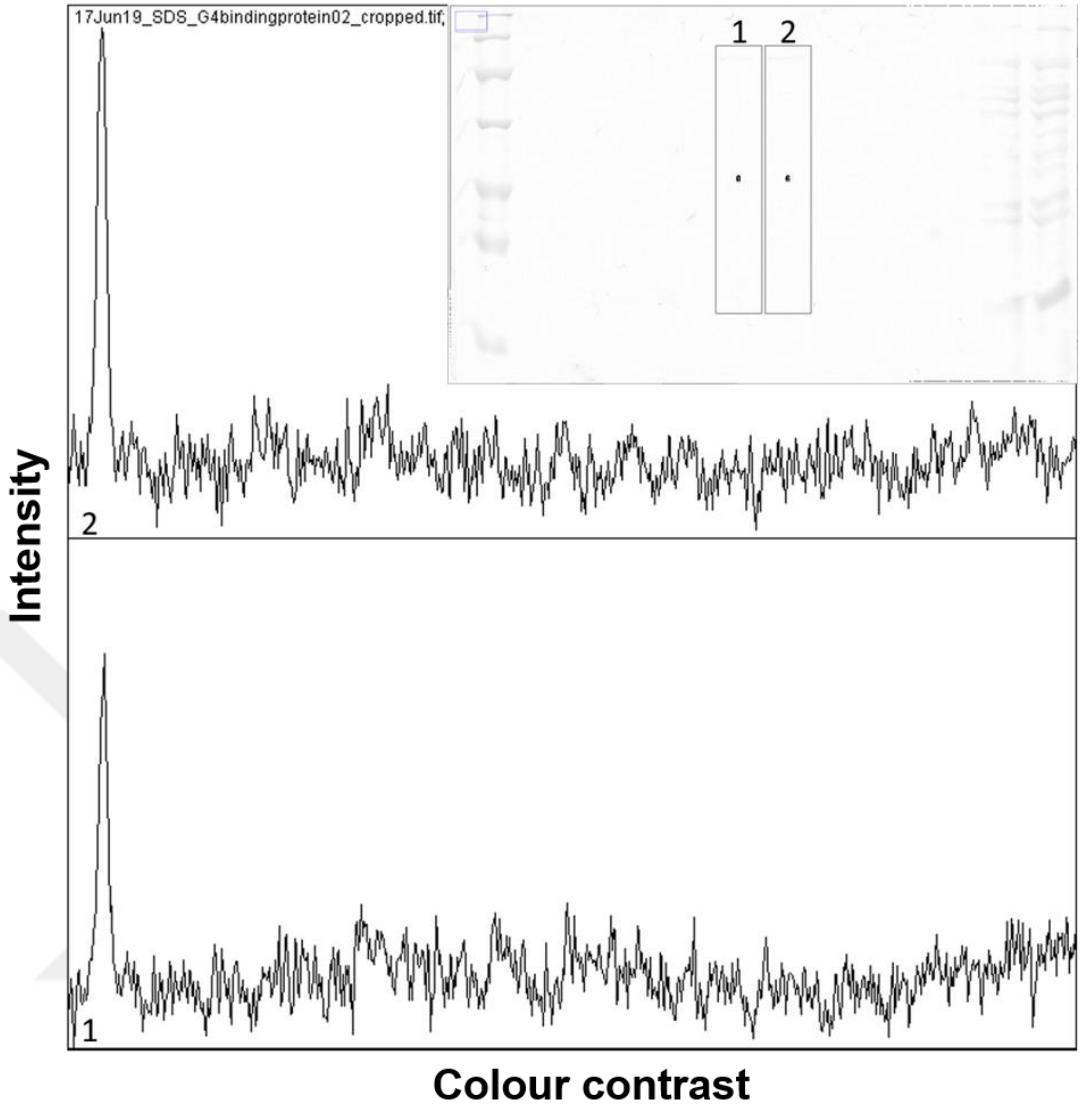

Figure 10. Band profiles of E3 elutions. The results were obtained via removing background from gel image and processing data for both bands on ImageJ software.

3.1.5. MALDI-TOF-TOF Results

It's expected that isolated protein samples by pull-down assays indicate the function of G-quadruplex structures in the *E. coli* genome. As such, large amounts of protein samples were produced within certain periods. In order to increase protein concentration, lyophilization techniques were carried out and protein pellets were observed in test tubes. MALDI-TOF-TOF analysis of the proteins were conducted in Mass Spectrometry Analysis Laboratory of Prof. Dr. Bekir Salih in the Chemistry Department of Hacettepe University. Samples were loaded to Bruker Rapiflex MALDI-TOF system after digestion with trypsin enzyme. The specific peaks of

protein samples and trypsin enzymes were observed on related intervals. Therefore, it was concluded that protein concentration was sufficient for analysis and samples were not containing contamination since that any nonspecific bands had not been observed. The results were analyzed in Swiss-Prot and NCBInr databases by detecting the matches between fingerprint peaks of the sample with published protein peaks. Table 1. and Table 2. are containing the results obtained from analysis with Swiss-Prot and NCBInr databases respectively.

Table 1. The results of analysis isolated protein by Swiss-Prot database after characterization with MALDI-TOF-TOF technique.

Fam.	**Mem.**	**Dat.**	**Sco.**	**Mass**	**Desc.**
3	1	Swiss-Prot	521	96580	Aldehyde-alcohol dehydrogenase
6	1	Swiss-Prot	379	155918	DNA-directed RNA polymerase subunit beta’
2	1	Swiss-Prot	775	37292	Outer membrane protein A
1	1	Swiss-Prot	1340	43427	Elongation factor Tu 1
7	1	Swiss-Prot	376	61235	30S ribosomal protein S1
11	1	Swiss-Prot	278	77704	Elongation factor G
4	1	Swiss-Prot	461	18963	Single-stranded DNA-binding protein
5	1	Swiss-Prot	389	18240	PTS system glucose-specific EIIA component
9	1	Swiss-Prot	314	9529	DNA-binding protein HU-alpha
10	1	Swiss-Prot	296	21987	Spermidine N(1)-acetyltransferase

Table 2. The results of analysis isolated protein by NCBInr database after characterization with MALDI-TOF-TOF technique.

Fam.	Mem.	Dat.	Sco.	Mass	Desc.
3	1	NCBInr	376	96579	Bifunctional acetaldehyde-CoA/alcohol dehydrogenase
3	2	NCBInr	353	96594	Bifunctional acetaldehyde-CoA/alcohol dehydrogenase
3	3	NCBInr	347	96688	MULTI: Bifunctional acetaldehyde-CoA/alcohol dehydrogenase
5	1	NCBInr	333	155928	Polymerase beta',RNA
2	1	NCBInr	548	37327	Chain A, The Host Outer Membrane Proteins Ompa and Ompc
1	1	NCBInr	1217	32279	Elongation factor Tu
1	4	NCBInr	1144	44993	Elongation factor Tu
1	3	NCBInr	1163	25626	Hypothetical protein L668_02045
4	1	NCBInr	350	14996	Single-stranded DNA-binding protein
9	1	NCBInr	212	77720	Elongation factor G

In order to achieve certain results, the data were scanned with *E. coli* taxonomy in both databases. The score values demonstrating the relations between protein sequence and matched mass spectrums were obtained through scanning on databases. Adequate score values revealed that data was reliable for analysis. Furthermore it's known that the NCBInr database is containing broader data then Swiss-Prot database. The similarities between results from both databases exhibit the consistency and purity of the protein samples. In both databases, two distinct proteins which are acetaldehyde alcohol dehydrogenase (adh) and RNA polymerase beta subunit were identified with great score values. In light of these findings, the possibility of the isolated protein being one those identified proteins is considered as quite enough. In particular, the 150 kDa molecular weight determined in SDS-Page results and the molecular weights of identified proteins which are found within 100-150 kDa range supports that possibility. For instance, the molecular weight of acetaldehyde alcohol dehydrogenase was specified as 100 kDa, and this value is rather close to the molecular weight of isolated protein. Besides, as a second alternative, RNA polymerase beta subunit is quite promising since that G-quadruplex structures could be found on promoter regions of many genes. In light of these information, characterization of isolated protein was completed by MALDI-TOF-TOF technique and scanning with both Swiss-Prot and

NCBInr databases, and two distinct proteins were identified with high possibility for being G-quadruplex binding protein.

3.2. Knocking-Out of G-quadruplexes by CRISPR/Cas9 Technology

3.2.1. pREDCas9, pCRISPR, and pCas9 Isolation

pREDCas9, pCRISPR, and pCas9 plasmids were isolated from their host strains in order to be used in CRISPR/Cas9 technique. While pREDCas9 is resistant to spectinomycin, pCRISPR is resistant to kanamycin, and pCas9 is resistant to chloramphenicol. As such, bacteria were grown in 50 µg/ml spectinomycin, 50 µg/ml kanamycin, and 25 µg/ml chloramphenicol concentrations by following recommendation of addgene. Plasmids were isolated from overnight growth cultures and they were observed with approximately 15,000 bp, 9,000 bp, and 2700 bp for pREDCas9, pCRISPR, and pCas9 respectively. While pREDCas9 has low copy number, remaining two plasmids have high copy numbers. Isolated plasmids were stored at -20 ^{0}C, and used on following applications.

3.2.2. sgRNA and Template Oligonucleotide Design

Genome modification with CRISPR/Cas9 technique requires sgRNA designing and repairing DSB with HDR mechanism requires template oligonucleotides carrying homology arms. In this scope, sgRNAs for G-quadruplex forming sequences on the upstreams and downstreams of hrpB, yhhP, yihS, TauD, adK, alx, cstA, and pstA genes were designed targeting at least two regions. Furthermore, designing of sgRNAs were completed as they contain BsaI restriction sites so that they can be inserted into pCRISPR plasmid. Additionally, homology arm template oligonucleotides were designed for each genomic region. While there are many online tools that could be used for sgRNA and template DNA design, in this study Benchling online tool were used and the oligonucleotides were selected considering their on-target and off-target scores provided by the algorithm of the tool. Although more then one sgRNAs had been designed, one of them were selected and used in the initial trials of CRISPR/Cas9 technique. Those sgRNAs and homology arm template oligonucleotides' sequences are listed on Table 3.

Table 3. The forward and reverse sequences of sgRNA forming oligonucleotides as well as the sequence of homology template oligonucleotides used in the study.

Gene Name	Forward Oligonucleotide	Reverse Oligonucleotide	Homology Template Oligonucleotide
adk	AAACGCTGATAAGTTTGCTTGTGCG	AAAACGCACAAGCAAACTTATCAGC	AAGTTCGCGCTGATCTGGAAAAAATACTCGGCTAATTCGAAAGCGCGCACAACAGGTCCGCACAAGCAAACTTATCAGCAATCTCAGGCCGGATATTCAT
alx	AAACGACAACATCAAAAAAGGACTG	AAAACAGTCCTTTTTTGATGTTGTC	GGATAATTTTTAATCTGCCTAAGCCGTGTACCCTGTCATTAACATGAGCAGAGTCCTTTTTTGATGTTGTCATCAGTCTGGAAGCCGCACGTTGGCTTTA
pstA	AAACCGCCTTTTTGACTCTGTACAG	AAAACTGTACAGAGTCAAAAAGGCG	GCGTTGTTTTTGCGAAGAATAAACACGGTTGATATTGCTGACACGGTTTTGAGCGCCTTTTTGACTCTGTACACAATTAACACTTTGCCGGATGCGGCGT
cstA	AAACTCAGTCCCCTCGCCCCACCGG	AAAACCGGTGGGGCGAGGGGACTGA	ACACTAAAGTCAGAGTGAGGGGGGGATGTTGGCGAATGTTGGCTTAGTGCGAGCGCGTTGATAGCATTTGTAGGCCGGATAAGGCGTTCACGCCGCATCC
tauD	AAACGCGTTCACCGTACTTTCAACG	AAAACGTTGAAAGTACGGTGAACGC	CGATCCTTGGGGATAAACCGTTTTATCGGGCGGGGTAATACGAGAGTGGATACTTTCAACAGGTTAACTCCCCCTTTCTGAGAGGAAACAAAATTAACGC
hrpB	AAACCGAAAAAGTATTCGTAAGGTG	AAAACACCTTACGAATACTTTTTCG	ACGACCCGGCAAATACTGCACCGACGCGACGGACGAAAAAGTATTCGTAATCAAGGGAGAGGGGACCGATCGAGCACAAATTTTGAGAGATATCTTCTTC
yihS	AAACGACCGTCCGGTGCAGTTTTTG	AAAACAAAAACTGCACCGGACGGTC	TGCTGGATATTAATGCGAAATAACTGATCTGAAAATAGGGTGCGAGTTTTGGTGCAGTTTTTGGGTTTGACATCAGCCTCACGGGAGGCTGTATGTCGTT
yhhP	AAACATCGCGCTCAATGTTGCGATG	AAAACATCGCAACATTGAGCGCGAT	GATTCGTAAAGGCGGTTGATAGGGGCTGATTGGCTTCGATGCCGCCTTTTCGATCGCGCTCAATGTTGCGATCAATTTGCCTTATCTCCTGCGCAACAAT

3.2.3. pREDCas9 transformation

Spectinomycin resistant, tracrRNA and ƛ-red genes expressing, 15,461 bp containing pREDCas9 plasmid had been transferred to Mg1655 and HME63 strains of *E. coli* via heat shock transformation. Transformed bacterial cultures were spreaded to 50 µg/ml spectinomycin containing LB-Agar plates and incubated at 37 ^{0}C overnight. Following day single colonies indicating pREDCas9 carrying bacteria were observed (Figure 11). One of those observed colonies were selected inoculated into 50 µg/ml spectinomycin containing liquid LB media. Overnight cultures were inoculated to fresh LB media, and the cultures reached exponential phase were stored in 25% glycerol for following applications at -80 ^{0}C.

Figure 11. Colonies of Mg1655 and HME63 strains of *E. coli* on spectinomycin containing LB-Agar plates after pREDCas9 transformation.

3.2.4. pCRISPR Digestion and Gel Purification

The strategy developed by Jiang's, et al. requires delivery of pCas9 plasmid and sgRNA inserted pCRISPR plasmid to bacterial cells. In order to insert sgRNA pCRISPR were designed with directed repeat regions which could be digested with the

BsaI restriction enzyme. Through the digestion, this region is removed from pCRISPR plasmid and designed sgRNAs could be ligated by the BsaI restriction sites. As such, by following the recommendation of the supplier company of BsaI enzyme (ThermoFisher Scientific), pCRISPR plasmid were digested and BsaI was inactivated for ceasing it's cleavage activity. Digested pCRISPR samples were loaded to 1% agarose gel and results were visualized with UV-visualization techniques (Figure 12). Furthermore it's expected that the removed fragment found in test tubes could reanneal to digested pCRISPR plasmid through the extant BsaI restriction sites. In order to get rid of these fragments from the samples, digested pCRISPR plasmids were isolated from agarose gel by utilizing MACHEREY-NAGEL NucleoSpin Gel and PCR Clean-up kit. Considering the arms of digested pCRISPR plasmids could not anneal to one another since the sites could not match, the gel isolated samples were stored at -20 ^{0}C for following applications.

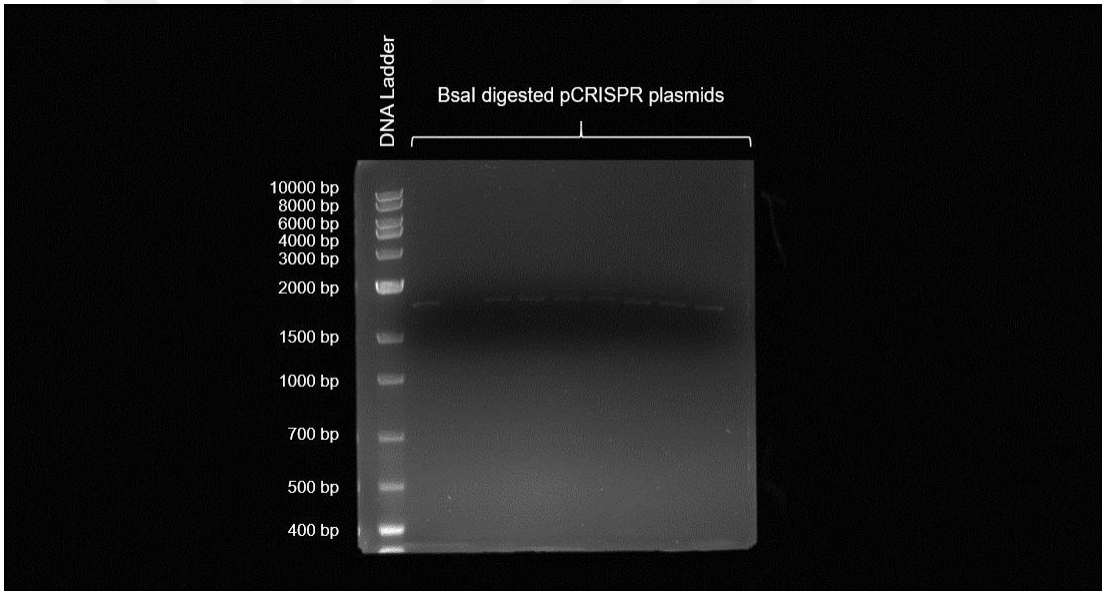

Figure 12. 1% agarose gel image containing pCRISPR plasmids digested with BsaI restriction enzyme for 16 hours.

3.2.5. PCR Applications for Control of sgRNA Insertion

sgRNA forming oligonucleotides were designed as forward and reverse sequences, and the ligation process requires phosphorylation and annealing of those oligonucleotides in order to insert into BsaI digested pCRISPR plasmid. As such, sgRNA forming forward and reverse primers were phosphorylated in a single test tube, annealed by incubation at 95 ^{0}C with 0.1 M NaCl, and diluted 10 times for enhancing

ligation efficiency. Ligation technique were conducted by following Jiang and his colleagues' protocol. In this scope, reaction mixes containing digested pCRISPR, annealed and diluted sgRNAs, T4 DNA ligase, and reaction buffer were incubated at room temperature. sgRNA insertion into digested pCRISPR might be controlled by PCR application since that forward oligonucleotide of sgRNAs could be used as forward primer with reverse primer recognizing pCRISPR plasmid which had been designed previously. As such, PCR reaction mixes for each sgRNA inserted pCRISPR plasmids were prepared as containing sgRNA:pCRISPR, sgRNA forward oligonucleotide, reverse primer, PCR MasterMix. In addition, the same reaction mixes were prepared with undigested pCRISPR plasmid instead of sgRNA:pCRISPRs as a negative control group. Due to that optimum Tm values of forward and reverse primers are not known, Touch-down PCR approach was carried out. While PCR protocols were separated into two distinct stages, annealing temperature was adjusted as starting from 65 ^{0}C and completing at 55 ^{0}C by 0.5 ^{0}C reducing in each cycle of totally 20 cycles. Besides, annealing temperature was kept at constant 55 ^{0}C in the second stage of PCR application. Once that Touch-down PCR had been completed, samples were loaded into 1% agarose gel and results were analyzed by UV-visualization. The bands obtained from Touch-down PCR are demonstrated on Figure 13.

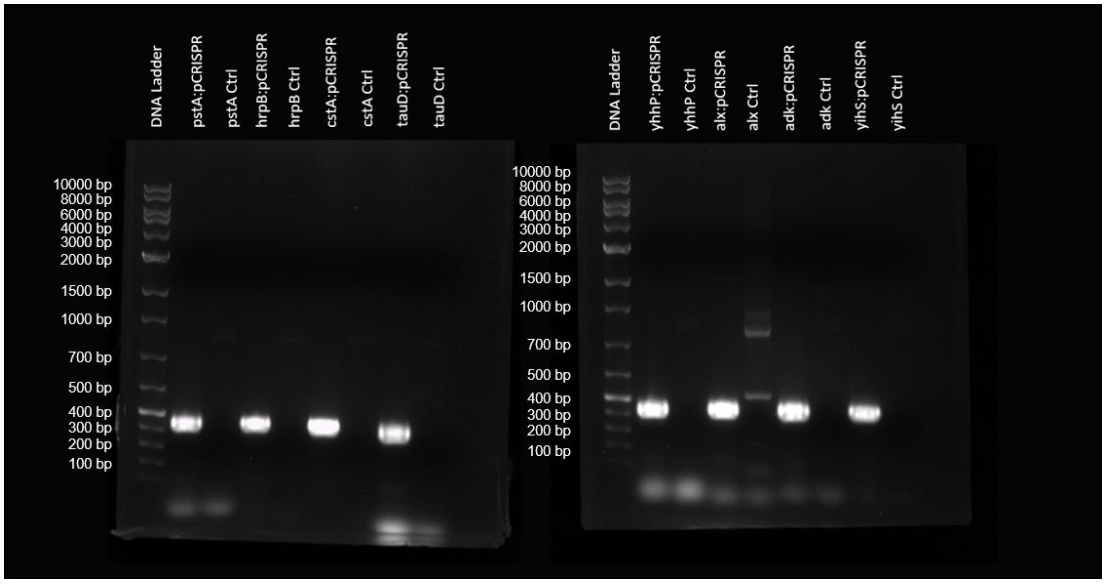

Figure 13. Control of sgRNA forming oligonucleotides into digested pCRISPR plasmids after Touch-down PCR application.

While positive bands indicating sgRNA insertion into digested pCRISPR plasmids were observed on sgRNA:pCRISPR samples, any bands were not encountered from PCR samples of undigested pCRISPR plasmids. Since that undigested pCRISPR plasmids do not contain sgRNA forming forward and reverse oligonucleotides, forward oligonucleotide of sgRNAs could not recognize any region of whole plasmid sequences. As such, it's concluded that insertion of sgRNA forming oligonucleotides into BsaI digested pCRISPR plasmids were accomplished with sufficient efficiency. On following steps, sgRNA:pCRISPR plasmids were delivered to competent DH5α in order to increase their amount. Afterwards all sgRNA:pCRISPR plasmids were isolated from bacterial colonies picked on 50 µg/ml kanamycin containing LB-Agar plates.

3.2.6. Transfer of sgRNA:pCRISPR and Template DNAs to pREDCas9 Containing Bacteria

Once that sgRNA insertion into BsaI digested pCRISPR had been confirmed, pREDCas9 containing Mg1655 and HME63 strains of *E. coli* were transformed with lacZ sgRNA:pCRISPR plasmids and G-quadruplex targeting sgRNA:pCRISPR plasmids with electroporation, respectively. As such, electroporation was conducted by following the recommended pre-set protocol of the Gene Pulser Xcell Electroporation System. In addition, sgRNA:pCRISPR plasmids were co-delivered with related homology template oligonucleotides which had been designed previously. In this wise, that sgRNA expressing from pCRISPR, Cas9 and ƛ-red genes expressing from pREDCas9, creation double strand break, and repairing DSB with homology template oligonucleotides were intended with this approach.

Mg1655 bacterial cells transformed with lacZ sgRNA:pCRISPR plasmid and template oligonucleotide were incubated at 37 ^{0}C, and spreaded over LB-Agar plates which contain 50 µg/ml spectinomycin, 50 µg/ml kanamycin, 0.1 mM IPTG, and 40 µg/ml X-gal compounds. Following overnight incubation of plates at 37 ^{0}C, colonies were observed. The colonies of E. coli Mg1655 culture transformed with lacZ sgRNA:pCRISPR and template DNA are shown on Figure 14-A. While lacZ expressing bacterial colonies are observed as blues, the bacterial colonies that do not express lacZ might be observed as white on the existence of X-gal and IPTG compounds. Due to the fact that, it's obviously concluded that lacZ gene was knocked-

out with 60-65% efficiency in our results. These results indicate that CRISPR/Cas9 technique, which intends to edit the genome by creating DSB and repairing with designed template oligonucleotides, was accomplished with enough efficiency in our approaches.

Once to observe lacZ knock-outs on Mg1655 strain, the procedure was applied on HME63 strain of *E. coli* for editing G-quadruplex forming sequences of its genome. It's highlighted that G-quadruplex forming sequences had been revealed on either upstreams or downstreams of various genes such as HrpB, TauD, adK, alx, cstA, pstA, yihS, yhhP. In this scope sgRNAs targeting the G-quadruplex structures related with those genes were designed and inserted into pCRISPR plasmid. These sgRNA:pCRISPR plasmids were co-delivered with relevant template DNAs to HME63 cultures. The cells were incubated at 37 ^{0}C, and spreaded over LB-Agar plates containing 50 µg/ml spectinomycin, and 50 µg/ml kanamycin. Following overnight incubation of plates at 37 ^{0}C, the colonies were encountered. The colonies indicating HME63 cells containing both sgRNA:pCRISPR and pREDCas9 plasmids are demonstrated on Figure 14-B. These results indicate that pREDCas9 containing bacterial cells were transformed with pCRISPR since that they might survive on the existence of both spectinomycin and kanamycin. However, due to the fact that editing G-quadruplex forming sequences could not be analyzed with naked eyes contrary to lacZ knock-outs, the editing efficiencies of those trials should have been analyzed with other approaches.

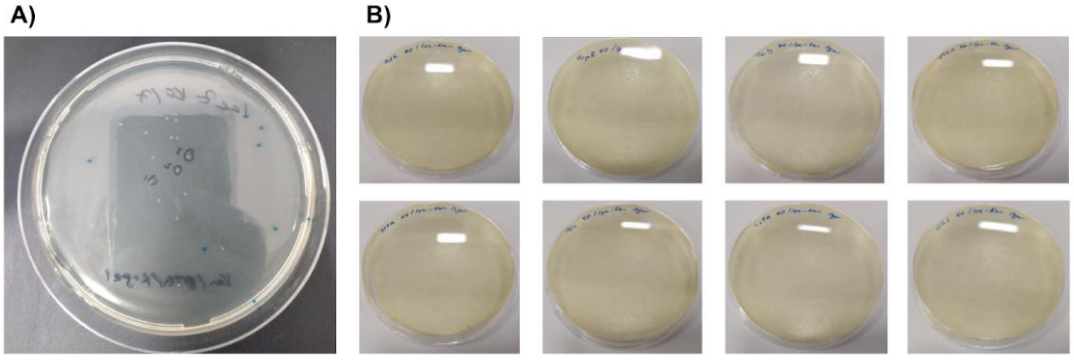

Figure 14. The colonies after CRISPR/Cas9 application A) white colonies indicating lacZ knock-outs and blue colonies indicating lacZ expressing bacteria B) colonies of G-quadruplex forming sequences adjacent to adk, hrpB, tauD, yhhP, pstA, alx, cstA, yihS genes knock-out trials.

3.2.7. Colony PCR Applications for Control of Genome Editing

In order to analyze editing of G-quadruplex forming sequences related to adjacent genes, colony PCR application were conducted. In the CRISPR/Cas9 approach of G-quadruplex editing, it's intended that a genomic region with the length between 50 bp - 80 bp is removed for each knock-out. The colonies of HME63 cultures containing sgRNA:pCRISPR, pREDCas9 and template DNA were selected and utilized in colony PCR application performed with forward and reverse sequencing primers. Furthermore, other reaction mixes were prepared with the genome of wild type HME63 bacteria with the same primers. Furthermore, since that optimum Tm values of primers had not been identified, Touch-down PCR technique were conducted as in sgRNA insertion into digested pCRISPR plasmids. Once the Touch-down PCR had been completed, samples were loaded into 2.5% agarose gel. Due to that differences between control groups and interested samples were expected between 50-80 bp, high concentrated agarose gel should have been used such as 2.5%. The results of PCR reactions including the samples and their controls are demonstrated on Figure 15. Any differences indicating knocking-out of G-quadruplexes on the genomic regions of *E. coli* were not observed between samples and control groups. Thus it's deduced that sgRNA:pCRISPR plasmids have been delivered, yet CRISPR/Cas9 technique has not worked on G-quadruplex forming sequence knock-outs. Besides, only picked colonies were analyzed in these experiments, considering the efficiency of CRISPR/Cas9 was 60-65% on lacZ knock-outs, many other colonies of HME63 bacteria should be investigated.

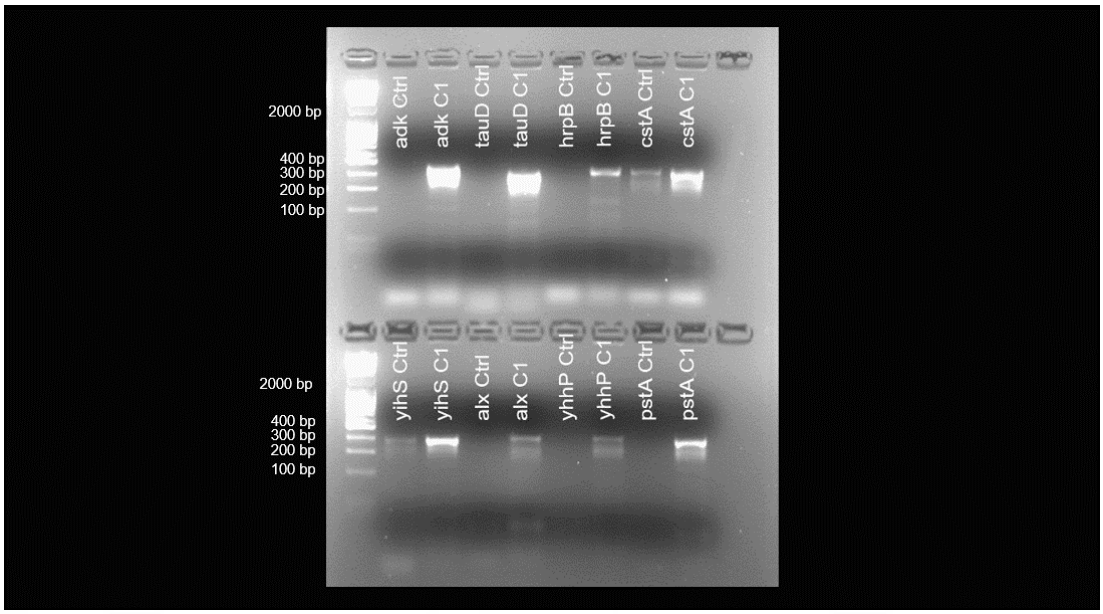

Figure 15. Colony PCR results of G-quadruplex forming sequences adjacent to adk, tauD, hrpB, cstA, yihS, alx, yhhP, pstA genes knock-out trials.

CHAPTER 4: DISCUSSION

In this study which is intended to investigate biological roles of genomic G-quadruplex structures of *E. coli* microorganism, various molecular biology techniques based on both traditional and novel approaches were carried out, and many results were obtained on the final stages of experiments. For instance, pull-down assay based on revealing of G-quadruplex binding cytoplasmic proteins of *E. coli* were conducted by following Jutras and his colleagues' protocol (Jutras, Verma, and Stevenson, 2012). In this scope, Mg1655 strain of E. coli bacteria were grown with great volume by incubating at 37 ^{0}C as optimum growth conditions highlighted in literature (Guyot et al., 2014). The turbidity indicating bacterial growth were observed, and the high number of growth bacterial cells were proven by observing bacterial pellets after centrifugation (Abu Bakar et al., 2015). Bacterial pellets were stored at -80 ^{0}C as recommended from many studies (Galluccio et al., 2020). Stored bacterial pellets were lysed with commercial B-Per II reagent whose lysis efficiency had been demonstrated with various applications (Islam, Aryasomayajula, and Selvaganapathy, 2017). Extracted cytoplasmic proteins via lysis were used in pull-down assays. As such, G-quadruplex forming biotin functionalized oligonucleotides were designed and incubated with Dynabead M-280 Streptavidin beads in order to create DNA-Bead complexes formed through interactions between streptavidin and biotin molecules. Bead-DNA complexes were used as bait for G-quadruplex binding proteins of *E. coli* lysates. Protein bound DNA-Bead complexes were collected with the help of strong magnets, and the incubations were repeated until saturation of DNA-Bead complexes with G-quadruplex binding proteins. Since that optimum salt concentrations for elution of G-quadruplex binding proteins, Protein-DNA-Bead complexes were incubated with distinct elution buffers containing different NaCl concentrations from 100 mM to 1 M. In order to analyze molecular existence and molecular weight of G-quadruplex binding proteins, samples were runned within polyacrylamide gel by performing SDS-Page electrophoresis technique. Furthermore, a bacterial lysate sample containing total cytoplasmic proteins was loaded into the last well of polyacrylamide gel. Results demonstrated that bacteria were lysed with quite efficiency since the many protein bands had been observed, and G-quadruplex binding proteins were encountered on E3 samples which are eluted in 25 mM Tris HCl, 300 mM NaCl buffer. The purity of protein bands were analyzed with ImageJ software. In addition, comparing the

commercial protein ladder loaded into the first well, it's concluded that the discovered G-quadruplex binding protein has approximately 150 kDa molecular weight.

Once that G-quadruplex binding protein had been observed with SDS-Page electrophoresis technique, it was characterized with the MALDI-TOF-TOF method. As such the low concentration of discovered protein was increased through repetition of Pull-down assays and lyophilization. Afterwards, protein was digested with trypsin enzyme and loaded to matrix for MALDI-TOF-TOF analysis in the Mass Spectrometry Analysis Laboratory in Hacettepe University. MALDI-TOF-TOF technique is such a mass spectrometry technique which is conducted for characterization of many molecules. During the process, protein samples are loaded into a special matrix, their ions are flown by application of ultraviolet laser, the fallings of ions are computed as mass/charge (m/z) ratio, and the samples are characterized according to their ions' time of flight. Therefore characteristic peaks of each protein might be identified, insomuch that various databases including the peaks of many samples were created and used for determination of proteins (Webster, and Oxley, 2012). In addition, this technique providing reliable results is also used for microorganism detection in several areas (Dingle, and Butler-Wu, 2013). In this scope, the discovered G-quadruplex binding protein was analyzed with scanning of NCBInr and Swiss-Prot databases. Consequently the acetaldehyde alcohol dehydrogenase (adh) and RNA polymerase beta subunit were identified by both NCBInr and Swiss-Prot databases. Furthermore, while the scores indicating possibilities of acetaldehyde alcohol dehydrogenase were determined as 521 and 376, RNA polymerase beta subunit's scores were determined as 379 and 333 with both Swiss-Prot and NCBInr databases respectively. Additionally, the molecular weights of acetaldehyde alcohol dehydrogenase and RNA polymerase beta subunit were identified as over 96 kDA and 150 kDA respectively in both databases (Lane, and Darst, 2010; Nair, Bennett, and Papoutsakis, 1994). Considering the scores, and the molecular weights, it's evaluated that the G-quadruplex binding protein discovered with molecular weight between 100-150 kDA might be one of those proteins. In the molecular mechanism of acetaldehyde alcohol dehydrogenase, it's revealed that adh acts the primary role on ethanol metabolizing to acetaldehyde (Cederbaum, 2012). Furthermore, it's also revealed that adh is overexpressed by *E. coli* in the low O2 supply. In this scope the great importance of adh in alcoholic fermentation of *E. coli* were highlighted with studies published in

literature (Holland-Staley et al., 2000). Attractively, the mRNA secondary structures' unwinding activity through interacting with 70S subunit of ribosomes which is responsible for mRNA entrance of adhE protein were demonstrated with the study of Shasmal and his colleagues at 2016 (Shasmal et al., 2016). Scientists have proven that this protein acts like mRNA helicase with their study published in the journal of Nature. Accordingly, it's demonstrated that the related protein could interact with nucleic acid chemistry, insomuch that it might unwind their secondary structures. In light of these information and considering the parallel molecular weights, it's concluded that the discovered G-quadruplex binding protein might be adh and the interested G-quadruplex structures might be one of subunits on alcohol fermentation process. Besides, a second possible alternative for being G-quadruplex binding protein of *E. coli* DNA directed RNA polymerase beta subunit was identified with 150 kDA molecular weight on both databases. It's demonstrated in the study of Kaplan, et al. G-quadruplex structures are able to form on either upstreams or downstreams of various genes of *E. coli* genome (Kaplan et al., 2016). This phenomena points that these non-canonical structures might be observed on functional regions such as promoters of those genes. In addition, it's previously emphasized in literature that G-quadruplexes might be found on many genes' promoters of the genome of many organisms including humans (Rigo, Palumbo, and Sissi, 2017). Consequently, RNA polymerase could interact with these structures. As such, that the possibility of discovered G-quadruplex binding protein with Pull-down assays could be RNA polymerase beta subunit is quite high. The case of discovered protein is determined as the RNA polymerase subunit will demonstrate that G-quadruplex structures might act a role on gene expression of the adjacent genes.

CRISR/Cas9 is one of the novel promising genome modification tools. Fundamentally, double strand breaks on the host genome are created via cleavage activity of Cas9 enzyme targeted with sgRNA and repaired with either NHEJ or HDR repair mechanisms. In this study CRISPR/Cas9 genome modification approach had been used in order to edit lacZ gene and G-quadruplex structures found on either upstreams or downstreams of tauD, hrpB, alx, adk, pstA, cstA, yihS, and yhhP genes of *E. coli*. As such, the pCRISPR and pREDCas9 plasmids whose activities had been demonstrated with the Jiang, et al. and Li, et al.'s studies respectively were used in our approach (Jiang et al., 2013; Li et al., 2015). While pCRISPR plasmid providing

directed repeats -insertion sites for sgRNA forming oligonucleotides-, crRNA, and kanamycin resistance, it's highlighted on the Jiang and his colleagues' study 65% efficiency had been obtained with co-usage of pCRISPR plasmids with pCas9 plasmid in order to insert desired mutations into bacterial genome. Furthermore, scientists emphasized that the sufficient efficiency was obtained with HME63 strain of *E. coli* since it's recombineering strain (Jiang et al., 2013). Furthermore, that 100% efficiency might be achieved by usage of pREDCas9 which is providing tracrRNA, ƛ-red genes, Cas9 enzyme, and spectinomycin resistance were shown with Li and his colleagues' research (Li et al., 2015). Due to ƛ-red genes of pREDCas9 plasmid, great efficiency might be observed on inserting desired mutations by manipulating HDR mechanism inside bacterial organisms. In light of these information, combinations of pCRISPR and pREDCas9 plasmids were utilized as a novel approach for genome modification in both Mg1655 and HME63 strains of *E. coli*. As such both strains were transformed with pREDCas9 plasmid with heat shock transformation and grown on spectinomycin containing LB-Agar plates. Selected single colonies containing pREDCas9 plasmid were used in the following applications. In the meantime, BsaI restriction sites carrying sgRNA forming oligonucleotides and the homology-armed template oligonucleotides were designed with Benchling online software. In addition, pCRISPR plasmids were digested with BsaI enzyme and isolated from 1% agarose gel by following protocol of Jiang, et al. While phosphorylated sgRNAs were inserted into BsaI digested pCRISPR plasmids, sgRNA insertions were proven with Touch-down PCR with proper primers. In addition, electrocompetent pREDCas9 containing bacteria cells were prepared by washing with distilled water according to Li, et al.'s protocol (Li et al., 2015). Bacterial cells were co-transformed with sgRNA inserted pCRISPR plasmids and related homology-armed template oligonucleotide and grown for spreading over appropriate LB-Agar plates.

In order to analyze genome modification efficiency of the developed approach consisting with co-usage of pCRISPR and pREDCas9 plasmids, lacZ gene were targeted to be knocked-out. lacZ gene is one of the components of lac operon of *E. coli* genome and which expresses β-galactosidase that is hydrolyzing beta-D-galactose's terminal residues to beta-D-galactosides (Juers, Matthews, and Huber, 2012). While HME63 does not express lacZ, Mg1655 strains were used for lacZ knock-out trials now that it expresses lacZ gene. Furthermore, while lacZ expressing bacterial colonies

are observed in blue colour in the existence of X-gal and IPTG compounds, lacZ knock-out bacterial cells are observed as white in the same conditions. While this phenomena is dependent on activity of β-galactosidase and metabolism of X-gal as substrate, a molecular biology technique with the name of blue/white screening was developed manipulating this process in order to analyze recombination in *E. coli* genome during many years (Padmanabhan, Banerjee, and Mandi, 2011). Once that pREDCas9 containing Mg1655 cells had been co-transformed with lacZ targeting pCRISPR and homology-armed oligonucleotide, they were spreaded over spectinomycin, kanamycin, IPTG, and X-gal containing LB-Agar plates. Blue colonies and white colonies were observed on overnight grown culture. While blue colonies which are negative results might be sourced from the bacterial cells transformed with only pCRISPR or bacterial cells that didn't allow the action of CRISPR/Cas9, white colonies indicate the lacZ edited cells. Comparing the number of white colonies and blue colonies, it's deduced that 65% efficiency was obtained with the developed CRISPR/Cas9 approach in genome editing on *E. coli* organisms.

Considering the study of Kaplan and his colleagues (Kaplan et al., 2016), the G-quadruplex structures found on upstreams or downstreams of various genes of *E. coli* were targeted with the CRISPR/Cas9 technique whose activity had been proven as 65%. Briefly, the targeted G-quadruplexes are found on the upstreams of hrpB, tauD, adk, alx, cstA, yhhP, and downstreams of pstA and yihS genes. The functions of these genes might be summarized as; *i)* hrpB expresses ATP dependent helicase enzyme (Granato et al., 2016), *ii)* tauD expresses alpha-ketoglutarate dependent taurine dioxygenase providing taurine metabolism as an alternative source in the absence of sulfate (Bollinger et al., 2005), *iii)* adk expresses adenylate kinase responsible for transfer of last sulfate group between ATP and AMP (Gutierrez, and Csonka, 1995), *iv)* alx expresses membrane dependent redox modulator (Zeinert et al., 2018), *v)* cstA expresses peptide transporter transfering peptides in the absence of carbon (Gasperotti et al., 2020), *vi)* pstA acts a role in phosphate specific transport system (Webb, Rosenberg, and Cox, 1992), *vii)* yihS expresses sulfoquinovose isomerase providing sulfoquinovose isomerization (Kaznadzey et al., 2018), and *viii)* yhhP act a significant role in *E. coli* physiognomy and it's demonstrated that without yhhP expression *E. coli* cells do not survive on media (Yamashino et al., 1998). In light of these information, it's expected that G-quadruplexes might have regulatory

roles on those genes possessing significant metabolic activities. Considering these, knocking-out those G-quadruplexes might cause alteration on expression levels of related genes. During G-quadruplex editing HME63 strains were used in order to increase efficiency compared to lacZ knock-outs. As such pREDCas9 containing HME63 cells were co-transformed with G-quadruplex targeting pCRISPR plasmid and homology-armed template oligonucleotide. The bacterial cultures were spreaded over spectinomycin and kanamycin containing LB-Agar plates. Theoretically between 50-80 bp is able to be removed from the upstreams or downstreams of related genes caused by design of homology-armed oligonucleotides in each editing. Following they single colonies indicating the bacterial cells containing both pREDCas9 and pCRISPR simultaneously were observed over LB-Agar plates and few of them were selected so that to be used in Colony-PCR. Genomic DNA of colonies were targeted with proper primers, and wild type HME63 genome were used as control. Performing Touch-down PCR the DNA samples were loaded into 2.5% agarose gel and results were displayed with UV-Visualization. However any difference between genomes of wild type HME63 and colonies were not observed. All these results have demonstrated that G-quadruplex structures might not be edited with CRISPR/Cas9 application. It's previously revealed that our developed method had exhibited 65% efficiency in genome editing. Taking into account the low efficiency, it's estimated that colonies which didn't enable genome editing might be selected in colony PCR application. On the contrary, knocked-out regions could be repaired with NHEJ repair mechanism or other strategy of HME63 cells. As such, other colonies should be investigated and those genomes are required to be sequenced in order to analyze possible mutations caused by different sources. However most prominently, it's known that G-quadruplexes have higher stability compared to canonical double stranded structures of DNA molecules (Lane et al., 2008). In addition, those G-quadruplexes must be unwinded by sgRNA loaded into Cas9 enzyme in genome editing since it could cleave dsDNA (Wu, Kriz, and Sharp, 2014). The main problem which is unwinding G-quadruplexes might affect the efficiency of the developed CRISPR/Cas9 approach in this study. As such, novel strategies should be developed and their efficiencies should be revealed in order to modify G-quadruplex forming structures.

CHAPTER 5: CONCLUSION

G-quadruplexes are one of the significant secondary structures of nucleic acids formed with stacking of G-tetrads consisting of adjacent guanine residues. While these specific structures could be formed as intramolecular and intermolecular, it's also uncovered that due to their four stranded structures they have higher stability compared to canonical dsDNA molecules. During many years various functions of G-quadruplexes had been discovered among various organisms. In particular the locations and the interactions between their own ligands identify biological roles of those non-canonical structures. Furthermore, the relations between G-quadruplexes with many important diseases such as cancer were revealed as well as other regulatory functions of them on metabolic reactions. Besides the significant roles of G-quadruplexes on mammalian organisms, their several functions on microorganism metabolisms were also discovered with few studies. For instance, the functions of G-quadruplexes on virulence strategies of various viruses indicates their significance on microorganisms. *E. coli* is one of the most prominent bacteria living in gastrointestinal tracts of mammals. While there are several pathogenic strains of *E. coli* causing mild diseases, it became a model organism in laboratory usage now that its genomic and proteomic structures were discovered for many years. However there are not many studies investigating G-quadruplex structures of the *E. coli* genome except the study of Kaplan and his colleagues. In that study 52 G-quadruplex forming structures on either upstreams or downstreams of various genes were discovered with a computational tool. Scientists have demonstrated the locations of those G-quadruplexes on *E. coli* genome map and emphasized that their findings indicate some regulatory roles of related G-quadruplex structure. In this scope biological roles of these novel G-quadruplexes discovered in the *E. coli* genome are investigated with this research. As such the interactions between this G-quadruplex forming sequence with the ligands belonging to *E. coli* organisms and the effect of absence of these structures on expression levels of adjacent genes are tried to be analyzed.

So far it's revealed that biological functions of many G-quadruplexes are depending on interactions with their own ligands such as various enzymes or protein complexes. G-quadruplex binding proteins of *E. coli* have not been demonstrated and characterized yet. While many molecular biology techniques were developed in order to investigate nucleic acid binding proteins, Pull-down assay which is based on usage

of nucleic acids as bait for proteins is one the most prominent of those techniques. Following this promising technique G-quadruplex binding protein of *E. coli* was collected from bacterial lysates and demonstrated with 100-150 kDa molecular weight by performing SDS-Page electrophoresis. Furthermore, isolated proteins were characterized with MALDI-TOF-TOF technique. While the fingerprint peaks of investigated proteins had been scanned in Swiss-Prot and NCBInr databases, two common proteins which are acetaldehyde alcohol dehydrogenase and DNA dependent RNA polymerase beta subunit were determined with reliable scores. Molecular weights of both proteins exhibit similarity with the molecular weight of the discovered G-quadruplex binding protein. It's known that G-quadruplexes could be found on promoter regions of many genes of various organisms. Accordingly the interaction between RNA polymerase with those G-quadruplexes might be predicted. However, revealing the interaction between beta subunit of RNA polymerase and G-quadruplexes of *E. coli* genome keeps light to regulatory roles of G-quadruplexes on expression levels of related genes. Besides, acetaldehyde alcohol dehydrogenase (adh) is a special enzyme that acts a role in alcohol fermentation in low O2. Interestingly secondary structures of several mRNAs might be unwinded through special interactions between ribosomes and adhE had been demonstrated in literature. While this phenomena reveal that adhE could work such an mRNA helicase, the revealing relations between adh and genomic G-quadruplexes of E. coli is going to be a significant discovery about the biological role of G-quadruplexes on alcohol fermentation process.

Additionally, investigating the effect of knocking-out of G-quadruplexes on expression level of their interrelated genes gives direction concerning biological roles of those non-canonical structures. As such, a novel CRISPR/Cas9 approach consists of a combination of two different methods whose efficiencies had been proven previously were developed. Considering the study of Kaplan, et al. sgRNA forming oligonucleotides and homology-armed template oligonucleotides were designed as targeting G-quadruplexes closer to tauD, hrpB, adk, alx, pstA, cstA, yihS, and yhhP genes as well as lacZ gene in the Benchling online tool. Mg1655 and HME63 strains were transformed with pREDCas9 plasmid by performing heat shock transformation. Once that pCRISPR plasmid had been digested with BsaI restriction enzyme, designed sgRNA forming oligonucleotides were inserted through the ligation process.

Subsequently, pREDCas9 containing bacterial cells were co-transformed with sgRNA inserted pCRISPR plasmid and homology-armed template oligonucleotide by electroporation. Blue/white screening technique had been carried out by spreading lacZ edited Mg1655 culture over IPTG, X-gal, spectinomycin, and kanamycin containing LB-Agar plates. Consequently the efficiency of the novel CRISPR/Cas9 approach was obtained as 65% with the results from lacZ editing on Mg1655 strain of *E. coli*. Even though HME63 bacteria colonies had been observed by spreading bacterial cultures over spectinomycin and kanamycin containing LB-Agar plates after CRISPR/Cas9 application targeting G-quadruplexes, any positive results were not obtained indicating G-quadruplex knock-outs with Colony PCR technique. The colonies show the adequate efficiency of electroporation in pCRISPR transformation since that they might survive on both spectinomycin and kanamycin containing LB-Agar plates. Furthermore it's predicted that double strand breaks could be repaired with other strategies of HME63 cells. In order to analyze possible mutations genomic DNAs of organisms should be sequenced with sequencing techniques. However it's established that the main reason for negative results is higher stabilities of G-quadruplexes due to they should be unwinded with sgRNAs found within Cas9 enzyme. In this case, other innovative CRISPR/Cas9 approaches should be developed.

Briefly, the functions of G-quadruplexes discovered in the *E. coli* genome were investigated in this study. The findings kept light on many metabolic activities of investigated G-quadruplexes. For instance, a G-quadruplex binding protein was discovered and characterization experiments demonstrated that it might be acetaldehyde alcohol dehydrogenase (adh) or DNA dependent RNA polymerase beta subunit. While in the case of adh it will enlighten the role of G-quadruplexes on alcohol fermentation, if protein is identified as RNA polymerase beta subunit it will indicate the regulatory function of G-quadruplexes on transcription of interrelated genes. Besides, Western Blotting technique will be conducted in order to verify the proteins that have been revealed with MALDI-TOF-TOF analysis. As such, the antibodies recognizing adhE and RNA polymerase Beta subunit will be acquired, SDS-Page technique will be repeated, and protein intensities will be analyzed by performing Western Blotting. Furthermore, due to the fact that G-quadruplexes possess higher stability than double helix structures of DNA molecules, they couldn't be knocked-out with the developed CRISPR/Cas9 approach. As such, novel other approaches will be

searched in order to edit G-quadruplexes on *E. coli* genome, and the effect of their knocking-out on expression levels of tauD, hrpB, cstA, pstA, adk, alx, yihS, and yhhP are going to be analyzed. Therefore, the biological role of G-quadruplexes on the metabolic reactions performed by respective genes will have been revealed. Terminally, G-quadruplex stabilizing molecules such as BRACO-19, Phen-DC 3, TMPyP4 will be delivered to bacterial cells, and the effect of G-quadruplex stabilizing on expression level of the same genes will be analyzed. While the findings will keep light on biological functions of genomic G-quadruplexes of *E. coli* organism, these discoveries will contribute significant impact to literature.

REFERENCES

Aasen, M. (1998) *Senatet blokkerte for bruk av aktiv dødshjelp i Australia*, Tidsskrift for den Norske Laegeforening, Vol. 118, pp. 440-445.

Abu Bakar, A., Mohd Rasol, R., Nordin, Y., Md Noor, N., and Bin Mohd Ali, M.K.F. (2015) *Turbidity method to measure the growth of anaerobic bacteria related to microbiologically influenced corrosion*, Solid State Phenomena, Vol. 227, pp. 298–301.

Adámik, M., Kejnovská, I., Bažantová, P., Petr, M., Renčiuk, D., Vorlíčková, M., and Brázdová, M. (2016) *p53 binds human telomeric G-quadruplex in vitro*, Biochimie, Vol. 128–129, pp. 83–91.

Adli, M. (2018) *The CRISPR tool kit for genome editing and beyond*, Nature Communications, Vol. 9, pp. 1-13.

Allocati, N., Masulli, M., Alexeyev, M.F., and Di Ilio, C. (2013) *Escherichia coli in Europe: An overview*, International Journal of Environmental Research and Public Health, Vol. 10, pp. 6235–6254.

Almeida, C.R. De, Dhir, S., Dhir, A., Sattentau, Q., Meinhart, A., Proudfoot, N.J., Almeida, C.R. De, Dhir, S., Dhir, A., Moghaddam, A.E., and Sattentau, Q. (2018) *RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination*, Molecular Cell, pp. 1–13.

Anzalone, A. V., Koblan, L.W., and Liu, D.R. (2020) *Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors*, Nature Biotechnology, Vol. 38, pp. 824–844.

Arroyo-Olarte, R.D., Bravo Rodríguez, R., and Morales-Ríos, E. (2021) *Genome editing in bacteria: Crispr-cas and beyond*, Microorganisms, Vol. 9, pp. 1-25.

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, pp. 261-275.

Bartas, M., Cutová, M., Brázda, V., Kaura, P., Šťastný, J., Kolomazník, J., Coufal, J., Goswami, P., Červeň, J., and Pečinka, P. (2019) *The presence and localization of*

G-Quadruplex forming sequences in the domain of bacteria, Molecules, Vol. 24, pp. 1-13.

Biswas, B., Kandpal, M., Jauhari, U.K., and Vivekanandan, P. (2016) *Genome-wide analysis of G-quadruplexes in herpesvirus genomes*, BMC Genomics, Vol. 17, pp. 1–16.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997) *The complete genome sequence of Escherichia coli K-12*, Science, Vol. 277, pp. 1453–1462.

Blount, Z.D. (2015) *The unexhausted potential of E. coli*, eLife, Vol. 4, pp. 1–12.

Bollinger, J.M., Price, J.C., Hoffart, L.M., Barr, E.W., and Krebs, C. (2005) *Mechanism of taurine: α-ketoglutarate dioxygenase (TauD) from Escherichia coli*, European Journal of Inorganic Chemistry, Vol. 2005, pp. 4245–4254.

Brázda, V., Cerveň, J., Bartas, M., Mikysková, N., Coufal, J., and Pečinka, P. (2018) *The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors*, Molecules, Vol. 23, pp. 1-16.

Brázda, V., Hároníková, L., Liao, J.C.C., and Fojta, M. (2014) *DNA and RNA quadruplex-binding proteins*, International Journal of Molecular Sciences, Vol. 15, pp. 17493–17517.

Cederbaum, A.I. (2012) *Alcohol Metabolism*, Clinics in Liver Disease, Vol. 16, pp. 667–685.

Chen, Z., and Xu, X.H. (2016) *Roles of nucleolin: Focus on cancer and anti-cancer therapy*, Saudi Medical Journal, Vol. 37, pp. 1312–1318.

Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K.K., and Xodo, L.E. (2010) *The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element*, Journal of Biological Chemistry, Vol. 285, pp. 22003–22016.

Cogoi, S., and Xodo, L.E. (2006) *G-quadruplex formation within the promoter of the*

KRAS proto-oncogene and its effect on transcription, Nucleic Acids Research, Vol. 34, pp. 2536–2549.

Croxatto, A., Prod'hom, G., and Greub, G. (2012) *Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology*, FEMS Microbiology Reviews, Vol. 36, pp. 380–407.

Cui, Y., Xu, J., Cheng, M., Liao, X., and Peng, S. (2018) *Review of CRISPR/Cas9 sgRNA Design Tools*, Interdisciplinary Sciences: Computational Life Sciences, Vol. 10, pp. 455–465.

Ding, Y., Fleming, A.M., and Burrows, C.J. (2018) *Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes*, Scientific Reports, Vol. 8, pp. 1–11.

Dingle, T.C., and Butler-Wu, S.M. (2013) *MALDI-TOF mass spectrometry for microorganism identification*, Clinics in Laboratory Medicine, Vol. 33, pp. 589–609.

Fajrial, A.K., He, Q.Q., Wirusanti, N.I., Slansky, J.E., and Ding, X. (2020) *A review of emerging physical transfection methods for CRISPR/Cas9-mediated gene editing*, Theranostics, Vol. 10, pp. 5532–5549.

Fujimoto, T., Miyoshi, D., Tateishi-Karimata, H., and Sugimoto, N. (2009) *Thermal stability and hydration state of DNA G-quadruplex regulated by loop regions*, Nucleic Acids Symposium Series, Vol. 53, pp. 237–238.

Galardin, M., Koumoutsi, A., Herrera-Dominguez, L., Varela, J.A.C., Telzerow, A., Wagih, O., Wartel, M., Clermont, O., Denamur, E., Typas, A., and Beltrao, P. (2017) *Phenotype prediction in an Escherichia coli strain panel*, bioRxiv, Vol. 6, pp. 141879.

Galluccio, M., Pantanella, M., Giudice, D., Brescia, S., and Indiveri, C. (2020) *Low temperature bacterial expression of the neutral amino acid transporters SLC1A5 (ASCT2), and SLC6A19 (B0AT1)*, Molecular Biology Reports, Vol. 47, pp. 7283–7289.

Gasperotti, A., Göing, S., Fajardo-Ruiz, E., Forné, I., and Jung, K. (2020) *Function and regulation of the Pyruvate transporter CstA in escherichia coli*, International

Journal of Molecular Sciences, Vol. 21, pp. 1–17.

González, V., Guo, K., Hurley, L., and Sun, D. (2009) *Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein*, Journal of Biological Chemistry, Vol. 284, pp. 23622–23635.

Graham, D.W. (2015) *MALDI-TOF mass spectrometry : an emerging technology for microbial identification and diagnosis* Vol. 6, pp. 1–16.

Granato, L.M., Picchi, S.C., De Oliveira Andrade, M., Takita, M.A., De Souza, A.A., Wang, N., and MacHado, M.A. (2016) *The ATP-dependent RNA helicase HrpB plays an important role in motility and biofilm formation in Xanthomonas citri subsp. Citri*, BMC Microbiology, Vol. 16, pp. 1–14.

Grinde, B. (2013) *Herpesviruses: latency and reactivation – viral strategies and host response*, Journal of Oral Microbiology, Vol. 5, pp. 22766.

Gutierrez, J.A., and Csonka, L.N. (1995) *Isolation and characterization of adenylate kinase (adk) mutations in Salmonella typhimurium which block the ability of glycine betaine to function as an osmoprotectant*, Journal of Bacteriology, Vol. 177, pp. 390–400.

Guyot, S., Pottier, L., Hartmann, A., Ragon, M., Hauck Tiburski, J., Molin, P., Ferret, E., and Gervais, P. (2014) *Extremely rapid acclimation of Escherichia coli to high temperature over a few generations of a fed-batch culture during slow warming*, MicrobiologyOpen, Vol. 3, pp. 52–63.

Hänsel-hertsch, R., Antonio, M. Di, and Balasubramanian, S. (2017) *DNA G-quadruplexes in the human genome: detection, functions therapeutic potential*, Nature Publishing Group, Vol. 5, pp. 279-284.

Harris, L.M., and Merrick, C.J. (2015) *G-Quadruplexes in Pathogens: A Common Route to Virulence Control?*, PLoS Pathogens, Vol. 11, pp. 1–15.

Holland-Staley, C.A., Lee, K., Clark, D.P., and Cunningham, P.R. (2000) *Aerobic activity of Escherichia coli alcohol dehydrogenase is determined by a single amino acid*, Journal of Bacteriology, Vol. 182, pp. 6049–6054.

Hrabák, J., Chudác'ková, E., and Walková, R. (2013) *Matrix-assisted laser desorption*

ionization-time of flight (MALDITOF) mass spectrometry for detection of antibiotic resistance mechanisms: From research to routine diagnosis, Clinical Microbiology Reviews, Vol. 26, pp. 103–114.

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 and Development, Vol. 31, pp. 2296–2309.

Huang, Z.L., Dai, J., Luo, W.H., Wang, X.G., Tan, J.H., Chen, S. Bin, and Huang, Z.S. (2018) *Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions*, Journal of the American Chemical Society, Vol. 140, pp. 17945–17955.

Idalia, V.-M.N., and Bernardo, F. (2017) Escherichia coli *as a Model Organism and Its Application in Biotechnology*, *Escherichia coli* - Recent Advances on Physiology, Pathogenesis and Biotechnological Applications, Vol. 7, pp. 801-815.

Islam, M.S., Aryasomayajula, A., and Selvaganapathy, P.R. (2017) *A review on macroscale and microscale cell lysis methods*, Micromachines, Vol. 8, pp. 83-110.

Jain, N., Mishra, S.K., Shankar, U., Jaiswal, A., Sharma, T.K., Kodgire, P., and Kumar, A. (2020) *G-quadruplex stabilization in the ions and maltose transporters gene inhibit Salmonella enterica growth and virulence*, Genomics, Vol. 112, pp. 4863–4874.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M.N., Yan, T., and Ishii, S. (2017) *Environmental Escherichia coli: ecology and public health implications—a review*, Journal of Applied Microbiology, Vol. 123, pp. 570–581.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013) *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*, Nature Biotechnology, Vol. 31, pp. 233–239.

Juers, D.H., Matthews, B.W., and Huber, R.E. (2012) *LacZ β-galactosidase: Structure and function of an enzyme of historical and molecular biological importance*, Protein Science, Vol. 21, pp. 1792–1807.

Jutras, B.L., Verma, A., and Stevenson, B. (2012) *Identification of Novel DNA-Binding proteins Using DNA-Affinity Chromatography/Pull Down*, Current Protocols

in Microbiology, Vol. 1, pp. 1–13.

Kaltashov, I.A., Bobst, C.E., and Abzalimov, R.R. (2013) *Mass spectrometry-based methods to study protein architecture and dynamics*, Protein Science, Vol. 22, pp. 530–544.

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* Vol. 44, pp. 9083–9095.

Katouli, M. (2010) *Population structure of gut Escherichia coli and its role in development of extra-intestinal infections*, Iranian Journal of Microbiology, Vol. 2, pp. 59–72.

Kaznadzey, A., Shelyakin, P., Belousova, E., Eremina, A., Shvyreva, U., Bykova, D., Emelianenko, V., Korosteleva, A., Tutukina, M., and Gelfand, M.S. (2018) *The genes of the sulphoquinovose catabolism in Escherichia coli are also associated with a previously unknown pathway of lactose degradation*, Scientific Reports, Vol. 8, pp. 1–12.

Khalil, A.M. (2020) *The genome editing revolution: review*, Journal of Genetic Engineering and Biotechnology, Vol. 18, pp. 1-16.

Kharel, P., Balaratnam, S., Beals, N., and Basu, S. (2020) *The role of RNA G-quadruplexes in human diseases and therapeutic strategies*, Wiley Interdisciplinary Reviews: RNA, Vol. 11, pp. 1–20.

Kikin, O., D'Antonio, L., and Bagga, P.S. (2006) *QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences*, Nucleic Acids Research, Vol. 34, pp. 676–682.

Kim, M.Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D., and Hurley, L.H. (2003) *The different biological effects of Telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures*, Cancer Research, Vol. 63, pp. 3247–3256.

Kosiol, N., Juranek, S., Brossart, P., Heine, A., and Paeschke, K. (2021) *G-quadruplexes: a promising target for cancer therapy*, Molecular Cancer, Vol. 20, pp.

1–18.

Kota, S., Dhamodharan, V., Pradeepkumar, P.I., and Misra, H.S. (2015) *G-quadruplex forming structural motifs in the genome of Deinococcus radiodurans and their regulatory roles in promoter functions,* Applied Microbiology and Biotechnology, Vol. 99, pp. 9761-9169.

Krafčíková, P., Demkovičová, E., Halaganová, A., and Víglaský, V. (2017) *Putative HIV and SIV G-Quadruplex Sequences in Coding and Noncoding Regions Can Form G-Quadruplexes*, Journal of Nucleic Acids, Vol. 2017, pp. 1-13.

Lane, A.N., Chaires, J.B., Gray, R.D., and Trent, J.O. (2008) *Stability and kinetics of G-quadruplex structures*, Nucleic Acids Research, Vol. 36, pp. 5482–5515.

Lane, W.J., and Darst, S.A. (2010) *Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis*, Journal of Molecular Biology, Vol. 395, pp. 671–685.

Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. jie, Chen, T., and Zhao, X. (2015) *Metabolic engineering of Escherichia coli using CRISPR-Cas9 meditated genome editing*, Metabolic Engineering, Vol. 31, pp. 13–21.

Lista, M.J., Martins, R.P., Billant, O., Contesse, M.A., Findakly, S., Pochard, P., Daskalogianni, C., Beauvineau, C., Guetta, C., Jamin, C., Teulade-Fichou, M.P., Fåhraeus, R., Voisset, C., and Blondel, M. (2017) *Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA*, Nature Communications, Vol. 8, pp. 1-13.

Liu, W., Zhu, H., Zheng, B., Cheng, S., Fu, Y., Li, W., Lau, T.C., and Liang, H. (2012) *Kinetics and mechanism of G-quadruplex formation and conformational switch in a G-quadruplex of PS2.M induced by Pb2+*, Nucleic Acids Research, Vol. 40, pp. 4229–4236.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010) *Comparison of 61 Sequenced Escherichia coli Genomes*, Microbial Ecology, Vol. 60, pp. 708–720.

Lundgren, M., Charpentier, E., and Fineran, P.C. (2015) *CRISPR: Methods and protocols*, CRISPR: Methods and Protocols, pp. 1–366.

Lyu, K., Chow, E.Y.C., Mou, X., Chan, T.F., and Kwok, C.K. (2021) *RNA G-quadruplexes (rG4s): Genomics and biological functions*, Nucleic Acids Research, Vol. 49, pp. 5426–5450.

Mao, S., Ghanbarian, A.T., Spiegel, J., Cuesta, S.M., Antonio, M. Di, Marsico, G., and Hänsel-hertsch, R. (2019) *Europe PMC Funders Group Europe PMC Funders Author Manuscripts DNA G-quadruplex structures mould the DNA methylome* Vol. 25, pp. 951–957.

Masai, H., and Tanaka, T. (2020) *G-quadruplex DNA and RNA: Their roles in regulation of DNA replication and other biological functions*, Biochemical and Biophysical Research Communications, Vol. 531, pp. 25–38.

McRae, E.K.S., Booy, E.P., Padilla-Meier, G.P., and McKenna, S.A. (2017) *On characterizing the interactions between proteins and guanine quadruplex structures of nucleic acids*, Journal of Nucleic Acids, Vol. 2017, pp. 1-12.

Meissner, B. (2019) *Chapter 30*, Spirits in the Grass, Vol. 1855, pp. 222–224.

Mendoza, O., Bourdoncle, A., Boulé, J.B., Brosh, R.M., and Mergny, J.L. (2016) *G-quadruplexes and helicases*, Nucleic Acids Research, Vol. 44, pp. 1989–2006.

Mishra, S.K., Shankar, U., Jain, N., Sikri, K., Tyagi, J.S., Sharma, T.K., Mergny, J.L., and Kumar, A. (2019) *Characterization of G-Quadruplex Motifs in espB, espK, and cyp51 Genes of Mycobacterium tuberculosis as Potential Drug Targets*, Molecular Therapy - Nucleic Acids, Vol. 16, pp. 698–706.

Moon, J., Han, J.H., Kim, D.Y., Jung, M. joon, and Kim, S.K. (2015) *Effects of deficient of the Hoogsteen base-pairs on the G-quadruplex stabilization and binding mode of a cationic porphyrin*, Biochemistry and Biophysics Reports, Vol. 2, pp. 29–35.

Muraki, K., Nyhan, K., Han, L., and Murnane, J.P. (2012) *Mechanisms of telomere loss and their consequences for chromosome instability*, Frontiers in Oncology, Vol. 2, pp. 1–13.

Nair, R. V., Bennett, G.N., and Papoutsakis, E.T. (1994) *Molecular characterization of an aldehyde/alcohol dehydrogenase gene from Clostridium acetobutylicum ATCC*

824, Journal of Bacteriology, Vol. 176, pp. 871–885.

Nakamura, S., and Minamino, T. (2019) *Flagella-driven motility of bacteria*, Biomolecules, Vol. 9, pp. 1-23.

Nakanishi, C., and Seimiya, H. (2020) *G-quadruplex in cancer biology and drug discovery*, Biochemical and Biophysical Research Communications, Vol. 531, pp. 45–50.

Neidle, S. (2016) *Quadruplex Nucleic Acids as Novel Therapeutic Targets*, Journal of Medicinal Chemistry, Vol. 59, pp. 5987–6011.

Neidle, S. (2010) *Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer*, FEBS Journal, Vol. 277, pp. 1118–1125.

Niu, K., Xiang, L., Jin, Y., Peng, Y., Wu, F., Tang, W., Zhang, X., Deng, H., Xiang, H., Li, S., Wang, J., Song, Q., and Feng, Q. (2019) *Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates*, Nucleic acids research, Vol. 47, pp. 7306–7320.

Okeke, I.N., Lamikanra, A., Steinrück, H., and Kaper, J.B. (2000) *Characterization of Escherichia coli strains from cases of childhood diarrhea in provincial southwestern Nigeria*, Journal of Clinical Microbiology, Vol. 38, pp. 7–12.

Onel, B., Lin, C., and Yang, D. (2014) *DNA G-quadruplex and its potential as anticancer drug target*, Science China Chemistry, Vol. 57, pp. 1605–1614.

Padmanabhan, S., Banerjee, S., and Mandi, N. (2011) *Screening of Bacterial Recombinants: Strategies and Preventing False Positives*, Molecular Cloning - Selected Applications in Medicine and Biology.

Perrone, R., Lavezzo, E., Palù, G., and Richter, S.N. (2017) *Conserved presence of G-quadruplex forming sequences in the Long Terminal Repeat Promoter of Lentiviruses*, Scientific Reports, Vol. 7, pp. 1–11.

Qin, R., Li, J., Liu, X., Xu, R., Yang, J., and Wei, P. (2020) *SpCas9-NG self-targets the sgRNA sequence in plant genome editing*, Nature Plants, Vol. 6, pp. 197–201.

Rajendran, A., Endo, M., Hidaka, K., Lan, P., Tran, T., Mergny, J., Gorelick, R.J., and Sugiyama, H. (2013) *HIV - 1 Nucleocapsid Proteins as Molecular Chaperones for*, Journal of American chemical society, Vol. 135, pp. 18575–18585.

Rath, D., Amlinger, L., Rath, A., and Lundgren, M. (2015) *The CRISPR-Cas immune system: Biology, mechanisms and applications*, Biochimie, Vol. 117, pp. 119–128.

Reshes, G., Vanounou, S., Fishov, I., and Feingold, M. (2008) *Cell shape dynamics in Escherichia coli*, Biophysical Journal, Vol. 94, pp. 251–264.

Rhodes, D., and Lipps, H.J. (2015) *Survey and summary G-quadruplexes and their regulatory roles in biology*, Nucleic Acids Research, Vol. 43, pp. 8627–8637.

Rigo, R., Palumbo, M., and Sissi, C. (2017) *G-quadruplexes in human promoters: A challenge for therapeutic applications*, Biochimica et Biophysica Acta - General Subjects, Vol. 1861, pp. 1399–1413.

Rodríguez-Rodríguez, D.R., Ramírez-Solís, R., Garza-Elizondo, M.A., Garza-Rodríguez, M.D.L., and Barrera-Saldaña, H.A. (2019) *Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review)*, International Journal of Molecular Medicine, Vol. 43, pp. 1559–1574.

Rosano, G.L., and Ceccarelli, E.A. (2014) *Recombinant protein expression in Escherichia coli: Advances and challenges*, Frontiers in Microbiology, Vol. 5, pp. 1–17.

Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E.P.C., and Bikard, D. (2021) *The impact of genetic diversity on gene essentiality within the Escherichia coli species*, Nature Microbiology, Vol. 6, pp. 301–312.

Ruggiero, E., and Richter, S.N. (2018) *Survey and summary G-quadruplexes and G-quadruplex ligands: Targets and tools in antiviral therapy*, Nucleic Acids Research, Vol. 46, pp. 3270–3283.

Saranathan, N., and Vivekanandan, P. (2018) *G-Quadruplexes: More Than Just a Kink in Microbial Genomes*, Trends in Microbiology, Vol. 148, pp. 1–16.

Satkunanathan, S., Thorpe, R., and Zhao, Y. (2017) *The function of DNA binding*

protein nucleophosmin in AAV replication, Virology, Vol. 510, pp. 46–54.

Sauer, M., and Paeschke, K. (2017) *G-quadruplex unwinding helicases and their function in vivo* , Biochemical Society Transactions, Vol. 45, pp. 1173–1182.

Shasmal, M., Dey, S., Shaikh, T.R., Bhakta, S., and Sengupta, J. (2016) *E. coli metabolic protein aldehyde alcohol dehydrogenase-E binds to the ribosome: a unique moonlighting action revealed*, Nature Publishing Group, Vol. 6, pp. 1–12.

Shulman, S.T., Friedmann, H.C., and Sims, R.H. (2007) *Theodor Escherich: the first pediatric infectious diseases physician?*, Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, Vol. 45, pp. 1025–1029.

Song, J., Perreault, J.-P., Topisirovic, I., and Richard, S. (2016) *RNA G-quadruplexes and their potential regulatory roles in translation*, Translation, Vol. 4, pp. 1-16.

Souza, V., Rocha, M., Valera, A., and Eguiarte, L.E. (1999) *Genetic structure of natural populations of Escherichia coli in wild hosts on different continents*, Applied and Environmental Microbiology, Vol. 65, pp. 3373–3385.

Stanssens, P., Zabeau, M., Meersseman, G., Remes, G., Gansemans, Y., Storm, N., Hartmer, R., Honisch, C., Rodi, C.P., Böcker, S., and van den Boom, D. (2004) *High-throughput MALDI-TOF discovery of genomic sequence polymorphisms*, Genome Research, Vol. 14, pp. 126–133.

Structures, H.T.G., Lin, C., and Yang, D. (2017) *Chapter 17* Vol. 1587, pp. 151-162.

Sui, H., Chen, Q., and Imamichi, T. (2020) *A pull-down assay using DNA/RNA-conjugated beads with a customized competition strategy: An effective approach to identify DNA/RNA binding proteins*, MethodsX, Vol. 7, pp. 1-5.

Sun, Z.Y., Wang, X.N., Cheng, S.Q., Su, X.X., and Ou, T.M. (2019) *Developing novel G-quadruplex ligands: From interaction with nucleic acids to interfering with nucleic acid–protein interaction*, Molecules, Vol. 24, pp. 396-425.

Thakur, R.K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J.L., Basundra, R., Kumar, A., and Chowdhury, S. (2009) *Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression*, Nucleic Acids Research, Vol. 37, pp. 172–183.

Tian, T., Chen, Y.Q., Wang, S.R., and Zhou, X. (2018) *G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting*, Chem, Vol. 4, pp. 1314–1344.

Tikhonova, P., Pavlova, I., Isaakova, E., Tsvetkov, V., Bogomazova, A., Vedekhina, T., Luzhin, A. V, Sultanov, R., Severov, V., Klimina, K., Kantidze, O.L., Pozmogova, G., Lagarkova, M., and Varizhuk, A. (2021) *DNA G-Quadruplexes Contribute to CTCF Recruitment*, International Journal of Molecular Sciences, Vol. 22, pp. 1–17.

Vento, J.M., Crook, N., and Beisel, C.L. (2019) *Barriers to genome editing with CRISPR in bacteria*, Journal of Industrial Microbiology and Biotechnology, Vol. 46, pp. 1327–1341.

Verma, S.C., Qian, Z., and Adhya, S.L. (2019) *Architecture of the Escherichia coli nucleoid*, PLoS Genetics, Vol. 12, pp. 1-35.

Vila, J., Sáez-López, E., Johnson, J.R., Römling, U., Dobrindt, U., Cantón, R., Giske, C.G., Naas, T., Carattoli, A., Martínez-Medina, M., Bosch, J., Retamar, P., Rodríguez-Banõ, J., Baquero, F., and Soto, S.M. (2016) *Escherichia coli: An old friend with new tidings*, FEMS Microbiology Reviews, Vol. 40, pp. 437–463.

Vlasenok, M., Levchenko, O., Basmanov, D., Klinov, D., Varizhuk, A., and Pozmogova, G. (2018) *Data set on G4 DNA interactions with human proteins*, Data in Brief, Vol. 18, pp. 348–359.

Walia, R., and Chaconas, G. (2013) *Suggested Role for G4 DNA in Recombinational Switching at the Antigenic Variation Locus of the Lyme Disease Spirochete*, PLoS ONE, Vol. 8, pp. 1-8.

Wang, Q., Liu, J., Chen, Z., Zheng, K., Chen, C., Hao, Y., and Tan, Z. (2011) *G-quadruplex formation at the 3 0 end of telomere DNA inhibits its extension by telomerase , polymerase and unwinding by helicase* Vol. 86, pp. 1–9.

Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A., and Huse, J.T. (2019) *G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma*, Nature Communications, Vol. 10, pp. 1-14.

Wang, Y., Hu, Y., Wu, T., Liu, H., Zhang, L., Zhou, X., and Shao, Y. (2015) *Specific*

G-quadruplex structure recognition of human telomeric RNA over DNA by a fluorescently activated hyperporphyrin, The Analyst, Vol. 140, pp. 5169–5175.

Webb, D.C., Rosenberg, H., and Cox, G.B. (1992) *Mutational analysis of the Escherichia coli phosphate-specific transport system, a member of the traffic ATPase (or ABC) family of membrane transporters. A role for proline residues in transmembrane helices*, Journal of Biological Chemistry, Vol. 267, pp. 24661–24668.

Webster, J., and Oxley, D. (2012) *Protein identification by MALDI-TOF mass spectrometry*, Methods in Molecular Biology, Vol. 800, pp. 227–240.

Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., and Eperon, I.C. (2018) *Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X*, Nucleic Acids Research, Vol. 46, pp. 886–896.

Westermann, L., Neubauer, B., and Köttgen, M. (2021) *Nobel Prize 2020 in Chemistry honors CRISPR: a tool for rewriting the code of life*, Pflugers Archiv European Journal of Physiology, Vol. 473, pp. 1–2.

Winnerdy, F.R., and Phan, A.T. (2020) *Quadruplex structure and diversity*, Annual Reports in Medicinal Chemistry, Vol. 54, pp. 45–73.

Wood, M.L., and Royle, N.J. (2017) *Chromosomally Integrated Human Herpesvirus 6: Models of Viral Genome Release from the Telomere and Impacts on Human Health.* Viruses, Vol. 7, pp. 184-198.

Woude, M.W. Van Der, and Ba, A.J. (2004) *Phase and Antigenic Variation in Bacteria*, Clinical Microbiology Reviews, Vol. 17, pp. 581–611.

Wu, X., Kriz, A.J., and Sharp, P.A. (2014) *Target specificity of the CRISPR-Cas9 system*, Quantitative Biology, Vol. 2, pp. 59–70.

Yamashino, T., Isomura, M., Ueguchi, C., and Mizuno, T. (1998) *The yhhP gene encoding a small ubiquitous protein is fundamental for normal cell growth of Escherichia coli*, Journal of Bacteriology, Vol. 180, pp. 2257–2261.

Yoshimi, K., Oka, Y., Miyasaka, Y., Kotani, Y., Yasumura, M., Uno, Y., Hattori, K., Tanigawa, A., Sato, M., Oya, M., Nakamura, K., Matsushita, N., Kobayashi, K., and

Mashimo, T. (2021) *Combi-CRISPR: combination of NHEJ and HDR provides efficient and precise plasmid-based knock-ins in mice and rats*, Human Genetics, Vol. 140, pp. 277–287.

You, H., Wu, J., Shao, F., and Yan, J. (2014) *Stability and Kinetics of c- MYC Promoter G - Quadruplexes Studied by Single-Molecule Manipulation,* American Chemical Society, Vol. 7, pp. 10–13.

You, J., Li, H., Lu, X.M., Li, W., Wang, P.Y., Dou, S.X., and Xi, X.G. (2017) *Effects of monovalent cations on folding kinetics of G-quadruplexes*, Bioscience Reports, Vol. 37, pp. 1–12.

Zeinert, R., Martinez, E., Schmitz, J., Senn, K., Usman, B., Anantharaman, V., Aravind, L., and Waters, L.S. (2018) *Structure–function analysis of manganese exporter proteins across bacteria,* Journal of Biological Chemistry, Vol. 293, pp. 5715–5730.

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A.K., Taboada, E., and Gannon, V.P.J. (2007) *Genome evolution in major Escherichia coli O157:H7 lineages*, BMC Genomics, Vol. 8, pp. 121-137.

Zhang, Z., Dai, J., Veliath, E., Jones, R.A., and Yang, D. (2009) *Structure of a two-G-tetrad intramolecular G-quadruplex formed by variant human telomeric sequence in K+ solution: Insights into the interconversion of human telomeric G-quadruplex structures*, Nucleic Acids Research, Vol. 38, pp. 1009–1021.