Illinois State University

ISU ReD: Research and eData

10-2-2017

The Efficiency Of Dna Repair Processes In Proximity Of G4 Dna Structures

Gloria Alvarado
Illinois State University, gloalvara@ilstu.edu

THE EFFICIENCY OF DNA REPAIR PROCESSES IN PROXIMITY

OF G4 DNA STRUCTURES

Gloria Alvarado

73 Pages

Guanine-rich segments of genomes promote the formation of a four-stranded DNA structure called G-quadruplex (G4 DNA). While evolutionarily conserved, guanine repeats show increased rates of mutagenesis and instability, and this is likely associated with G4 formation. In this thesis, I sought to clarify the molecular sources of genetic instability by testing the repair, excision, and synthesis through guanine-rich DNA. Using *E. coli* UDG and human SMUG1, I found that uracil is inefficiently repaired in proximity to G4 DNA *in vitro*. Investigation of *E. coli* exonuclease III and exonuclease I activity indicated that G4 DNA is a block to activity and complete DNA digestion. Finally, I studied the role of the Y-family translesion polymerase DPO4 in G4 DNA replication bypass. Consistent with known translesion synthesis activities, experiments using Y-family translesion polymerase DPO4 showed an increased fidelity across from structured DNA compared to single-stranded templates. Together these data indicate a possible source of increased mutagenesis around G4 DNA, suggesting that lower levels of DNA repair may promote base substitutions within guanine repeats. In order to limit replication stalling at G4 DNA, the Y-family translesion polymerases may be critical for ensuring faithful replication through DNA repeats and loss of such activities would increase the potential for recombination and genome instability at repetitive loci.

THE EFFICIENCY OF DNA REPAIR PROCESSES IN PROXIMITY

OF G4 DNA STRUCTURES

GLORIA ALVARADO

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

School of Biological Sciences

ILLINOIS STATE UNIVERSITY

2017

THE EFFICIENCY OF DNA REPAIR PROCESSES IN PROXIMITY

OF G4 DNA STRUCTURES

GLORIA ALVARADO

COMMITTEE MEMBERS:

Erik Larson, Chair

Tom Hammond

Alysia Vrailas – Mortimer

ACKNOWLEDGMENTS

CONTENTS

TABLES

FIGURES

CHAPTER I: BACKGROUND ON G-QUADRUPLEX

Introduction

The human genome contains thousands of sequence motifs with tandem repeats of guanine that support the intracellular formation of four-stranded DNA structures called G-quadruplex (G4 DNA). G4 DNA was originally recognized as a structure that forms at telomeres, but it now seems clear that guanine-rich loci throughout the human genome are not only abundant but also fold into G4 structures (Bochman, et al., 2012; Rhodes and Lipps, 2015). Those structures are highly stable, more so than duplex DNA. Relevant to disease, G4 DNA appears to form near or within cancer-related genes. In particular, cancer genes prone to rearrangements in T-cell leukemia and follicular lymphoma have been shown to adopt G4 conformations near the DNA break points. (Nambiar, et al., 2011; Nambiar, et al., 2013; Williams, et al., 2015).

In general, the minimal sequence needed to support G4 formation is at least four sets of three tandem guanines, each separated by one to seven non-guanine bases, shown in Figure 1A (Huppert and Balasubramanian, 2005; Todd, et al., 2005). The guanines present in each set interact through Hoogsteen bonding, forming a tetrad. Tetrads stack on top of each other to form the four-stranded DNA structure (Figure 1B). Hoogsteen bonds are inherently thermostable and each tetrad requires the presence of a monovalent cation located at the center of the tetrad. The monovalent cation acts on the inward facing oxygen of self-paired guanines (Bochman, et al., 2012; Maizels, 2006; Rhodes and Lipps, 2015). Different monovalent cations present distinctive stabilization properties to G4, with potassium and sodium generally being the most stabilizing and lithium and cesium being the least (Sen and Gilbert, 1990; Phan, et al., 2006; Rhodes and

Lipps, 2015). Both potassium and sodium are found within the cell at concentrations that are supportive of spontaneous G4 DNA formation.

The degree of G4 DNA stability depends largely on the sequence motif, stabilizing monovalent cations, loop size between guanines, and guanine alignment (Bochman, et al., 2012; Rhodes and Lipps, 2015). Structure fluidity and diversity occurs for G4 DNAs folded *in vitro* (Lee, et al., 2005). That diversity in G4 DNA conformation creates the possibility that many structures coexist at equilibrium (Figure 2), though structure conformation and stability is largely dependent on sequence, with stability dependent on the quantity of guanines present (Maizels, 2006; Bochman, et al., 2012). G4 DNA can form within a single DNA strand forming an intramolecular structure (Figure 2A), or between two or four DNA strands forming an intermolecular structure (Figure 2B). Intramolecular G4 DNA forms independent of sequence concentration while intermolecular G4 DNA requires a high concentration of DNA *in vitro* (Maizels, 2006). G4 DNA structure formation can be controlled to an extent by adjusting the reaction temperature and ions present (Sen and Gilbert, 1990). Sen and Gilbert found that the formation of complex four-stranded intermolecular G4 DNA is dependent on the presence of sodium not potassium. The resulting structure is highly thermostable and does not require stabilizing salts in solution once folded, as discussed in Chapter II. This highlights the incredible stability of G4 DNA compared to duplexes of similar nucleotide content.

Whether G4 DNA forms *in vivo* was highly debated, but it is now widely recognized to as a component of normal cell physiology. In order for G4 DNA to form in the cell, the guanine-rich sequence needs to be freed from complement. While primarily double-stranded (DS) and condensed in nucleosomes, the processes of replication, transcription, and DNA repair does generate transiently single-stranded DNA (Bochman, et al., 2012; Maizels, 2006; Rhodes and

Lipps, 2015), providing an opportunity for alternate structure formation. Initial evidence of *in vivo* formation of G4 DNA came from studying telomeres (Schaffitzel, et al., 2001; Paeschke, et al., 2005). Human telomeres contain the repeat $(TTAGGG)_{1-1500}$ that has the potential to form G4 DNA (Bochman, et al., 2012; Rhodes and Lipps, 2015). Treatment of human cells with G4 DNA stabilizing molecules leads to telomere shortening, indicating the presence of activities that resolve G4 DNA during replication, while also suggesting the structures serve as a cap for the ends of linear chromosomes (Sun, et al. 1997; Rhodes and Lipps, 2015). However, even given the probable formation of G4 DNA at telomeres, the functional contributions of G4 formation at other genome sites is less well understood.

Utilizing the minimal G4 definition (Figure 1A), computation analyses identified over 300,000 potential G4-capable sequences throughout the human genome, indicating that telomeres are not the sole location of these sequences (Huppert and Balasubramanian, 2005; Todd, et al., 2005; Capra, et al., 2010; Vasquez and Wang, 2013). These studies found that G4-capable sequences are highly conserved in humans and between different species, and the genomic locations are non-random (Capra, et al., 2010; Huppert and Balasubramanian, 2005; Todd, et al., 2005). Data showed G4-capable sequences upstream of the transcriptional start sites for 50% of human genes, in ribosomal DNA, and in immunoglobulin heavy-chain switch regions (Huppert and Balasubramanian, 2005; Todd, et al., 2005; Capra, et al., 2010; Maizels, 2006). Furthermore, bioinformatic analyses revealed a higher likelihood of G4-capable sequences at promoters for human oncogenes and regulatory genes compared to tumor-suppressor and housekeeping genes, and G4-capable sequences are more likely to be on the non-template strand (Bochman, et al., 2012). Another study found the mismatch repair protein MutSα heterodimer binds to G4 DNA, implying a role for the factor in Ig class switch recombination (Larson, et al.,

2005). These data support the hypothesis that G4-capable sequences have been evolutionarily retained, possibly due to their ability to regulate gene expression by either blocking or enhancing gene transcription rates (Eddy and Maizels, 2009; Bochman, et al., 2012; Rhodes and Lipps, 2015; Vasquez and Wang, 2013).

Given that G4-capable sequences are found at promoter regions, it seems likely that pathways for the regulation of G4 DNA formation exist. Indeed, several helicases are now known to be associated with the unwinding G4 DNA. This includes FANCJ, PIF1, and BLM, and WRN, which have been associated with increased cancer risk and premature aging disease (Bochman, et al., 2012; Vasquez and Wang, 2013; Rhodes and Lipps, 2015). Genome instability during replication is probably prevented by G4 unwinding activities because the loss of FANCJ leads to large guanine deletions (London, et al., 2008; Wu, et al., 2008; Bochman, et al., 2012). Furthermore, chaperone proteins identified in *S. lemnae* and *S. cerevisiae* promote G4 DNA formation (Bochman, et al., 2012; Rhodes and Lipps, 2015), revealing the presence of another DNA structure-specific cellular activity.

Even with chaperones and helicases regulating G4 folding, failures in those pathways are probably linked to improper G4 formations that lead to site-specific genome instability (Maizels, 2015; Bochman, et al., 2012; Rhodes and Lipps, 2015; Vasquez and Wang, 2013). G4-capable sequences have been found within 500 bp of double-strand break sites and at hotspots for genetic instability in human disease (Vasquez and Wang, 2013). G4 DNA is also present at major translocation break sites responsible for blood cell cancers (Nambiar, et al., 2010; Nambiar, et al., 2013; Williams, et al., 2015). For example, a recent study found a large density of G4 DNA clustered at a common translocation break site t(1;19) in the oncogenic *TCF3* gene, a factor associated with a subset of acute lymphoblastic leukemias (Williams, et al., 2015). There is also

evidence that G4 DNA inhibits DNA replication, leading to stalled or collapsed replication forks, deletions, and complex rearrangements (Rhodes and Lipps, 2015; Vasquez and Wang, 2013). In short, improper or unregulated formation of G4 DNA at guanine-rich loci has the potential to cause mutagenesis and recombination (Maizles, 2015; Shah and Mirkin, 2015).

It would be of great value to increase our understanding of how DNA repair, replication, and processing occurs at or near G4 structures. Disruption of any of those activities is well known to result in mutagenesis and genomic instability. In this thesis, I investigate how G4 DNA affects base excision repair, exonuclease function, and replication. The genome is under constant attack by both endogenous and exogenous sources of DNA damage and the cell must be able to repair that damage independent of DNA conformational status. In order to better define how base excision repair (BER) responds to lesions near G4, I investigated uracil DNA glycosylase (UDG) activity next to those structures. Mismatch repair (MMR) and BER both require nuclease activities, so I extended my analysis to include exonuclease activity assays. Finally, I tested the polymerization capabilities of translesion polymerase DPO4 across from G4 DNA. The goal of my research is to gain a better understanding of how G4 DNA effects regular DNA metabolic activities, specifically DNA repair, to better clarify the molecular sources of mutagenesis and instability at guanine-rich loci.

References

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics*, *13*(11), 770-780.

Capra, J. A., Paeschke, K., Singh, M., & Zakian, V. A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS computational biology*, *6*(7), e1000861.

Eddy, J., & Maizels, N. (2009). Selection for the G4 DNA motif at the 5′ end of human genes. *Molecular carcinogenesis*, *48*(4), 319-325.

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, *33*(9), 2908-2916.

Larson, E. D., Duquette, M. L., Cummings, W. J., Streiff, R. J., & Maizels, N. (2005). MutSα binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Current biology*, *15*(5), 470-474.

Lee, J. Y., Okumus, B., Kim, D. S., & Ha, T. (2005). Extreme conformational diversity in human telomeric DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 18938-18943.

London, T. B., Barber, L. J., Mosedale, G., Kelly, G. P., Balasubramanian, S., Hickson, I. D., Boulton, S. J., & Hiom, K. (2008). FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *Journal of Biological Chemistry*, *283*(52), 36132-36139.

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature structural & molecular biology*, *13*(12), 1055.

Maizels, N. (2015). G4-associated human diseases. *EMBO reports*, e201540607.

Nambiar, M., Goldsmith, G., Moorthy, B. T., Lieber, M. R., Joshi, M. V., Choudhary, B., Hosur, R. V., & Raghavan, S. C. (2010). Formation of a G-quadruplex at the BCL2 major breakpoint region of the t (14; 18) translocation in follicular lymphoma. *Nucleic acids research*, *39*(3), 936-948.

Nambiar, M., Srivastava, M., Gopalakrishnan, V., Sankaran, S. K., & Raghavan, S. C. (2013). G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t (10; 14) translocation in T-cell leukemia. *Molecular and cellular biology*, *33*(21), 4266-4281.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., & Lipps, H. J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature structural & molecular biology*, *12*(10), 847.

Phan, A. T., Kuryavyi, V., & Patel, D. J. (2006). DNA architecture: from G to Z. *Current opinion in structural biology*, *16*(3), 288-298.

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, *43*(18), 8627-8637.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., & Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences*, *98*(15), 8572-8577.

Sen, D., & Gilbert, W. (1990). A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, *344*(6265), 410-414.

Shah, K. A., & Mirkin, S. M. (2015). The hidden side of unstable DNA repeats: Mutagenesis at a distance. *DNA repair*, *32*, 106-112.

Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C.,
Neidle, S. and Hurley, L.H. (1997). Inhibition of human telomerase by a G-quadruplex-
interactive compound. *Journal of medicinal chemistry*, *40*(14), 2113-2116.

Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence
motifs in human DNA. *Nucleic acids research*, *33*(9), 2901-2907.

Vasquez, K. M., & Wang, G. (2013). The yin and yang of repair mechanisms in DNA structure-
induced genetic instability. *Mutation Research/Fundamental and Molecular Mechanisms
of Mutagenesis*, *743*, 118-131.

Williams, J. D., Fleetwood, S., Berroyer, A., Kim, N., & Larson, E. D. (2015). Sites of instability
in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures in vitro. *Frontiers in
genetics*, *6*, 177.

Wu, Y., Shin-ya, K., & Brosh, R. M. (2008). FANCJ helicase defective in Fanconia anemia and
breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Molecular and
cellular biology*, *28*(12), 4116-4128.

A.

$$GGG_+N_{1-7}GGG_+N_{1-7}GGG_+N_{1-7}GGG_+$$

B.
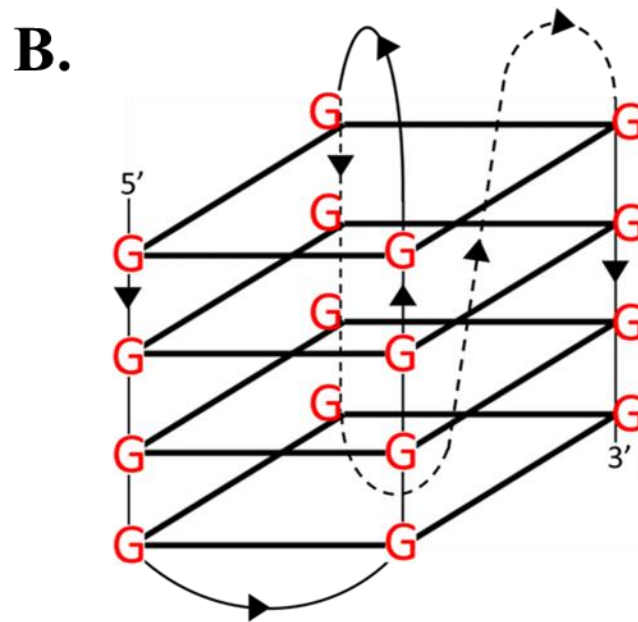
Figure 1. Generic G-Quadruplex (G4) Sequence and Structure. (A) The minimal G4-capable

sequence used when searching the genome for potential G4 DNA locations. (B) A model of a G4

structure.
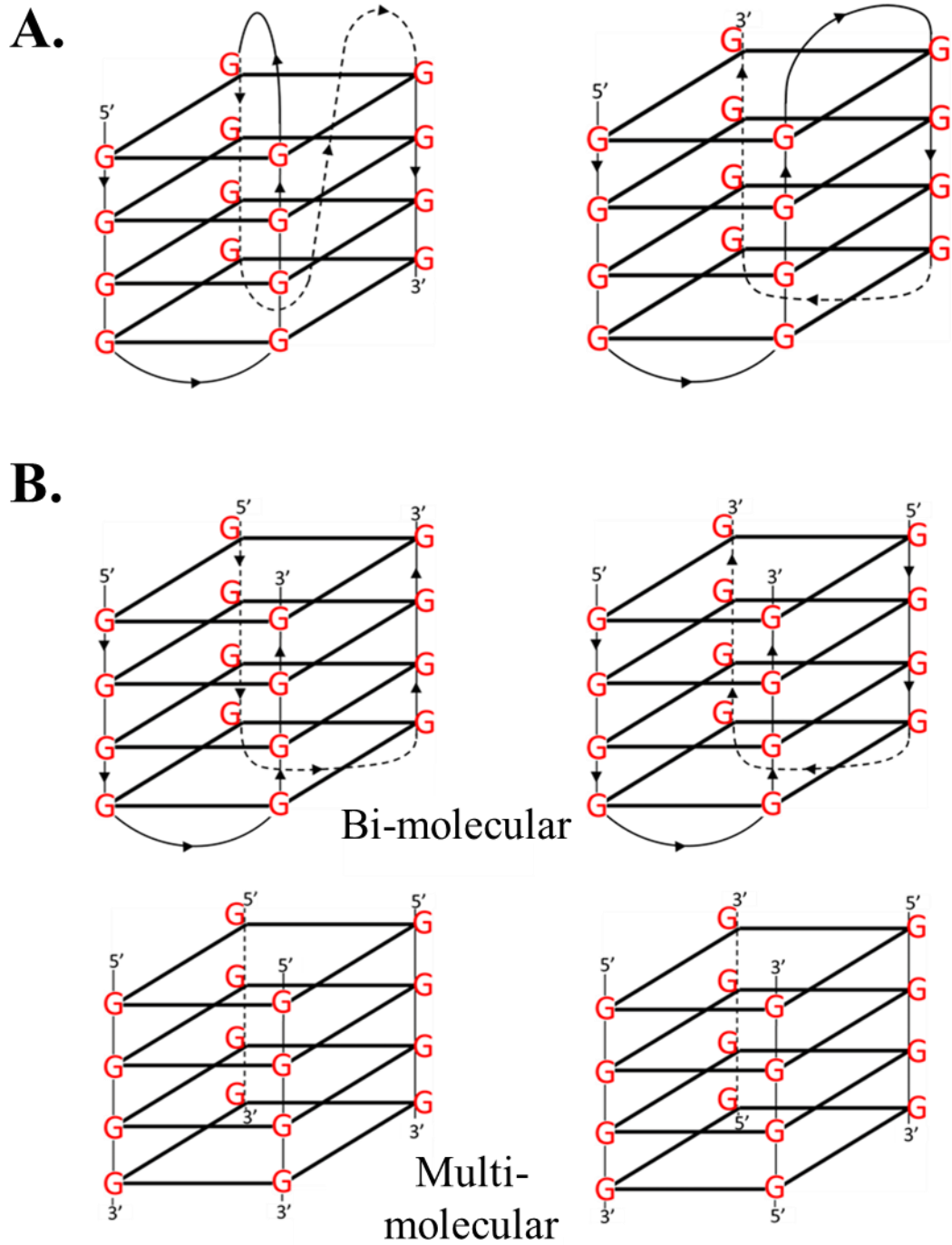
Figure 2. Intramolecular and Intermolecular G4 DNA. (A) Intramolecular G4 DNA conformations. (B) Intermolecular G4 DNA conformations with bi-molecular conformations on top and multi-molecular conformations on bottom.

CHAPTER II: BASE EXCISION REPAIR SURROUNDING G4

Introduction

All organisms must correct DNA damage in order to preserve genome integrity and the faithful transmission of genetic information to the next cell generation. In this chapter, I present my results on how G4 DNA impacts uracil repair activities through the Base Excision Repair (BER) pathway. Using *E. coli* uracil DNA glycosylase (UDG) and human single-strand-selective monofunctional uracil DNA glycosylase (hSMUG1) as models, I tested the extent that G4 DNA affects local uracil glycosylase activity.

Excision Repair

All excision repair pathways share the common feature of; 1., recognizing DNA damage, 2., excising the damage, and 3., DNA re-synthesis and ligation to repair the break. Therefore, a failure in lesion recognition leads to mutagenesis because repair is not initiated. In BER, glycosylases are the repair-initiating enzymes, and they are essential for reducing mutagenesis caused by the most common type of DNA damage in the genome, base damage (Krokan, et al., 2002). When core aspects of BER are lost, there is a predisposition to cancer, immune deficiencies, and neurological impacts (Krokan, et al., 2002; Brenerman, et al., 2014). DNA glycosylases come in two flavors: monofunctional with only base cleavage ability, and bifunctional with both base cleavage and endonuclease abilities (Sung and Demple, 2006; Brenerman, et al., 2014). Glycosylases are lesion specific, and they remove the altered DNA base to leave an abasic site, which then initiates BER (Figure 3). The abasic site is subsequently cleaved by an AP endonuclease, typically AP endonuclease 1 (APE1). The DNA break is resynthesized by DNA polymerase β (POLβ) and ligated by Ligase IIIα (LIGIIIα) to complete repair (Brenerman, et al., 2014). An alternate BER pathway, termed long-patch BER, removes a

11

longer section of DNA. This occurs when POLβ or replicative polymerases synthesize a new DNA strand while displacing the existing strand. The displaced strand is cleaved by flap-endonuclease FEN1 and either Ligase I or Ligase III re-joins the DNA (Sung and Demple, 2006). The type of BER pathway utilized depends on the type of lesion present and the responding glycosylase (Sung and Demple, 2006).

BER and MMR both correct genomic uracil, one of the most common types of DNA damage (Larson, et al., 2008; Krokan, et al., 2002). Uracil can be introduced into the human genome through misincorporation during replication, enzymatically by activation-induced cytosine deaminase (AID) activity in B cells, or from spontaneous cytosine deamination (Visnes, et al., 2009; Krokan, et al., 2002). Cytosine is an easy target for deamination because the reaction is hydrolytic (Lindahl, 1993). Spontaneous cytosine deamination in double-stranded (DS) DNA occurs at about 0.6% the rate of single-stranded (SS) DNA (Lindahl, 1993). As mentioned in Chapter I, the genome is SS during replication, transcription, or during DNA repair activities (Bochman, et al., 2012; Maizels, 2006; Rhodes and Lipps, 2015), and that exposes DNA to conditions favorable for cytosine deamination. Overall, any given cell sees 70-500 spontaneous cytosine deamination events per day, each with the potential to cause C→T substitutions (An, et al., 2005; Lindahl, 1993; Visnes, et al., 2009). This is a common source of mutagenesis in human cells and has the potential to accumulate in proliferating cells (Krawczak, et al., 1998; Brenerman, et al., 2014; Krokan, et al., 2002; Visnes, et al., 2009). Although uracil is generally mutagenic when introduced into the genome, it is an essential part of programmed immune system rearrangements (Petersen-Mahrt, et al., 2002).

In general, uracil glycosylases function by flipping the uracil base into their active site for cleavage of the *N*-glyosidic bond between the base and deoxyribose, leaving an abasic site

(Brenerman, et al., 2014). Humans have several uracil glycosylases for uracil repair (Krokan, et al., 2002; Visnes, et al., 2009). Uracil-*N*-glycosylase 1 and 2 (UNG1 and UNG2) are specific to uracil and function in the mitochondrial and nucleus, respectively. UNG1 is the only uracil glycosylase acting on mitochondrial DNA, but both UNG1 and UNG2 are active on SS and DS DNA. Single-strand-selective monofunctional uracil DNA glycosylase (hSMUG1) cleaves uracil along with 5-hydroxymethyluracil (HmU). Though the name implies hSMUG1 acts exclusively on SS DNA, it has comparable efficiency on both DS and SS DNA. T/U mismatch DNA glycosylase (TDG) cleaves both uracil and thymine opposite a mismatch with a strong preference for uracil cleavage. Finally, methyl-CpG-binding domain Protein 4 (MBD4) also cleaves uracil and thymine however, MBD4 repairs uracil and thymine present due to deamination of cytosine and guanine (Krokan, et al., 2002; Visnes, et al., 2009).

The importance of uracil base removal from the genome is underscored by the fact that organisms across all domains of life have evolved repair pathways for removal (Krokan et al., 2002). Evidence suggests that the major glycosylases in humans are UNG2 and SMUG1, probably because they can excise uracil from both SS and DS DNA (Kavli, et al., 2002; An, et al., 2005). While they perform a similar function, they act on DNA through distinct mechanisms (Pettersen, et al., 2007). Knockout/knockdown studies in mice confirm this finding, indicating that their functions are non-overlapping and a loss of either enzyme results in increased mutagenesis and disease (An, et al., 2005).

UNG2

Among enzymes, UNG2 has one the highest catalytic efficiencies and activity is specifically targeted to uracil lesions (Lindahl, et al., 1977; Drohat 1999, Visnes, et al., 2009). Originally thought to be absent in archaea, UNG2 is the most widespread and abundant of the

13

uracil glycosylases, and present in all domains of life (Krokan, et al., 2002). UNG2 is highly

conserved, with a striking conservation across Kingdoms. For instance, there is over 70%

identity between the human and bacterial protein (Olsen, et al., 1989). UNG2 can remove uracil

from U:G mismatches, which occur from the spontaneous deamination of cytosine, however, its

major role seems to be after replication (Kavli, et al., 2002; Visnes 2009). UNG2 is upregulated

during S-phase, degraded during the G2-phase, and co-localizes with replication foci (Kavli, et

al., 2002; Haug, et al., 1998; Visnes, et al., 2009). APE1 produces a stimulatory effect to UNG2,

and the presence of Mg^{++} increases its affinity to SS DNA 10-fold, further supporting roles in

genome maintenance in proliferating cells (Kavli, et al., 2002; Pettersen, et al., 2007).

SMUG1

SMUG1 has a much lower catalytic efficiency when compared to UNG2, but in addition

to uracil, it can excise 5-hydroxymethyluracil, $3,N^4$-ethenocytosine, and 5-fluorouracil lesions, in

preferential order (Kavli, et al., 2002; Pettersen, et al., 2007). The gene is mainly confined to

vertebrates and insects, but it is also found in some bacteria in place of UDG (Nilsen, et al.,

2001; Pettersen, et al., 2007). There is a high degree of conservation between active site and

lesion-binding domains for both UNG and SMUG1, however there is a DNA helix wedge motif

that is unique to SMUG1 (Figure 4) (Wibley, et al., 2003; Pettersen, et al., 2007). This DNA

helix wedge motif inserts itself into the minor groove and makes contact with both DNA strands

and adjacent nucleotide bases (Wibley, et al., 2003; Pettersen, et al., 2007). SMUG1 is the major

uracil glycosylase in non-proliferating cells, and that is due to the DNA binding properties of

SMUG1's DNA helix wedge motif (An, et al., 2005; Kavli, et al., 2002; Pettersen, et al., 2007;

Nilsen, et al., 2001). For example, the wedge motif allows for stabilized binding on abasic sites

that occurs preferentially across from guanines, indicating protection against DNA breaks and

repair of deaminated cytosines (Wibley et al., 2003; Kavli, et al., 2002; Pettersen, et al., 2007). In addition, evidence indicates APE1 has a stimulatory effect on SMUG1, even though SMUG1's high affinity for abasic sites may also inhibit APE1 activity (Kavli, et al., 2002; Pettersen, et al., 2007). Furthermore, while UNG2 localizes to replication foci, SMUG1 is found evenly dispersed throughout the cell, and its presence will arrest cell proliferation (Nilsen, et al., 2001; Kavli, et al., 2002; Pettersen, et al., 2007).

BER and G4 DNA

Opportunities for cytidine deamination coincide with G4 formation because both are enhanced when DNA is single stranded (Bochman, et al., 2012; Maizels, 2006; Rhodes and Lipps, 2015; Lindahl, 1993). Given that uracil in the genome is potentially mutagenic, it is important to know how G4 DNA affects repair factors that recognize and remove uracil. Therefore, a better understanding of how human cells deal with G4 DNA is important for defining the mechanisms of sequence-specific mutagenesis.

UNG2 activity is inhibited at uracils directly adjacent to G4 DNA but not at lesions located three bases away (Holton and Larson, 2015). Those findings imply that G4 DNA physically blocks uracil repair, but the extent of inhibition is not well defined. Here, I clarify the impact of G4 DNA on nearby uracil repair, testing activities of *E. coli* UDG and human SMUG1 (hSMUG1) at uracils placed next to G4 tetrads *in vitro*. I find that glycosylase activity is reduced at uracils adjacent to G4 DNA, but activity is gradually restored as a function of nucleotide distance from the structure. Both UDG and hSMUG1 showed reduced activity at G4 DNA, indicating that the inhibition is not enzyme specific. In this thesis, I will present results supporting the model that G4 DNA promotes mutagenesis, in part, because it physically interferes with nearby uracil repair.

<center>**Methods**</center>

Circular Dichroism (CD) – Experimental and control DNAs were synthetic oligonucleotides synthesized by Operon (Huntsville, AL) and Integrated DNA Technologies (Coralville, IA) respectively (Table 1 and Table 2). Experimental DNAs consisted of guanine-rich oligonucleotides with single uracil substitutions at various locations surrounding guanines. Control DNAs consisted of experimental oligonucleotide sequence with thymine substitutions at guanines participating in G4 formation, with single uracil substitutions at locations identical to those in experimental sequences. Two additional positive controls were used consisting of the experimental sequence scrambled and thymine interrupted sequence scrambled. An Aviv model 215 CD spectrometer was used and spectra was taken in a 1 cm path quarts cell at 37 °C. DNAs for assays were prepared at a concentration of 50 µM in 10 mM Tris-HCl pH 7.6, 1 mM EDTA, and 100 mM KCl. Molar ellipticity was measured from 220 to 300 nm for three scans in 1 nm increments with a 1 sec. averaging time for each sample. Three scans were averaged to obtain final reading.

Uracil Cleavage Assay Sample Preparation – Experimental and control DNAs are oligonucleotide based and were tested for structure formation using CD (Table 1 and Table 2). Experimental (G4 containing) DNAs were folded in 950 mM NaCl and 50 mM KCl at a concentration of 500 mM. The folding reaction was placed in a BioRad MJ Mini Thermocycler (Hercules, CA) and incubated at 98 °C for 2 minutes, 60 °C for 90 minutes and 37 °C for 7 hours. Once folded, substrates were diluted to 13 µM. Substrates were 5' end labeled with 32P-γ-ATP purchased from Perkin Elmer (Waltham, MA) using T4 polynucleotide kinase purchased from New England Biolabs (NEB) (Ipswich, MA) and passed through illustra™ Microspin™ G-25 columns purchased from GE Healthcare (Wauwatosa, WI) to remove unincorporated label.

<center>16</center>

Thymine interrupted controls were 5' end labeled with ^{32}P-γ-ATP identical to the G4 oligonucleotides. Once labeled, thymine interrupted controls were annealed to their respective complementary oligonucleotide (Table 3) in a reaction containing five-fold excess of unlabeled compliment by heating to >95 °C in a water bath and allowing to drop to room temperature.

Uracil Cleavage Assays – All cleavage assays contain 260 fmol of DNA substrate in both hSMUG1 and UDG reactions. Purified hSMUG1 and UDG were purchased from NEB and assays followed manufacture recommended reaction conditions. The amount of enzyme added to cleavage assays was determined empirically by assaying the units required to attain over 80% cleavage product on control (unstructured) oligonucleotides (about 3.6 nM UDG and about 33.5 nM hSMUG1). Uracil excision was performed at 37 °C for 30 minutes. Cleavage of oligonucleotides was performed via the addition of NaOH to a final concentration of 333 mM and placed at 55 °C for 10 minutes. The reaction was stopped by adding a solution of 20 mM EDTA, 60 µg/mL Proteinase K, and 0.4% SDS incubated at 37 °C for 10 minutes followed by an equal volume of 90% formamide. Cleavage assays were resolved using a 7 M urea 15% PAGE run at 450V for 20 minutes-5.5 hours depending on the oligonucleotide used.

Image Collection and Data Analysis – Gels were exposed to a phosphor screen and images captured using a Typhoon FLA 7000 (GE Healthcare). Uracil cleavage assays were quantified using ImageQuant TL software (GE Healthcare) by dividing the pixel density of the resolved cleavage product (CP) band by the pixel density of the sum of the cleavage product and full length (FL) bands [CP/(CP+FL)]. All uracil cleavage assays were performed a minimum of four times.

Results

G4 DNA Formation. The model G4 DNA sequence used in this study was characterized

by Sen and Gilbert in 1990 and determined to be an intermolecular structure (Sen and Gilbert,

1990). A parallel multi-molecular structure forms from four separate oligonucleotides (Figure

5A) in the presence of sodium and potassium, although there is also a variable amount of anti-

parallel, bi-molecular structure (Figure 5B). Once folded, the structure is extremely stable and

does not require additional stabilizing salt, which is ideal for use in hSMUG1 uracil cleavage

assay reactions because this enzyme is salt sensitive (data not shown). Complex sequences

capable of forming intermolecular G4 DNA, like the one used here, form better in the presence

of sodium because potassium stabilizes structure intermediates and creates more complex

mixtures of G4 species (Sen and Gilbert, 1990). For this reason, and to ensure intermolecular G4

DNA formation, our sequence was folded in 950 mM NaCl and only 50 mM KCl.

The aim of this study is to establish the footprint of inhibition imparted by G4 DNA for

UDG and hSMUG1. To accomplish this, several synthetic oligonucleotides carrying a known G4

sequence motif were synthesized to include a single uracil substitution located 5' of a G4 tetrad

(5' #U), in between two G4s (GUG #), and 3' of G4 (3' #U) (Table 1). Control substrates (Figure

5C) were derived from the same model G4 DNA sequence, but guanine repeats were interrupted

with thymine (Table 2) and paired with complement (Table 3). These sequences are annotated as

GT DNA throughout. Two positive controls were used alongside GT DNA controls to assay

uracil excision on unstructured DNA. First, I constructed a scrambled sequence using the base

composition from the G4 DNA sequence, and second, using the previous scrambled sequence, I

developed an oligonucleotide with random guanines exchanged for thymines (SCR and SCR GT

respectively). Circular dichroism (CD) scans were performed on all sequences to ensure G4

18

DNA formation or the lack thereof (Figure 6). Even though the presence of uracil near G4 DNA does not affect structure formation, CD scans were used to ensure G4 DNA formation in these sequences (Holton and Larson, 2015). Our model G4 DNA sequences and controls displayed characteristic spectra showing a peak at about 265 nm in G4 substrates and about 278 nm in GT substrates indicating G4 DNA formation and SS DNA respectively (Balagurumoorthy, et al., 1992; Kypr, et al., 2009).

G4 Formation Creates an Inhibitory Footprint for UDG Activity. Recent data shows UDG cleavage activity is limited when the uracil lesion is located directly adjacent to G4 DNA, but not at lesions three nucleotides away (Holton and Larson, 2015). Here I have expanded this analysis, comparing the extent of UDG and hSMUG1 inhibition around a G4 DNA structure. Representative images of each denaturing PAGE for UDG cleavage assays at each uracil location are shown in Figure 7. Cleavage products (CP) will migrate faster than the full-length substrate (FL) and so there are two possible bands in each cleavage reaction. The percentage of DNA cleaved in a reaction which reflects enzyme cleavage activity, is calculated by dividing the pixel density of the cleavage product band by the pixel density of the sum of complete substrate and cleavage product bands [CP/(CP+FL)]. Quantitation of cleavage activity was attained from the average of four separate reactions for each uracil location, i.e., 5', in between, or 3' to G4 DNA (Figure 8).

UDG activity is modestly inhibited compared to the thymine interrupted and SCR positive controls when uracil is positioned directly 5' to G4 DNA (5'1 U) (69% compared to 91%; Figure 8A). For uracils positioned three bases away from G4 paired guanines, cleavage efficiency was identical to unstructured control (Figure 8A), similar to previous results (Holton and Larson, 2015). Activity directly 3' to G4 DNA is mildly inhibited compared to thymine

interrupted and SCR positive controls with an observed cleavage activity of 82% directly adjacent to G4 DNA, but when uracil is moved three bases away cleavage is essentially identical to unstructured controls (Figure 8C). Activity in between G4 DNA was consistently inhibited regardless of the nucleotide position compared to thymine interrupted and SCR positive controls, with observed cleavage activities of 66%, 72%, 72%, and 56% (GUG 1, GUG 2, GUG 3, and GUG 4 respectively) (Figure 8B). UDG inhibition was similar, but milder, compared to the study completed by Holton and Larson, who observed about 10% and 20% activity 5' and 3' to G4 DNA, respectively. However, those studies utilized different G4-forming oligonucleotides. Full activity was observed on SCR control and DS thymine interrupted controls with an observed average cleavage activity over 90%. No significant difference in enzyme activity was observed between SS and DS positive controls, SCR and SCR GT respectively.

hSMUG1 is More Sensitive to G4 DNA than UDG. The effect of G4 DNA on SMUG1 has not been investigated until this point. hSMUG1 assays were performed identically to UDG. Representative images of each denaturing PAGE for hSMUG1 cleavage assays are shown in Figure 9. hSMUG1 cleavage activity was also calculated identically to UDG. Quantification of cleavage activity was attained by averaging four separate reactions, and were separated by their uracil placement group i.e., 5', in between, or 3' to G4 DNA (Figure 10). Similar to UDG, hSMUG1 activity was inhibited directly 3' of G4 DNA compared to thymine interrupted and SCR positive controls with an observed cleavage activity of 54% (Figure 10C). At uracil lesions three bases away from G4, excision returned to levels indistinguishable from unstructured controls (Figure 10C). Activity at uracils located in between G4 tetrads was consistently inhibited compared to thymine interrupted and SCR positive controls regardless of uracil position, with observed cleavage activity of 32%, 32%, 38%, and 29% (GUG 1, GUG 2, GUG 3,

and GUG 4 respectively) (Figure 10B). Uracil excision activity directly 5' to G4 DNA was inhibited compared to thymine interrupted and SCR positive controls, with 37% cleaved compared to 87% for the thymine substituted and SCR substrates (Figure 10A). Interestingly, SMUG1 cleavage activity was reduced to 65% at uracil placed 5 bases away from G4 DNA, indicating that G4 DNA inhibits SMUG1 efficiency even at this nucleotide distance (Figure 10A). This drop in activity was much more dramatic for hSMUG1 compared to UDG, indicating that hSMUG1 is more sensitive to G4 DNA compared to UDG. Full activity was observed on DS thymine interrupted controls, with the exception of 3' 2U, though there was still an average cleavage activity of over 80% observed. No significant difference in enzyme activity was observed between SS and DS positive controls, SCR and SCR GT respectively.

Discussion

This work focused on establishing the impact of G4 DNA on bacterial UDG and human SMUG1 (hSMUG1), measuring precisely the nucleotide distance required to reduce the inhibitory effect of structure formation on uracil excision activity. G4 DNA is known to be associated with genomic instability (Cea, et al, et al., 2015; Maizels, 2015), so it is important to define the actual molecular sources of mutagenesis at sequences near those participating in structure formation. Previous work has shown that the activity of UDG and the human homolog hUNG2 is reduced directly adjacent 5' or 3' to G4 DNA (Holton and Larson, 2015). Here I expand that analysis to clarify the footprint of enzyme inhibition around G4 DNA. Consistent with previous work, UDG activity was reduced at uracil lesions located adjacent to G4 DNA and a greater inhibition 5' compared to 3' to G4 DNA was also observed (Figure 8A and 8C). However, the level of overall repair inhibition observed in this study was modest in comparison. Differences observed could be explained by altered reaction conditions and substrates used.

Specifically, the sequence used in this study leads to the formation of predominantly inter- and bi-molecular G4. Once folded, the structure is stable in glycosylase reaction conditions. In contrast, the DNAs used in experiments reported by Holton and Larson were a mixture of intramolecular and intermolecular structures and reactions contained either 50 mM or 100 mM potassium chloride or lithium chloride. Since UDG does not have apparent salt sensitivity, the difference in activity is most likely associated with the different structures. My results therefore imply that some G4s are more inhibitory to UDG activity than others. I could only test SMUG1 activity on one type of G4, because it is salt sensitive, so it is unknown if some G4 conformations are less inhibitory than others. However, results from my research suggest that uracil excision by UDG may be dependent on the specific G4 structure formed, with inter-molecular G4's less inhibitory than intra-molecular G4. In the cell, G4 is likely to fold within, rather than among, DNAs, and thus intramolecular G4 is probably more of a physiological conformation. This has not been established, but the synthesis of existing results on uracil excision activity at G4 DNAs suggest that UDG is reluctant to remove uracils next to guanines participating in G4 formation. The same pattern may also be present in humans since hUNG2 results mirrored UDG (Holton and Larson, 2016), but hUNG2 activity was not tested on intermolecular G4.

There are a few explanations for why G4 DNA inhibits UDG activity. First, G4 DNA may physically obstruct substrate recognition. UDGs have two distinctive motifs targeted to DNA binding, one 5' and one 3' of the uracil (Huffman, et al., 2005). If the 3'-specific motif requires more room than the 5'-specific motif it is possible that G4 DNA is a barrier preventing efficient UDG binding and activity. As a corollary, UDG damage recognition and *N*-glycosyl bond cleavage mechanisms may be compromised in the presence of G4 DNA. It is thought that

UDG searches for damage in a "hopping" manner where DNA binding is short lived, unless a uracil is identified and flipped into the active site for excision (Stivers, et al., 1999). UDG slightly bends DNA upon non-specific binding, but fully bends and kinks DNA when the uracil is flipped into the active site (Parikh, et al., 2000; Wibley, et al., 2003; and Huffman, et al., 2005). It is possible there is a slight asymmetry in this DNA bending, affecting lesions differently depending on whether they are 3' or 5' of G4 DNA. G4 DNA is a highly stable non-duplex structure, and it is possible that productive DNA bending altogether is inhibited because of the structure. Data of UDG activity in between two G4 DNAs supports this idea (Figure 8B). If G4 DNA served as a barrier to UDG, you would expect positions GUG 2 and GUG 3 to have similar cleavage activities to 5' 2U. Instead, all uracils positioned in between G4 DNA show a ~60% reduction in activity compared to unstructured SCR controls, regardless of position. Further work on UDG mechanics at G4 DNA is required to elucidate specific cause of UDG inhibition near G4 DNA. It would also be of interest to investigate this phenomenon with hUNG2 or with substrates consisting of; A., uracils placed between guanine tetrads that are further apart or, B., with substrates containing multiple intramolecular structures as opposed to multiple sets of guanine tetrads in an intermolecular conformation.

I have demonstrated the consequences of G4 formation on hSMUG1 cleavage activity. Compared to UDG, hSMUG1 is more sensitive to the presence of G4 DNA. hSMUG1 activity 3' to G4 DNA returns to levels comparable to unfolded DNA three bases away, and activity between G4 DNAs is modest (Figure 10B and Figure 10C). Reduced activity 3' to G4 DNA could be caused by a slight block in hSMUG1 DNA binding. Crystal structures of xSMUG1 show there is disruption of the adjacent base pair 5' to uracil by the wedge motif (Wibley, et al.,

2003). Using our model sequence, the base pair 5' to the uracil at position 3' 1U would be a guanine participating in G4 DNA formation.

Even though hSMUG1 is part of the same protein family as UDG it has less than 10% sequence identity (Fromme, et al., 2004; Wibley, et al., 2003). Additionally, the wedge motif is unique to SMUG1 and the mechanism of cleavage distinct is from UDG (Pettersen, et al., 2007). Given the differences between the two enzymes, it is not surprising that there is a large difference between activities 5' to G4 DNA, though the specific causes remain unclear. Since activity returns to maximum for uracils placed 3' to G4 DNA, is possible the wedge motif in hSMUG1 plays a large role 3' to uracil lesions and G4 DNA is acting as a barrier, preventing binding and cleavage. Alternatively, a significant portion of hSMUG1 could require the area 3' to uracil lesions and G4 DNA is again a barrier to binding and cleavage. It cannot be fully ruled out that hSMUG1 activity around G4 DNA is partially influenced by sequence specificity. I observed a slight variation in cleavage activity depending the base that flanks uracil, consistent with studies showing preference for 5' – GUT (Pettersen, et al., 2007). At uracil located 5' 4U, cleavage activity increases from 5' 3U and drops again at position 5' 5U (Figure 9A). Position 5'4U contains uracil within the sequence 5' – GUT, providing an explanation for the sudden increase in activity. Further studies are required in order to quantify how strong this preference is and whether it changes as a function to distance from G4 DNA.

Similar to UDG, hSMUG1 may possess a DNA scanning mechanism to search for and identify uracil lesions in DNA. While the uracil base flipping mechanism is conserved in these uracil glycosylases, little is known of the search mechanisms. It has been shown that UDG displays a hopping mechanism and recent data indicates a limited scanning ability on either strand of about 10 base pairs once bound however, the mechanism for SMUG1 lesion searching

remains unclear (Stivers, et al., 1999; Friedman and Stivers, 2010). A scanning mechanism with a 3'→5' directionality could account for the increased inhibition observed at uracils placed 5' to G4 DNA compared to uracils placed 3' (Figure 10A and Figure 10C). It would be of great interest to investigate lesion search mechanics of hSMUG1 in future studies, and longer DNA substrates may be required to identify the consequence of G4 DNA on lesion identification.

Regardless of the specific causes, this data indicates G4 DNA is an impediment to proper repair of uracil by UDG and hSMUG1 (Figure 11), with increased inhibition the closer a uracil is to G4 DNA (Figure 8 and Figure 10). These data have different implications depending on whether the cell is proliferating or not. UDG is mainly present during replication with low levels at other stages in the cell cycle (Kavli, et al., 2002; Visnes, et al., 2009). Presumably, G4 DNA has an opportunity to form during replication because the DNA is transiently denatured, also predisposing the DNA to cytosine deamination events or misincorporation of deoxyuracil by replicative polymerases (Lindahl, 1993; Krokan, et al., 2002; Bochman, et al., 2012; Maizels, 2006; Rhodes and Lipps, 2015). Should uracil appear near guanine repeats and G4 DNA, removal and repair is likely to be less efficient, potentially increasing mutagenesis at those sites. With over 300,000 G4-capable sites (Huppert and Balasubramanian, 2005; Todd, et al., 2005), mutations could quickly accumulate in a proliferating cell. G4 DNA also has an opportunity to form during transcription. Inhibition of hSMUG1 has potential repercussions for highly transcribed loci, since the presence of hSMUG1 inhibits cell growth and proliferation, limiting its involvement in replication (Pettersen, et al., 2007). SMUG1 is a much less processive enzyme compared to UDG, partially due the unique wedge motif and prolonged DNA binding (Wibley, et al., 2003; Pettersen, et al., 2007). My data argues cytosine deamination within guanine-rich and transcribed loci may be particularly prone to reduced repair, resulting in C→T transitions.

In conclusion, I have demonstrated that UDG was modestly sensitive to the presence of G4 DNA folded in the oligonucleotides tested, with the reductions in activity dependent upon proximity to G4 DNA. Likewise, hSMUG1 was more sensitive, and activity is inhibited largely in the proximity to G4 DNA, suggesting that uracil bases located near guanine-rich loci capable of forming G4 structures may be at risk of mutagenesis. Furthermore, my data indicated UDG and hSMUG1 had distinct inhibition zones implicating a greater accumulation of uracil in non-proliferating cells where hSMUG1 is most active. Inhibited uracil excision may partially account for the genomic instability observed at G4-capable loci.

References

An, Q., Robins, P., Lindahl, T., & Barnes, D. E. (2005). C→ T mutagenesis and γ-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *The EMBO journal*, *24*(12), 2205-2213.

Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., & Sasisekharan, V. (1992). Hairpin and parallel quartet structures for telomeric sequences. *Nucleic acids research*, *20*(15), 4061-4067.

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics*, *13*(11), 770-780.

Brenerman, B. M., Illuzzi, J. L., & Wilson III, D. M. (2014). Base excision repair capacity in informing healthspan. Carcinogenesis, 35(12), 2643-2652.

Cea, V., Cipolla, L., & Sabbioneda, S. (2015). Replication of Structured DNA and its implication in epigenetic stability. *Frontiers in genetics*, *6*.

Drohat, A. C., Jagadeesh, J., Ferguson, E., & Stivers, J. T. (1999). Role of electrophilic and general base catalysis in the mechanism of Escherichia coli uracil DNA glycosylase. *Biochemistry*, *38*(37), 11866-11875.

Friedman, J. I., & Stivers, J. T. (2010). Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry*, *49*(24), 4957-4967.

Fromme, J. C., Banerjee, A., & Verdine, G. L. (2004). DNA glycosylase recognition and catalysis. *Current opinion in structural biology*, *14*(1), 43-49.

Haug, T., Skorpen, F., Aas, P. A., Malm, V., Skjelbred, C., & Krokan, H. E. (1998). Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. *Nucleic acids research*, *26*(6), 1449-1457.

Holton, N. W., & Larson, E. D. (2015). G-quadruplex DNA structures can interfere with uracil glycosylase activity in vitro. *Mutagenesis*, *31*(4), 385-392.

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, *33*(9), 2908-2916.

Huffman, J. L., Sundheim, O., & Tainer, J. A. (2005). DNA base damage recognition and removal: new twists and grooves. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *577*(1), 55-76.

Imai, K., Slupphaug, G., Wen, L., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H. E., Ochs, H. D., Fischer, A., & Durandy, A. (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nature immunology*, *4*(10), 1023.

Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., & Slupphaug, G. (2002). hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *Journal of Biological Chemistry*, *277*(42), 39926-39936.

Krawczak, M., Ball, E. V., & Cooper, D. N. (1998). Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *The American Journal of Human Genetics*, *63*(2), 474-488.

Krokan, H. E., Drabløs, F., & Slupphaug, G. (2002). Uracil in DNA-occurrence, consequences and repair. *Oncogene*, *21*(58), 8935.

Kypr, J., Kejnovská, I., Renčiuk, D., & Vorlíčková, M. (2009). Circular dichroism and conformational polymorphism of DNA. *Nucleic acids research*, *37*(6), 1713-1725.

Larson, E. D., Bednarski, D. W., & Maizels, N. (2008). High-fidelity correction of genomic uracil by human mismatch repair activities. *BMC molecular biology*, 9(1), 94.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*(6422), 709-715.

Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B., & Sperens, B. D. N. A. (1977). DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *Journal of Biological Chemistry*, *252*(10), 3286-3294.

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature structural & molecular biology*, *13*(12), 1055.

Maizels, N. (2015). G4-associated human diseases. *EMBO reports*, e201540607.

Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., & Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil–DNA glycosylase. *The EMBO journal*, *20*(15), 4278-4286.

Olsen, L. C., Aasland, R., Wittwer, C. U., Krokan, H. E., & Helland, D. E. (1989). Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme. *The EMBO journal*, *8*(10), 3121.

Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., & Tainer, J. A. (2000). Uracil-DNA glycosylase–DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proceedings of the National Academy of Sciences*, *97*(10), 5083-5088.

Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, *418*(6893), 99.

Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E., & Kavli, B. (2007). Uracil–DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic acids research*, *35*(12), 3879-3892.

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, *43*(18), 8627-8637.

Sen, D., & Gilbert, W. (1990). A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, *344*(6265), 410-414.

Stivers, J. T., Pankiewicz, K. W., & Watanabe, K. A. (1999). Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry*, *38*(3), 952-963.

Sung, J. S., & Demple, B. (2006). Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA. *The FEBS journal*, *273*(8), 1620-1629.

Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research*, *33*(9), 2901-2907.

Visnes, T., Doseth, B., Pettersen, H. S., Hagen, L., Sousa, M. M., Akbari, M., Otterlei, M., Kavli, B., Slupphaug, G., & Krokan, H. E. (2009). Uracil in DNA and its processing by different DNA glycosylases. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1517), 563-568.

Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L., & Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Molecular cell*, *11*(6), 1647-1659.
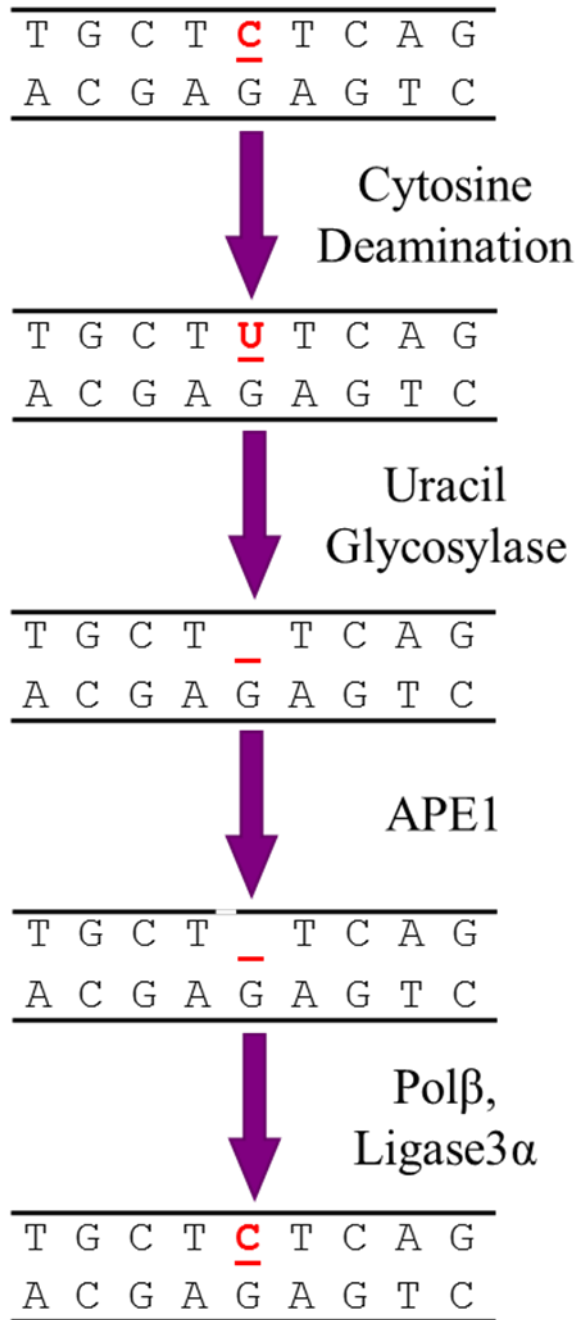
Figure 3. Short Patch Base Excision Repair Pathway. Non-bulky DNA lesions like uracil are repaired by BER. A lesion specific glycosylase, in this case a uracil glycosylase, will cleave the base, leaving an abasic site. APE1 will cleave the phosphodiester bond, allowing DNA polβ to resynthesize the DNA, and ligation by ligase3α completes repair.
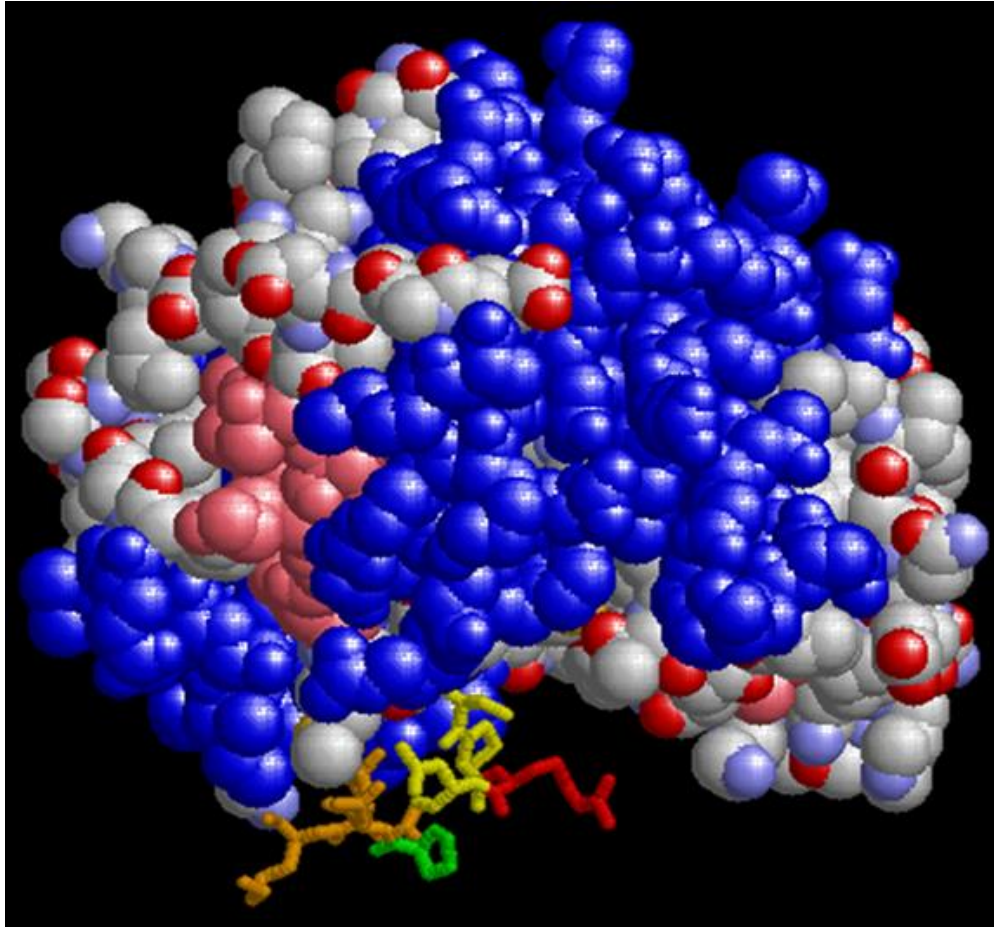
Figure 4. Crystal Structure of xSMUG1. SMUG1 from *Xenopus laevis,* which shares 66% sequence identity with human SMUG1, was crystalized and the structure solved. The unique SMUG1 wedge motif is shown as a ball-and-stick form consisting of a loop (yellow) and small α-helix (orange). This motif is thought to facilitate tight binding at AP sites by inserting itself into the minor groove. An arginine (red) in the α-helix replaces uracil after it is flipped into the active site, further stabilizing SMUG1 binding. A proline (green) in the loop stabilizes guanines on the complimentary strand indicating a large role in non-proliferating cells (1OE6).

Table 1.

Cleavage Assay, Uracil-G4 Oligonucleotide Sequences. Guanines participating in G-quadruplex are shown in bold and italicized. The location of uracil bases is shown in bold and double underlined.

Name:	Sequence (5' – 3'):
SCR	ATAGAGGCGACGGCGGTGAGAATAGAGTG<u>U</u>CGGCGAGTGATAGTCGCGG
5' 5U	TGGACCAGACCTAGCA<u>U</u>CTAT**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
5' 4U	TGGACCAGACCTAGCAG<u>U</u>TAT**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
5' 3U	TGGACCAGACCTAGCAGC<u>U</u>AT**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
5' 2U	TGGACCAGACCTAGCAGCT<u>U</u>T**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
5' 1U	TGGACCAGACCTAGCAGCTA<u>U</u>**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
GUG 1	TGGACCAGACCTAGCAGCTAT**GGGGG**<u>U</u>GCT**GGGG**AAGGT**GGG**AATGTGA
GUG 2	TGGACCAGACCTAGCAGCTAT**GGGGG**A<u>U</u>CT**GGGG**AAGGT**GGG**AATGTGA
GUG 3	TGGACCAGACCTAGCAGCTAT**GGGGG**AG<u>U</u>T**GGGG**AAGGT**GGG**AATGTGA
GUG 4	TGGACCAGACCTAGCAGCTAT**GGGGG**AGC<u>U</u>**GGGG**AAGGT**GGG**AATGTGA
3' 1U	TGGACCAGACCTAGCAGCTAT**GGGGG**AGCT**GGGG**AAGGT**GGG**<u>U</u>ATGTGA
3' 2U	TGGACCAGACCTAGCAGCTAT**GGGGG**AGCT**GGGG**AAGGT**GGG**A<u>U</u>TGTGA
3' 3U	TGGACCAGACCTAGCAGCTAT**GGGGG**AGCT**GGGG**AAGGT**GGG**AA<u>U</u>GTGA

Table 2.

Cleavage Assay, Thymine Substituted Oligonucleotide Sequences. The location of uracil bases is shown in bold and double underlined. The location of previous G-quadruplex sites from uracil-G4 sequences is shown in bold and thymine bases interrupting structure formation are italicized.

Name:	Sequence (5' – 3'):
SCR (GT)	ATAGA**_TT_**CGACG**_T_**CG**_T_**TGAGAATAGAGTG**U**C**_TT_**CGAGTGATAGTCGCGG
5' 5U (GT)	TGGACCAGACCTAGCA**U**CTAT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
5' 4U (GT)	TGGACCAGACCTAGCAG**U**TAT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
5' 3U (GT)	TGGACCAGACCTAGCAGC**U**AT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
5' 2U (GT)	TGGACCAGACCTAGCAGCT**U**T**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
5' 1U (GT)	TGGACCAGACCTAGCAGCTA**U**G**_TTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
GUG 1 (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**G**U**GCT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
GUG 2 (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GA**U**CT**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
GUG 3 (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GAG**U**T**_GTT_**GAGGT**_G_**T**_G_**AATGTGA
GUG 4 (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GAGC**U**G**_TT_**GAGGT**_G_**T**_G_**AATGTGA
3' 1U (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**G**U**ATGTGA
3' 2U (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**GA**U**TGTGA
3' 3U (GT)	TGGACCAGACCTAGCAGCTAT**_GTTT_**GAGCT**_GTT_**GAGGT**_G_**T**_G_**GAA**U**GTGA

34

Table 3.

Cleavage Assay, Complement Oligonucleotide Sequences. Reverse complimentary sequences of
thymine substituted oligonucleotides, allowing the production of double stranded control
substrates. The location of G-quadruplex in parent (uracil-G4) molecules, is shown in bold and
italicized for comparison. The adenosine compliment to uracil is shown bolded and underlined.

Name:	Sequence (5' – 3'):
SCR RC(GT)	CCGCGACTATCACTCGAAG**A**CACTCTATTCTCAACGACGTCGAATCTAT
5' 5U RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGCT**CAAAC**ATAG**A**TGCTAGGTCTGGTCCA
5' 4U RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGCT**CAAAC**ATA**A**CTGCTAGGTCTGGTCCA
5' 3U RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGCT**CAAAC**AT**A**GCTGCTAGGTCTGGTCCA
5' 2U RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGCT**CAAAC**A**A**AGCTGCTAGGTCTGGTCCA
5' 1U RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGCT**CAAAC**A**TAGCTGCTAGGTCTGGTCCA
GUG 1 RC(GT)	TCACATT**CAC**ACCTT**CAAC**AGC**A**CAAACATAGCTGCTAGGTCTGGTCCA
GUG 2 RC(GT)	TCACATT**CAC**ACCTT**CAAC**AG**A**TCAAACATAGCTGCTAGGTCTGGTCCA
GUG 3 RC(GT)	TCACATT**CAC**ACCTT**CAAC**A**A**CTCAAACATAGCTGCTAGGTCTGGTCCA
GUG 4 RC(GT)	TCACATT**CAC**ACCTT**CAAC**A**A**GCTCAAACATAGCTGCTAGGTCTGGTCCA
3' 1U RC(GT)	TCACAT**A**CACACCTT**CAAC**AGCT**CAAAC**ATAGCTGCTAGGTCTGGTCCA
3' 2U RC(GT)	TCACA**A**TCACACCTT**CAAC**AGCT**CAAAC**ATAGCTGCTAGGTCTGGTCCA
3' 3U RC(GT)	TCAC**A**TT**CAC**ACCTT**CAAC**AGCT**CAAAC**ATAGCTGCTAGGTCTGGTCCA

A. 5′...CA-**G-C-T-A-T**-G-G-G-G-G-**A-G-C-T**-G-G-G-G-A-A-G-G-T-G-G-G-**A-A-T**-GTGA 3′

B.

3′ AGTGTAAGGGTGGAA   GTATCGAC 5′
5′ ..CA**GCTAT**G   AAGGTGGG**AAT**GTGA 3′

C. 5′-TGGACCAGACCTAGCA**GCTAT**GTTTG**AGCT**GTTGAAGGTGTG**AAT**GTGA
3′-ACCTGGTCTGGATCGTCGATACAAACTCGACAACTTCCACACTTACACT
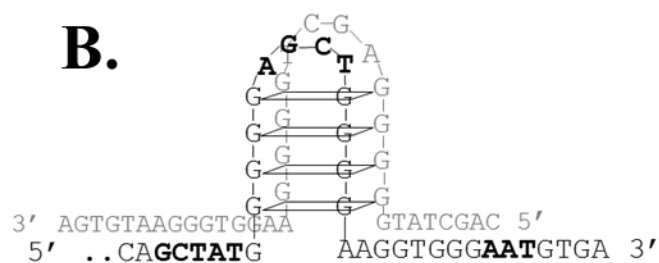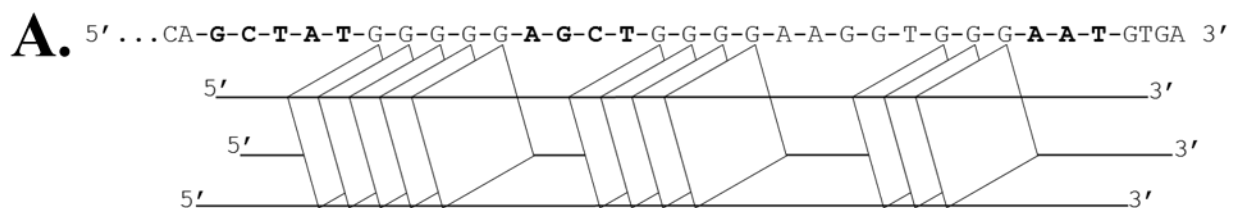
Figure 5. G4 DNA Model. (A) Intermolecular G-quadruplex structure composed of four
separate oligonucleotides in parallel orientation. (B) A secondary bi-molecular G-quadruplex is
formed, but is present as a minor product. (C) Double-stranded control substrates consisting of
G4-capable sequence with guanine runs interrupted with thymines. Bolded bases show the
positions where uracil is substituted in individual oligonucleotides.
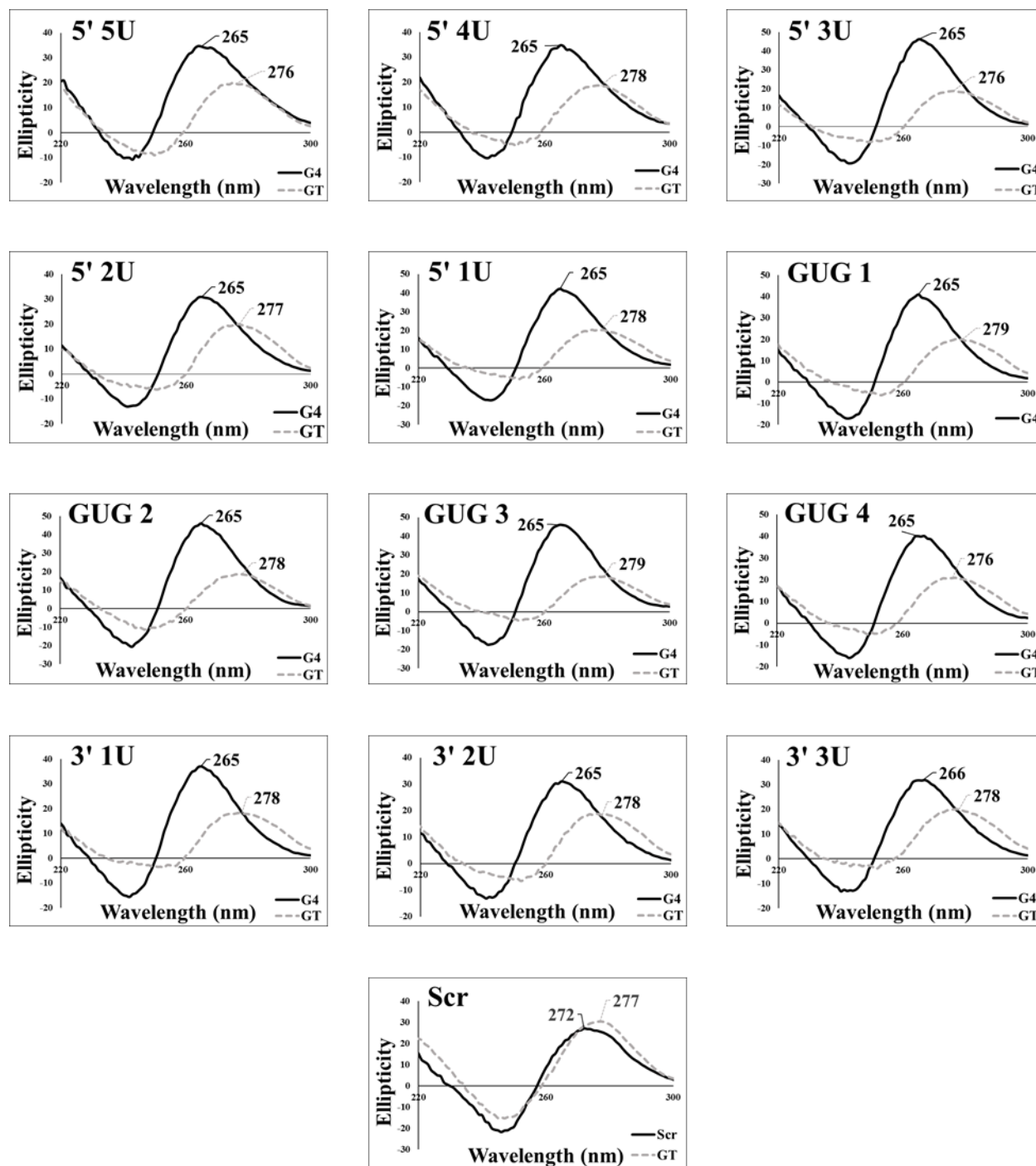
Figure 6. Circular Dichroism Analysis of Cleavage Assay Substrates. A CD scan is shown for every oligonucleotide containing uracil. Ellipticity is on the y-axis and wavelength (nm) is on the x-axis. G4 DNA oligonucleotides are shown in solid black while thymine interrupted control oligonucleotides are shown in dotted grey.

Figure 7. UDG Cleavage Assays. A representative denaturing PAGE is shown for the resolved products after UDG cleavage for uracil position. Four total UDG reactions were completed for each substrate. The location of the uracil relative to the G tetrads is shown on top. Lanes alternate between G4 DNA substrates and GT control substrates. Negative controls (no enzyme) are annotated with (-) and UDG supplemented reactions annotated with (+).
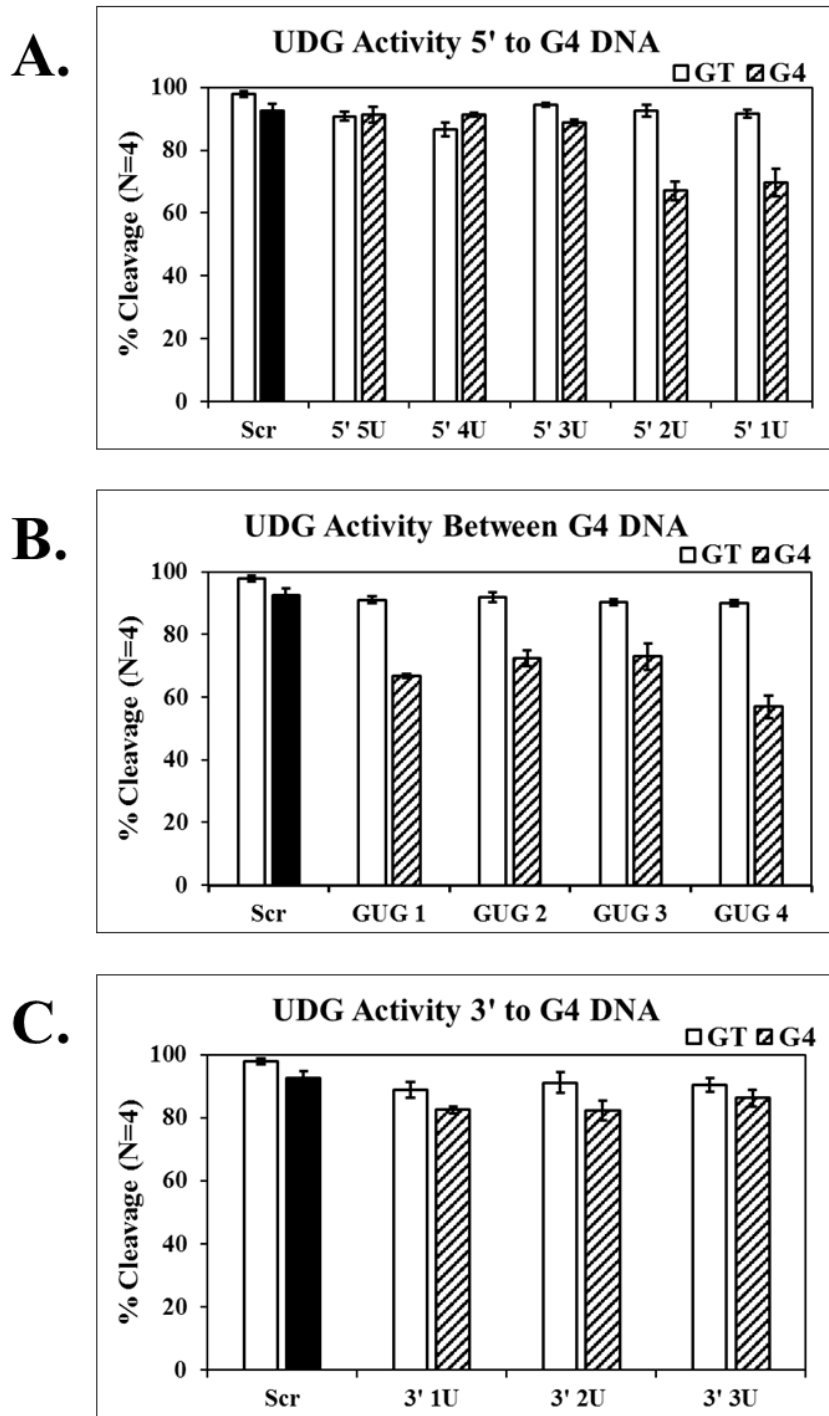
Figure 8. Quantification of UDG Cleavage Assays. Average cleavage of four reactions for each uracil position are shown as a percentage, with standard deviation. The scrambled control is shown in black and its GT counterpart on each graph. G4 DNA reactions are stripped and all GT reactions are white.
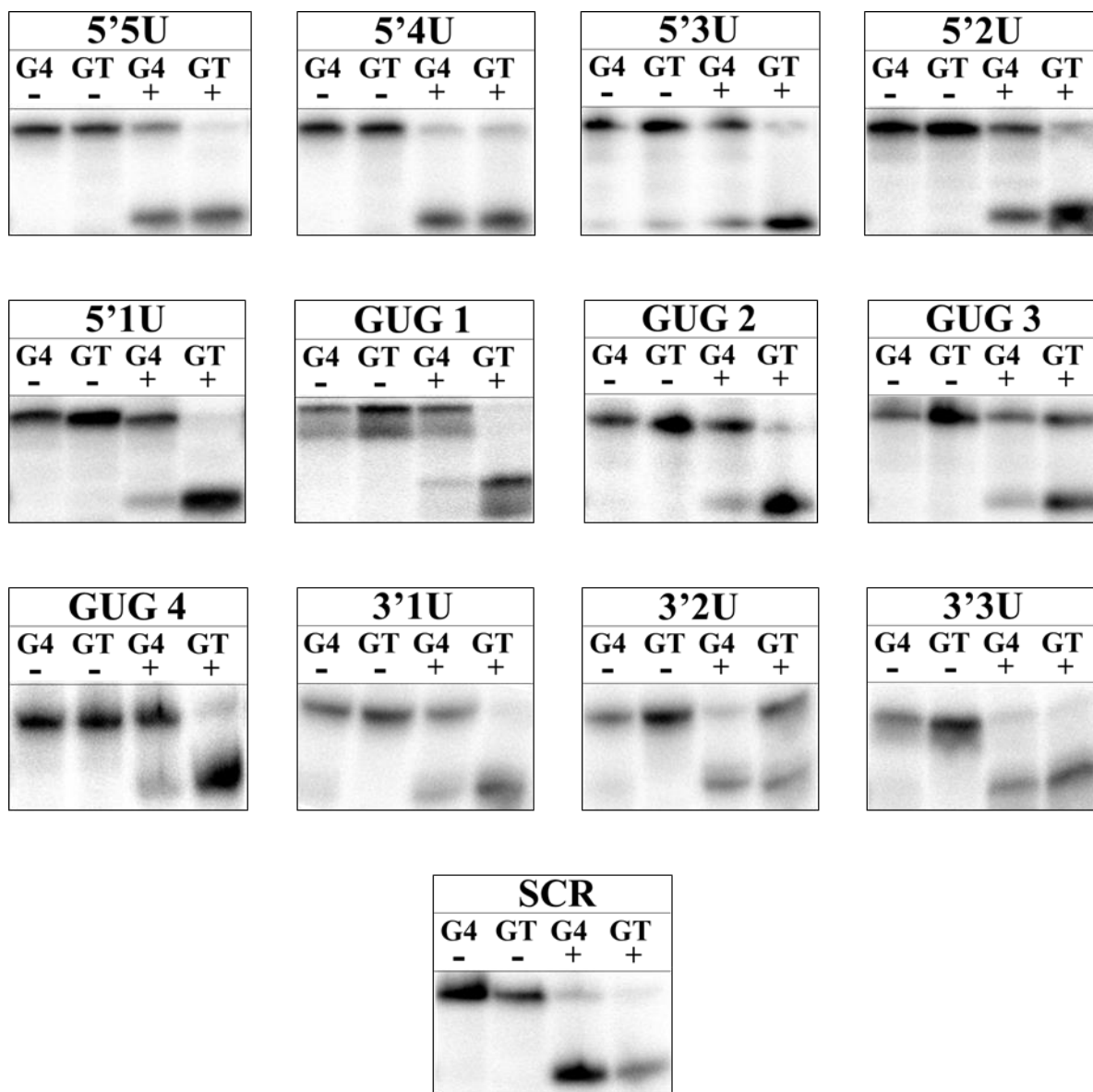
Figure 9. hSMUG1 Cleavage Assays. A representative denaturing PAGE is shown for the resolved products after hSMUG1 cleavage for uracil position. Location name is shown on top. Lane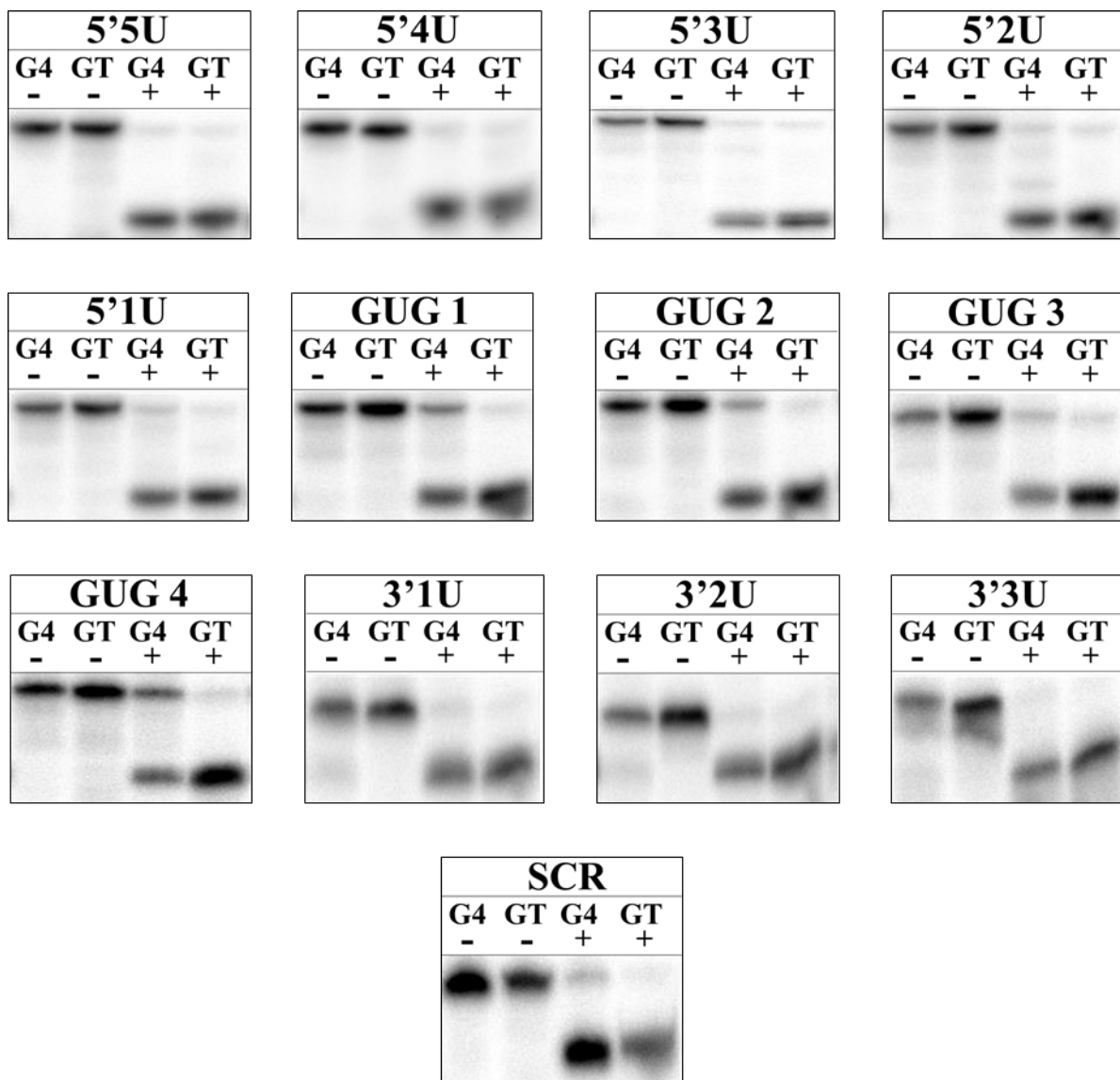s alternate between G4 DNA substrates and GT control substrates. Negative controls (no enzyme) are annotated with (-) and hSMUG1 supplemented annotated with (+).
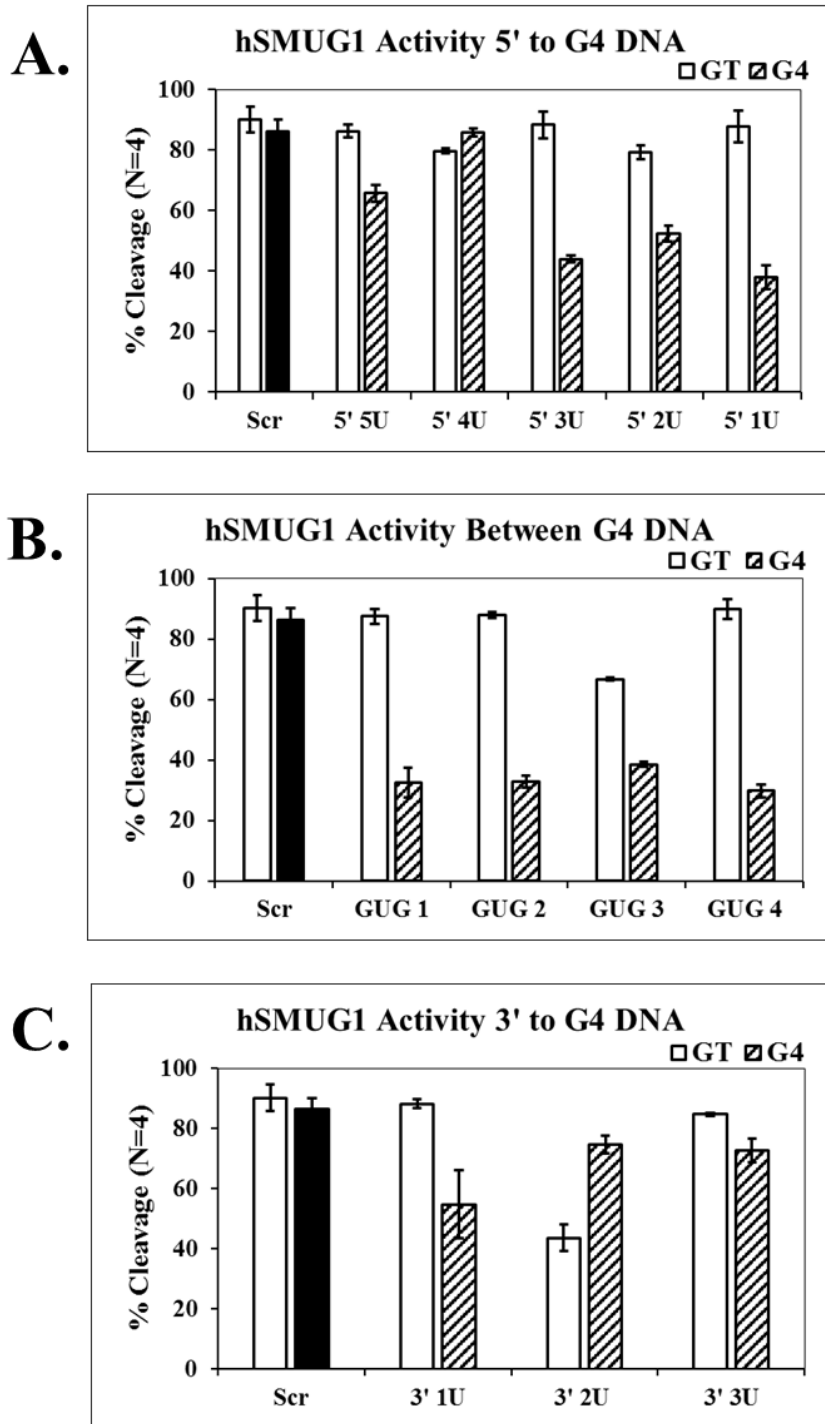
Figure 10. Quantification of hSMUG1 Cleavage Assays. The average cleavage of four reactions for each uracil position are annotated in percentage with standard deviation. The scrambled control is shown in black and its GT counterpart on each graph. G4 DNA reactions are shown stripped and all GT reactions are white.
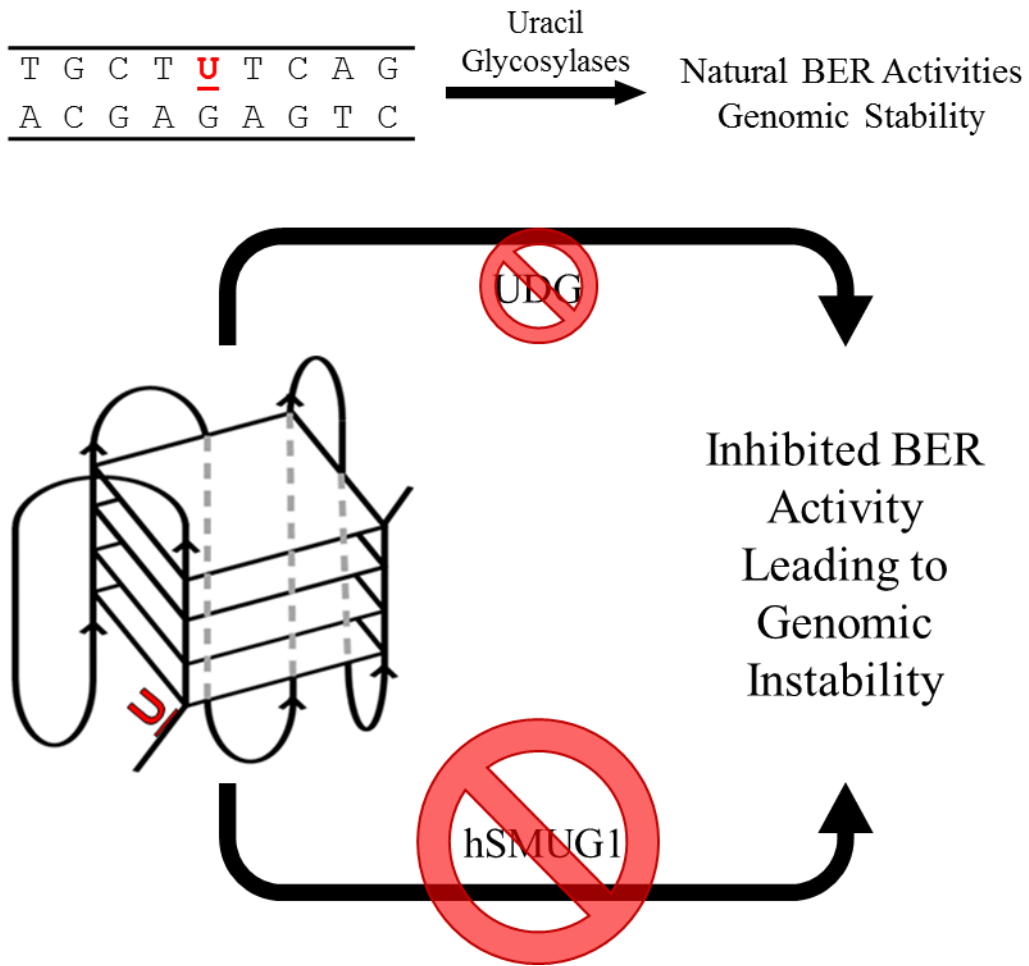
Figure 11. Model of G4 DNA Induced Genome Instability Concerning BER. Under B-form DNA conditions, uracil glycosylases retain full function. However, when G4 DNA is present there is inhibited activity by UDG and hSMUG1 leading to genomic instability with a greater effect on hSMUG1 than UDG.

CHAPTER III: EXONUCLEASE ACTIVITY AT G4

Introduction

Nucleases are a class of enzymes that either cleave DNA polymers internally (endonuclease) or excise nucleotides from one end (exonuclease). These activities are essential for all excision repair pathways, and repair of DNA damage ultimately requires the generation of a free 3' OH for repair polymerases (Mason and Cox, 2012). While a wide range of nuclease activities facilitate DNA repair in the cell, here I focused on characterizing activity of *E. coli* exonuclease III and exonuclease I at G-quadruplex structures.

Exonuclease III and Exonuclease I

Exonuclease III (ExoIII) is a Mg^{++}- dependent enzyme that excises nucleotides with 3'→5' directionality (Lovett, 2011). ExoIII disassociates frequently, leading to low processivity (Thomas and Olivera, 1978). While mainly an exonuclease, ExoIII also has endonuclease capabilities at abasic sites, cleaving the backbone 5' to the lesion. In *E. coli*, ExoIII is the major abasic endonuclease accounting for 80% of activity, indicating it is essential in base excision repair (Figure 3) (Lovett, 2011; Weiss, 1976; Weiss, 1981). ExoIII is also able to excise various damaged bases due to radiation or oxidation and is able to degrade RNA in DNA:RNA hybrid molecules, further implicating its importance in DNA damage repair and replication (Lovett, 2011; Weiss, 1981).

Exonuclease I (ExoI) is also a 3'→5' exonuclease dependent on Mg^{++}, however there is a strong specificity for single stranded DNA (Brody, et al., 1986; Lovett, 2011). ExoI is a highly processive enzyme with the ability to degrade 10,000 nucleotides/minute (Lovett, 2011; Thomas and Olivera, 1978). It interacts with single strand binding protein, SSB, in *E. coli* as part of a large network of replication and DNA repair proteins and is also known to be part of several

recombination pathways (Butland, et al., 2005; Friedman-Ohana and Cohen, 1998; Thoms, et al., 2008; Thoms and Wackernagel, 1998; Viswanathan, et al., 2001). Disruption of ExoI, leads to an increase in genetic rearrangements and mutation frequency, possibly due to its role in promoting mismatch repair (Viswanathan and Lovett, 1998; Burdett, et al., 2001; Cooper, et al., 1993; Viswanathan, et al., 2001).

Exonucleases and G4 DNA

Exonucleases have important roles in replication and DNA repair, instances where G4 DNA has the opportunity to form (Mason and Cox, 20012; Bochman, et al., 2012; Maizels, 2006; Rhodes and Lipps, 2015). Since G4 DNA formation correlates with sites of genome instability, the impact that the structure has on repair activities is important for elucidating mechanisms of mutagenesis (Maizels, 2015; Bochman, et al., 2012; Rhodes and Lipps, 2015; Vasquez and Wang, 2013). Exonuclease activities are required for excision repair. Here I show that both ExoIII and ExoI are unable to excise oligonucleotides contain G4 structures compared to identical sequences that to not contain the structure. My data argues that the genomic instability present in guanine-rich loci could partially be due to inhibited excision repair.

<div align="center">

Methods

</div>

Circular Dichroism – DNAs used for excision assays were oligonucleotides synthesized by Integrated DNA technologies (Coralville, IA) and page purified (Table 4). An Aviv model 215 CD spectrometer was used and spectra was taken in a 1 cm path quarts cell at 37 °C. Substrates were prepared at a concentration of 12 µM of in 10 mM Tris-HCl pH 7.6, 1 mM EDTA, and 100 mM KCl. Molar ellipticity was measured from 220 to 300 nm for three scans in 1 nm increments with a 1 s averaging time for each sample. Three scans were averaged to obtain final reading.

Sample Preparation – Oligonucleotides verified by CD scans were used in excision assays (Table 4). Each oligonucleotide was 5' end labeled with ^{32}P-γ-ATP, purchased from Perkin Elmer (Waltham, MA), using T4 polynucleotide kinase, from New England Biolabs (NEB) (Ipswich, MA). Unincorporated label was removed by column chromatography with illustra™ Microspin™ G-25 columns, purchased from GE Healthcare (Wauwatosa, WI).

Exonuclease III Assays – Exonuclease III was purchased from New England Biolabs (NEB) (Ipswich, MA). All exonuclease III reactions contained 20 nM of substrate and were placed in a hot water bath with either 100 mM KCl or 100 mM LiCl and allowed to cool to room temperature. Exonuclease III time range assays contained 30 units of exonuclease III and followed manufacturer recommended reaction conditions. Reactions were placed at 37 °C for 0, 10, 20, 30, 40, 50, or 90 minutes. Reactions were stopped at 37 °C for 10 minutes with 20 mM EDTA, 60 µg/mL Proteinase K, and 0.4% SDS. All exonuclease III assays were resolved with a 7 M urea 12% PAGE run at 400V for 45 minutes. Gels were exposed to phosphor screens and images captured using a Typhoon FLA 7000 (GE Healthcare).

Exonuclease I Assays – Exonuclease I was purchased from NEB. All exonuclease I reactions contained 20 nM of substrate and were placed in a hot water bath with either 100 mM KCl or 100 mM LiCl and allowed to cool to room temperature. Exonuclease I time range assays contained 2 units of exonuclease I and followed manufacturer recommended reaction conditions. Reactions were placed at 37 °C for 0, 2, 4, 6, 8, 10, or 20 minutes. Reactions were stopped with 20 mM EDTA, 60 µg/mL Proteinase K, and 0.4% SDS at 37 °C for 10 minutes. All exonuclease I assays were resolved with a 7 M urea 12% PAGE run at 400V for 45 minutes. Gels were exposed to a phosphor screens and image captured using a Typhoon FLA 7000 (GE Healthcare).

Results and Discussion

G4 DNA Formation. The G4-forming sequence I used here (5' –

GGGTGGGTTGGGTGGGG) has been previously characterized in *N. gonorrhoeae,* and G4

formation is required for proper recombination (Cahoon and Seifert, 2009). The specific G4

conformation was verified by circular dichroism (CD), and characterized (Cahoon and Seifert,

2009). I used this same 17 base pair sequence (Table 4) and further verified structure formation

by CD spectroscopy (Figure 12). As expected, the guanine-rich, G4-capable oligonucleotide

(G4) displayed a peak of ellipticity at 265 nm, a spectrum characteristic of G-quadruplex

formation. Controls consisted of a DNA complimentary to G4, (called G4 Com.) and a DNA

sequence similar to the G4 DNA, but with thymines substituted to disrupt G4 formation (called

GT). GT displayed a peak of ellipticity at 275 nm and G4 Com. displayed a peak of ellipticity at

279 nm, which indicates single stranded DNA (Balagurumoorthy, et al., 1992; Kypr, et al.,

2009).

Exonuclease III Activity. The G4 structure completely inhibited ExoIII activity (Figure

13). All substrates were in the presence of either potassium, favoring G4 DNA formation, or

lithium, which is less permissive to G4 DNA formation. While ExoIII easily excised past G4 in

reactions containing lithium, it was unable to in the presence of potassium, even after 90

minutes, indicating that G4 DNA inhibited activity (Figure 13A). Complete nucleotide excision

was observed for the G4 complimentary sequence and GT sequence, and full activity was

observed in both Li^+ and K^+ ionic conditions (Figure 13B and Figure 13C). This strongly

suggests that G4 DNA blocks ExoIII activity. The presence of various sized reaction products

reflects the dissociative properties of ExoIII (Thomas and Olivera, 1978). ExoIII participates in

base excision repair along with other mechanisms of DNA repair (Lovett, 2011; Weiss, 1976;

Weiss, 1981). Because of that, inhibition of exonuclease activity could exacerbate instability at guanine-rich loci.

Exonuclease I Activity. Similar to ExoIII, ExoI was reluctant to excise through G4 DNA. Complete nucleotide excision was observed almost instantly past G4 in the presence of lithium however, but not potassium even after 20 minutes (Figure 14A). Identical to ExoIII assays, DNAs that were unstructured, or unable to form G4, were completely processed by ExoI. Activity was insensitive to K+ and Li+, indicating that the reduction in activity observed for G4 (Figure 14A) is not due to the ion present, but rather due to the G4 structure (Figure 14B and Figure 14C). Almost instantaneous digestion of single stranded, or unstructured DNA, is indicative of ExoI efficiency (Lovett, 2011; Thomas and Olivera, 1978). ExoI is part of a large complex of proteins involved in replication, DNA repair and recombination, and my results indicate that impaired recombination and DNA repair at G4 DNA due to a block in ExoI activity, possibly accounting for at least a fraction of mutagenesis and genome observed at G4-capable sites.

In conclusion, I have found that both *E. coli* ExoIII and ExoI activity is reduced on DNAs containing G4 structure, and in conditions that support the formation of G4. Excision activity increases when the ion is changed to Li$^+$, (Figure 13 and Figure 14), and so it is not the sequence that is inhibitory, rather the formation of G4 inhibits exonuclease activity. Furthermore, the presence of Li$^+$ is not inhibitory, since full activity was observed for all DNAs tested (Figures). Since nucleases are required for all three excision repair pathways, this data provides another rationale for the higher levels of instability observed at guanine-rich loci. Failed DNA excision at G4 DNA may produces unresolved DNA breaks that promote recombination or mutagenesis.

References

Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., & Sasisekharan, V. (1992). Hairpin and parallel quartet structures for telomeric sequences. *Nucleic acids research*, *20*(15), 4061-4067.

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics*, *13*(11), 770-780.

Brody, R. S., Doherty, K. G., & Zimmerman, P. D. (1986). Processivity and kinetics of the reaction of exonuclease I from Escherichia coli with polydeoxyribonucleotides. *Journal of Biological Chemistry*, *261*(16), 7136-7143.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T., & Modrich, P. (2001). In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences*, *98*(12), 6765-6770.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., & Emili, J. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, *433*(7025), 531-537.

Cahoon, L. A., & Seifert, H. S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science*, *325*(5941), 764-767.

Cooper, D. L., Lahue, R. S., & Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry*, *268*(16), 11823-11829.

Friedman-Ohana, R., & Cohen, A. (1998). Heteroduplex joint formation in Escherichia coli recombination is initiated by pairing of a 3′-ending strand. *Proceedings of the National Academy of Sciences*, *95*(12), 6909-6914.

Genschel, J., Bazemore, L. R., & Modrich, P. (2002). Human exonuclease I is required for 5′ and 3′ mismatch repair. *Journal of Biological Chemistry*, *277*(15), 13302-13311.

Kypr, J., Kejnovská, I., Renčiuk, D., & Vorlíčková, M. (2009). Circular dichroism and conformational polymorphism of DNA. *Nucleic acids research*, *37*(6), 1713-1725.

Lovett, S. T. (2011). The DNA exonucleases of escherichia coli. *EcoSal Plus*, *4*(2).

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature structural & molecular biology*, *13*(12), 1055.

Maizels, N. (2015). G4-associated human diseases. *EMBO reports*, e201540607.

Mason, P. A., & Cox, L. S. (2012). The role of DNA exonucleases in protecting genome stability and their impact on ageing. *Age*, *34*(6), 1317-1340.

Shao, H., Baitinger, C., Soderblom, E. J., Burdett, V., & Modrich, P. (2014). Hydrolytic function of Exo1 in mammalian mismatch repair. *Nucleic acids research*, *42*(11), 7104-7112.

Thomas, K. R., & Olivera, B. M. (1978). Processivity of DNA exonucleases. *Journal of Biological Chemistry*, *253*(2), 424-429.

Thoms, B., Borchers, I., & Wackernagel, W. (2008). Effects of single-strand DNases ExoI, RecJ, ExoVII, and SbcCD on homologous recombination of recBCD+ strains of Escherichia coli and roles of SbcB15 and XonA2 ExoI mutant enzymes. *Journal of bacteriology*, *190*(1), 179-192.

Thoms, B., & Wackernagel, W. (1998). Interaction of RecBCD enzyme with DNA at double-strand breaks produced in UV-irradiated Escherichia coli: requirement for DNA end processing. *Journal of bacteriology*, *180*(21), 5639-5645.

Vasquez, K. M., & Wang, G. (2013). The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *743*, 118-131.

Viswanathan, M., & Lovett, S. T. (1998). Single-strand DNA-specific exonucleases in Escherichia coli: roles in repair and mutation avoidance. Genetics, 149(1), 7-16.

Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P., & Lovett, S. T. (2001). Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair. *Journal of Biological Chemistry*, *276*(33), 31053-31058.

Weiss, B. (1976). Endonuclease II of Escherichia coli is exonuclease III. *Journal of Biological Chemistry*, *251*(7), 1896-1901.

Weiss, B. (1981). 12 Exodeoxyribonucleases of Escherichia coli. *The enzymes*, *14*, 203-231.

Table 4.

Exonuclease Assay Sequences. Guanines participating in G-quadruplex are shown in bold.

Cytosines complimentary to G4 guanines are shown in bold and italicized. Thymines

interrupting G4 guanines are underlined.

Name:	Sequence (5' – 3'):
G4	GTATCTCACTTCTCT**GGG**T**GGG**TT**GGG**T**GGGG**ATCATTACGCTATTCTT
G4 Com.	AAGAATAGCGTAATGAT***CCCC*A*CCC*AA*CCC*A*CCC***AGAGAAGTGAGATAC
GT	GTATCTCACTTCTCT**GT̲GTGT̲GTTGT̲GTGT̲T̲G**ATCATTACGCTATTCTT
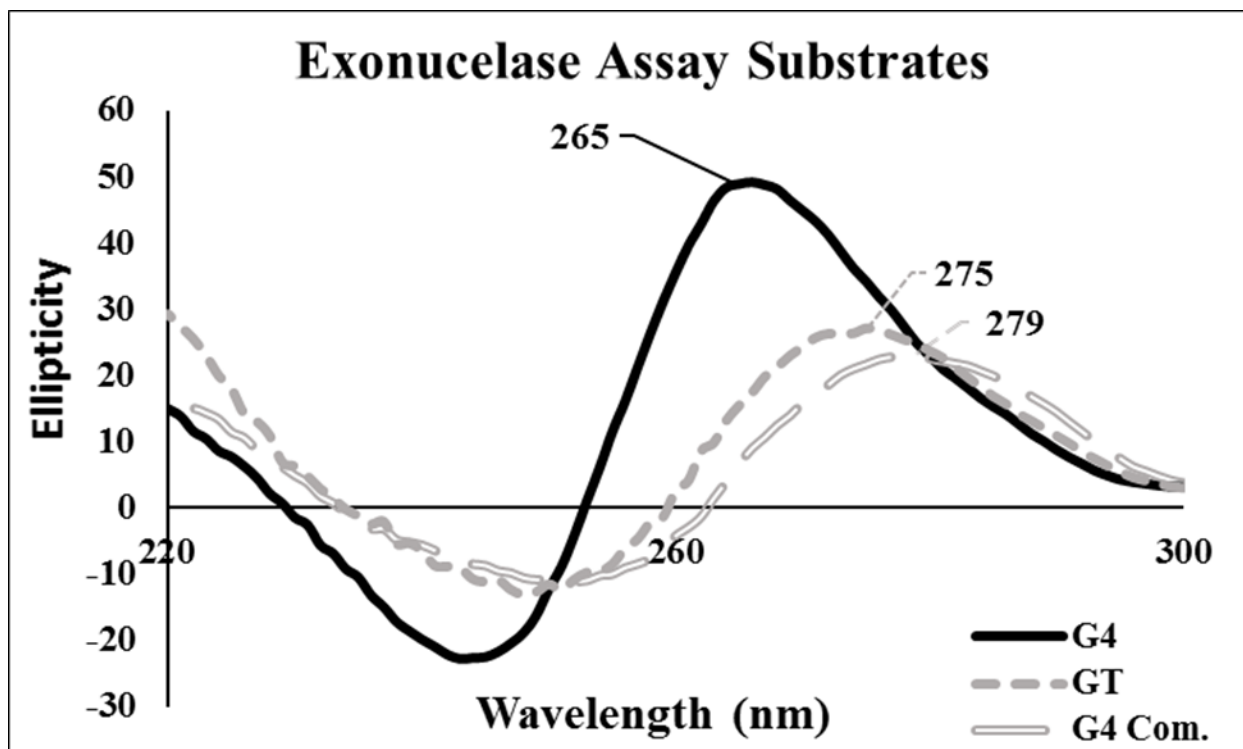
Figure 12. Circular Dichroism Analysis of Oligonucleotides used in Exonuclease Assays. A CD scan is shown for every oligonucleotide containing uracil. Ellipticity is on the y-axis and wavelength (nm) is on the x-axis. G4 DNA oligonucleotides are shown in solid black while thymine interrupted control oligonucleotides are shown in dotted grey and G4 compliment is shown in dotted outline grey.
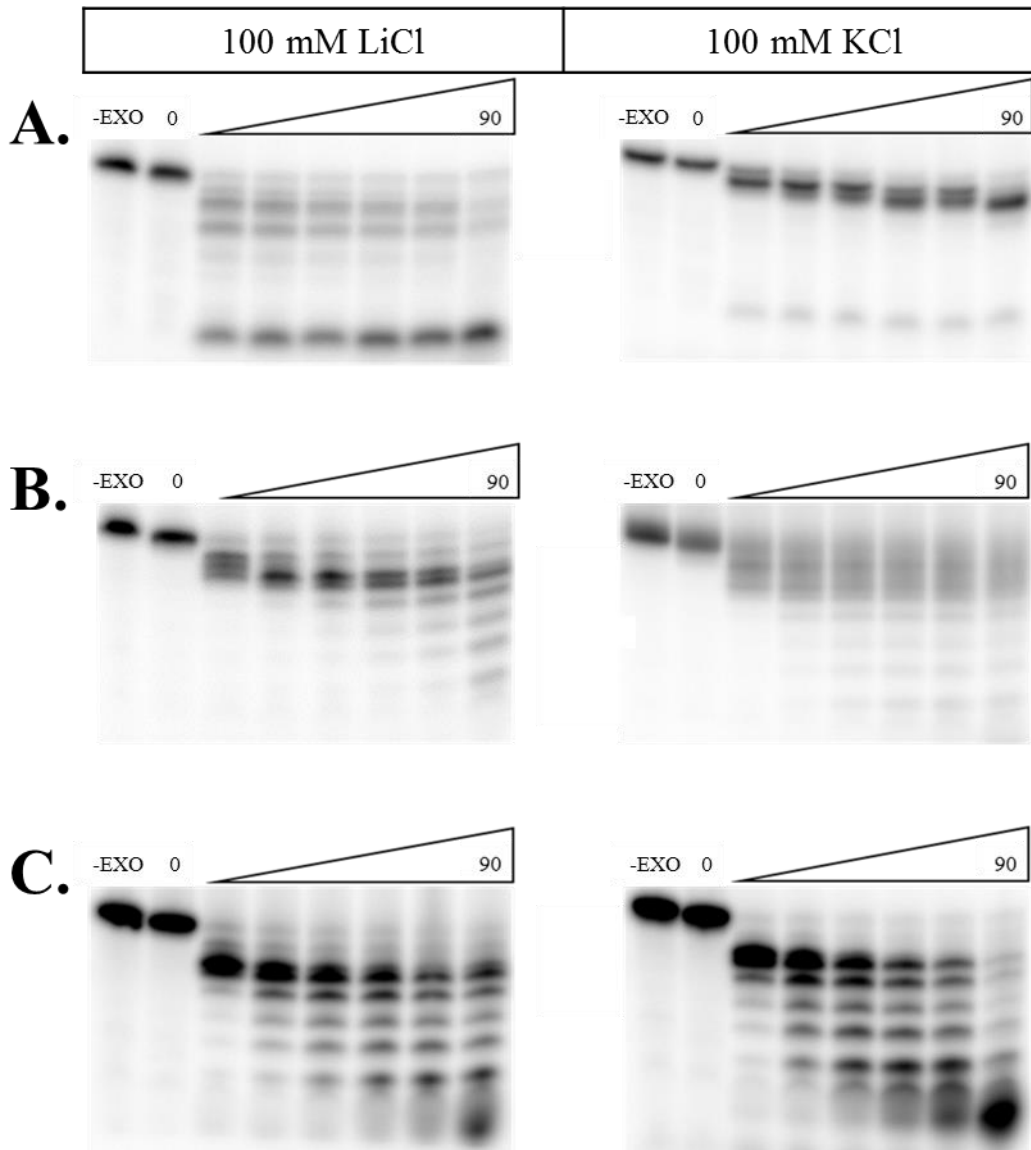
Figure 13. Exonuclease III Activity. Representative images of ExoIII assays in G4 DNA destabilizing conditions (LiCl, left) or G4 DNA stabilizing conditions (KCl, right). On top of each gel are lanes labeled as a negative control (-EXO), a reaction at 0 minutes (0), and a triangle representing increasing time with a 90 minute max. (A) G4 Assays. (B) G4 Com. Assays. (C) GT Assays.
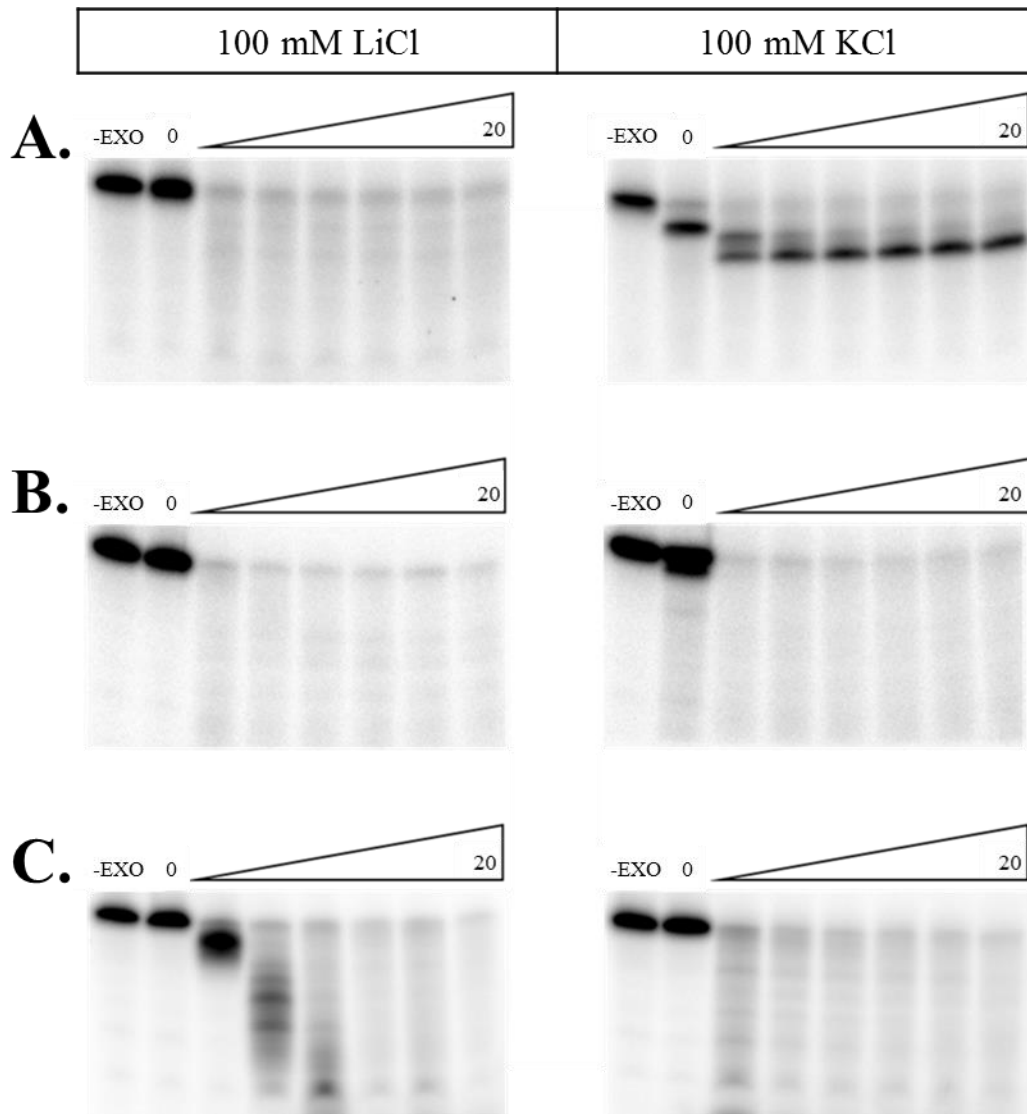
Figure 14. Exonuclease I Activity. Representative images of ExoI assays in G4 DNA destabilizing conditions (LiCl, left) or G4 DNA stabilizing conditions (KCl, right). On top of each gel are lanes labeled as a negative control (-EXO), a reaction at 0 minutes (0), and a triangle representing increasing time with a 20 minute max. (A) G4 Assays. (B) G4 Com. Assays. (C) GT Assays.

CHAPTER IV: DPO4 TRANSLESION DNA POLYMERASE ACTIVITY AT G4

Introduction

Our DNA is vulnerable to both exogenous and endogenous sources of DNA damage. Damaged DNA, particularly when it is bulky, can inhibit polymerase progression and stall an advancing replication fork. The resulting DNA breaks promote genome instability. Using translesion polymerases, cells are able to bypass those replication blockades, which prevents fork stalling. Here, I investigated the fidelity of translesion bypass for DPO4, the model Y-family polymerase from *Sulfolobus sulfataricus*, across sequences that form G-quadruplex structures (G4 DNA).

Translesion Synthesis (TLS)

Translesion synthesis is an activity whereby a polymerase replicates across from damaged DNA templates, and this is a pathway for preventing replication fork stalling during replication. The most common and abundant translesion polymerases are those in the Y-family, which include human polymerase eta, kappa, and iota, all of which have homologs across all domains of life (Ohmori, et al., 2001; Sale, et al., 2012). The active sites for translesion polymerases are larger, more flexible, and able to accommodate bulky lesions or DNA distortions. On undamaged DNA, this results in lower replication fidelity (Sale, et al., 2012). The model Y-family polymerase, DNA polymerase IV (DPO4), comes from a thermostable archaeon, *Sulfolobus sulfataricus,* and its structure was one of the first to be solved for an enzyme of this type (Ling, et al., 2001). DPO4 efficiently bypasses abasic sites and *cys-syn* cyclobutane thymine dimers (CPDs), along with acetyl aminoflourene adducted guanine and cisplatinated guanine (Sale, et al., 2012; Boudsocq, et al., 2001).

Translesion Synthesis of G4 DNA

G4 DNA inhibits replication, but how the cell deals with G4 DNA during replication has not been defined (Weitzmann, et al., 1996). Since it is known that the Y-family translesion polymerases replicate across from large DNA lesions, we reasoned that they may also assist in replicating templates that form G4 DNA. This model enjoys some experimental support. A growing amount of evidence indicates translesion polymerases reduce replication fork stalling when replicating G4 DNA, reducing genome instability. For example, more DSBs were observed when polymerases eta or kappa are absent in cells capable of forming G4 DNA and treated with G4 DNA stabilizing ligands, indicating a role for Y-family translesion polymerases in G4 DNA bypass (Bétous, et al., 2009). In addition, polymerase eta has been shown to have a high affinity for G4 DNA *in vitro,* and recognizes G4 DNA found within the human c-MYC gene (Eddy, et al., 2015). That same study showed polymerase eta had increased fidelity across G4 DNA, indicating a role in G4 DNA bypass (Eddy, et al., 2015). Given evidence on polymerase eta, I sought to clarify the role that fidelity has on G4 DNA bypass by DPO4. Furthermore, using a Y12A DPO4 mutant, I asked how lowered nucleotide selectivity effected G4 DNA bypass capabilities (Sherrer, et al., 2010). Since DPO4 shares kinetic and structural similarities to polymerase eta and polymerase kappa, data could provide a general understanding of how Y-family polymerases bypass G4 DNA (Sale, et al., 2012).

<center>**Methods**</center>

DNA Templates – Polymerase extension reaction templates were synthetic oligonucleotides purchased from Operon (Huntsville, AL) and were page purified. Polymerase extension reaction primers and ladder markers were also purchased from Operon and were salt-free purified. Primers were 5' end labeled with 32P-γ-ATP purchased from Perkin Elmer

<center>56</center>

(Waltham, MA) using T4 polynucleotide kinase purchased from New England Biolabs (Ipswich, MA) and passed through illustra™ Microspin™ G-25 columns purchased from GE Healthcare (Wauwatosa, WI) to remove unincorporated label. Polymerase extension reaction template and primer sequences can be found in Table 5.

Enzymes – The wild type DPO4 vector (pET22B(+)/Dpo4-NHis) was provided by Peter Guengerich from the Vanderbilt University School of Medicine (Zang 2005). A Tyr12Ala point substitution was introduced into pET22B(+)Dpo4-NHis using site directed mutagenesis and verified through DNA sequencing. Both wild type DPO4 and Y12A DPO4 vectors were transformed into *E. coli* strain BL21(DE3)pLysS purchased from Promega (Madison, WI) and purified by Ni^{++} affinity chromatography at the 1 L scale (Zang 2005 and Boudsocq 2001). Cultures were allowed to grow to saturation in LB supplemented with 100 µg/mL ampicillin. Expression was induced with 0.2 mM IPTG for 3 hours at 37°C when cell cultures reached an OD_{600} of 0.3. Cultures were pelleted and resuspended in 25 mL of lysis buffer (20 mM Tris pH 7.6, 10% glycerol, 300 mM NaCl, 0.1% PMFS and 5 mM imidazole) then placed at 85°C for 5 minutes to inactivate *E. coli* proteins. Supernatant was separated by centrifugation for 30 minutes at 20,000 x G. 1 mL of Ni^{++} agarose purchased from Genscript (Piscataway, NJ) was added to supernatant and mixed for 30 minutes at 4°C. Protein-bound Ni^{++} was collected on a column and washed with lysis buffer, 20 column volumes of lysis buffer plus 15 mM imidazole, and 20 column volumes of lysis buffer plus 30 mM imidazole. Proteins were eluted with 5 column volumes of lysis buffer plus 200 mM imidazole. Imidazole was removed with a 3-4 hour dialysis in 20 mM Tris, 25 mM NaCl, and 1mM DTT using a Slide-A-Lyzer® cassette purchased from Thermo Scientific (Waltham, MA). Protein solution was brought to 50% glycerol and a 4-20%

SDS PAGE was used to determine proteins to be >95% pure. Protein concertation was determined using a BSA standard (Berroyer, 2016).

Primer Extension Reactions – Reactions contained 150 fmols of 5' end labeled primer and 100 fmols of template. Primer and templates were annealed and folded by placing in boiling water bath allowed to cool to room temperature. Folding was done in either 50 mM LiCl or KCl and DPO4 polymerase reaction conditions consisting of 125 µM of indicated dNTPs, 1mM DTT, 10 mM Tris-HCl (pH 8.0), and 5 mM $MgCl_2$. The amount of enzyme added was the amount required to extend control template to completion. Extension reactions were conducted at 55 °C for 15 minutes and stopped at 37 °C for 10 minutes with 20 mM EDTA, 60 µg/mL Proteinase K, and 0.4% SDS. Reactions were denatured by adding an equal reaction volume of 90% formamide and incubation at 90 °C for 5 minutes. Reaction products were resolved with 15% denaturing PAGE with 7 M urea and 0.5x TBE. Gels were imaged using a Typhoon FLA 7000 (GE Healthcare).

<div align="center">

Results and Discussion

</div>

The goal of this part of my thesis was to characterize the activity and fidelity of Y12A DPO4 on templates containing G4 structures. Protein purification and characterization was established previously (Berroyer, 2016). The template used is modeled on a 17 base pair sequence (5' – GGGTGGGTTGGGTGGGG) from *N. gonorrhoeae,* which folds into G4 and is required for proper recombination in that organism (Cahoon and Seifert, 2009). Confirmation of G4 DNA formation through circular dichroism was performed previously (Berroyer, 2016), and not shown here.

Replication fidelity is a term used to describe the ability to correctly synthesize DNA, where complementary nucleotides are joined together. Translesion polymerases generally have

low fidelity on undamaged DNA compared to genome replication polymerases, and this is due to their large and flexible active sites (Sale, et al., 2012). While fidelity of DPO4 across from G4 DNA has not been characterized, there is evidence that a human homolog, human pol eta, has high fidelity when replicating G4 DNA templates (Eddy, et al., 2015). The primer extension reactions presented here test fidelity of both wild type and Y12A DPO4 by controlling the nucleotides present in each individual reaction. If there is a large amount of incorporation observed for reactions containing only one nucleotide, the interpretation is that replication fidelity is low since the polymerase extension required misincorporations. The ability of Y-family polymerases to synthesize across from G4 DNA suggests a role in G4 DNA bypass during replication.

Representative images of WT DPO4 primer extension reactions are shown in Figure 16. WT DPO4 is able to complete extension in the presence of all four nucleotides and can misincorporate several nucleotides shown in individual single nucleotide lanes in GT control template primer extension (Figure 16A). Full extension of G4 experimental template by WT DPO4 drops significantly, however a faint band is still present in four-nucleotide reaction (Figure 16B). Single nucleotide lanes showing misincorporation also drops significantly when across G4 DNA with no produced past the 39th position expect in the presence of cytosine where a long smear is present indicative of different length products (Figure 16B). These experiments indicate WT DPO4 has low fidelity on unstructured DNA templates and increased fidelity on structured DNA allowing for bypass of G4 DNA as shown in cytosine lane of Figure 16B. This data is in line with data on human polymerase eta showing increased fidelity across if G4 DNA (Eddy, et al., 2015).

Representative images of Y12A DPO4 primer extension reactions are shown in Figure 17. Similar to WT, Y12A DPO4 is able to complete extension in the presence of all four nucleotides and can misincorporate several nucleotides as shown in single nucleotide lanes in GT control template primer extension (Figure 17A). Unlike WT however, there is incomplete extension of G4 experimental template in reaction containing all four nucleotides, observed as a short smear of bands (Figure 17B). The level of misincorporation also apparently drops, with no extension past the 38th position with the exception of the thymine lane, where a faint band is observed, and cytosine, where it appears extension stops after second run of guanines. The mutation Y12A in DPO4 decreases sugar selectivity and increases nucleotide misincorporation (Sherrer, et al., 2010), implying that synthesis of G4 templates is likely high fidelity.

The importance of the 39th position in G4 DNA template is due to the possible structures present in solution. The first guanine run in the template contains four guanines while all others contain three. This creates the possibility that either the first guanine or the fourth guanine are participating in G4 DNA formation while the other is not. The second guanine in the first run of guanines at the 39th position should participate in G-quadruplex formation regardless of the specific structure that forms, so synthesis products at or past the 39th position must be the result of structure bypass by the polymerase.

In conclusion, my results suggest that WT DPO4 synthesis fidelity increases when templates form G4 DNA structures, indicating that structured templates are synthesis substrates and that the Y-family polymerases may be recruited during replication to ensure accurate bypass. This not only prevents fork collapse, but also reduces the potential for mutagenesis at guanine-rich loci. That model is supported by results showing incomplete extension by Y12A DPO4 on the G4 DNA template, when any nucleotide other than cytosine was supplied to the reaction

(Figure 16B). In other words, the loosening of restrictions on nucleotide usage through Y12A

substitution mutagenesis did not permit structure bypass by the DPO4 polymerase. Furthermore,

when reactions were supplied with only cytosine nucleotides, synthesis through the G4-folded

template was observed, but not when reactions were supplemented with the one of the other three

nucleotides (Figure 15B). During genome replication, the Y-family polymerases may be

recruited when replication forks stall at G4 DNA structures. High-fidelity synthesis across from

G4-folded guanines permits fork progression and accurate replication to reduce the potential for

mutagenesis. Therefore, in addition to DNA damage bypass, the Y-family polymerases may be

important for accurate replication through repetitive DNA sequences.

References

Berroyer, A. N. (2016) *Translesion Synthesis of G-Quadruplex DNA Structures* (Master's Thesis). Illinois State University, Normal, IL.

Bétous, R., Rey, L., Wang, G., Pillaire, M.-J., Puget, N., Selves, J., Biard, D. S. F., Shin-ya, K., Vasquez, K. M., Cazaux, C., and Hoffmann, J.-S. (2009). Role of TLS DNA polymerases eta and kappa in processing naturally occurring structured DNA in human cells. *Molecular Carcinogenesis*, *48*(4), 369–378.

Boudsocq, F., Iwai, S., Hanaoka, F., & Woodgate, R. (2001). Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polη. *Nucleic Acids Research*, 29(22), 4607-4616.

Cahoon, L. A., & Seifert, H. S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science*, *325*(5941), 764-767.

Eddy, S., Maddukuri, L., Ketkar, A., Zafar, M. K., Henninger, E. E., Pursell, Z. F., & Eoff, R. L. (2015). Evidence for the kinetic partitioning of polymerase activity on G-quadruplex DNA. *Biochemistry, 54*(20), 3218-3230.

Ling, H., Boudsocq, F., Woodgate, R., & Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell, 107*(1), 91-102.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L, Prakash, S., Todo, T., Walker, G. C., Wang, Z., & Woodgate, R. (2001). The Y-family of DNA polymerases. *Molecular cell, 8*(1), 7-8.

Sale, J. E., Lehmann, A. R., & Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nature reviews. *Molecular cell biology, 13*(3), 141.

Sherrer, S. M., Beyer, D. C., Xia, C. X., Fowler, J. D., & Suo, Z. (2010). Kinetic basis of sugar selection by a Y-family DNA polymerase from Sulfolobus solfataricus P2. *Biochemistry*, 49(47), 10179-10186.

Weitzmann, M. N., Woodford, K. J., & Usdin, K. (1996). The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *Journal of Biological Chemistry, 271*(34), 20958-20964.

Table 5.

Primer Extension Template and Ladder Marker Sequences. Guanines participating in G-quadruplex are shown in bold.

Name:	Sequence (5' – 3'):	Length:
pilE GT	GTCGTCTCGTCTCGTTGCTCTGTGTTGGTAGTTTGTT TGTTGTTGTTTCATGGTCATAGCTGT	69 nt
38 GT	CAGGAAACAGCTATGACCATGAAACAACAACAAAC AAA	38 nt
39 GT	CAGGAAACAGCTATGACCATGAAACAACAACAAAC AAAC	39 nt
42 GT	CAGGAAACAGCTATGACCATGAAACAACAACAAAC AAACTAC	42 nt
43 GT	CAGGAAACAGCTATGACCATGAAACAACAACAAAC AAACTACC	43 nt
44 GT	CAGGAAACAGCTATGACCATGAAACAACAACAAAC AAACTACCA	44 nt
pilE G4	GTCTTCTCTCTTTCT**GGG**T**GGG**TT**GGG**T**GGGG**ATTT CTTTTCTTTTTTTCATGGTCATAGCTGTTTCCTG	69 nt
37 G4	CAGGAAACAGCTATGACCATGAAAAAGAAAAGAA AT	37 nt
38 G4	CAGGAAACAGCTATGACCATGAAAAAGAAAAGAA ATC	38 nt
39 G4	CAGGAAACAGCTATGACCATGAAAAAGAAAAGAA ATCC	39 nt
42 G4	CAGGAAACAGCTATGACCATGAAAAAGAAAAGAA ATCCCCA	42 nt
46 G4	CAGGAAACAGCTATGACCATGAAAAAGAAAAGAA ATCCCCACCCA	46 nt

A.

GT Template

5'* ——38 nt——AA
————————TT**G**ATGGTTGTGTCTCGT...

G4 Template

5'* ——37 nt——T
————————AG**G**GGTGGGTTGGGTGGG...

B.

Left gel: WT GT   L (arrow pointing left for complete extension product)
Lanes: dNTP 4d dA dT dG dC
Markings: 39 nt / 38 nt

Right gel: WT G4   L (arrow pointing left)
Lanes: dNTP 4d dA dT dG dC
Markings: 39 nt / 38 nt / 37 nt

Figure 15. WT DPO4 Fidelity at G4 DNA. Representative images of extension reactions are shown with the template annotated along the top and the running start primer length. Complete extension products are marked with an arrow. Nucleotide contents for each reaction/lane are annotated along the bottom. The last lane contains a ladder. (A) Single nucleotide extension reactions of WT DPO4 on unstructured control template. Primer was 38 nucleotides long and ladder contains both 38 and 39 oligonucleotide lengths. (B) Single nucleotides extension reactions of WT DPO4 on structured experimental template. Primer was 37 nucleotides long and ladder contained 37, 38, and 39 oligonucleotide lengths.

A.

GT Template

5'* ———38 nt———AA
      ————————TT**G**ATGGTTGTGTCTCGT...

G4 Template

5'* ———37 nt———T
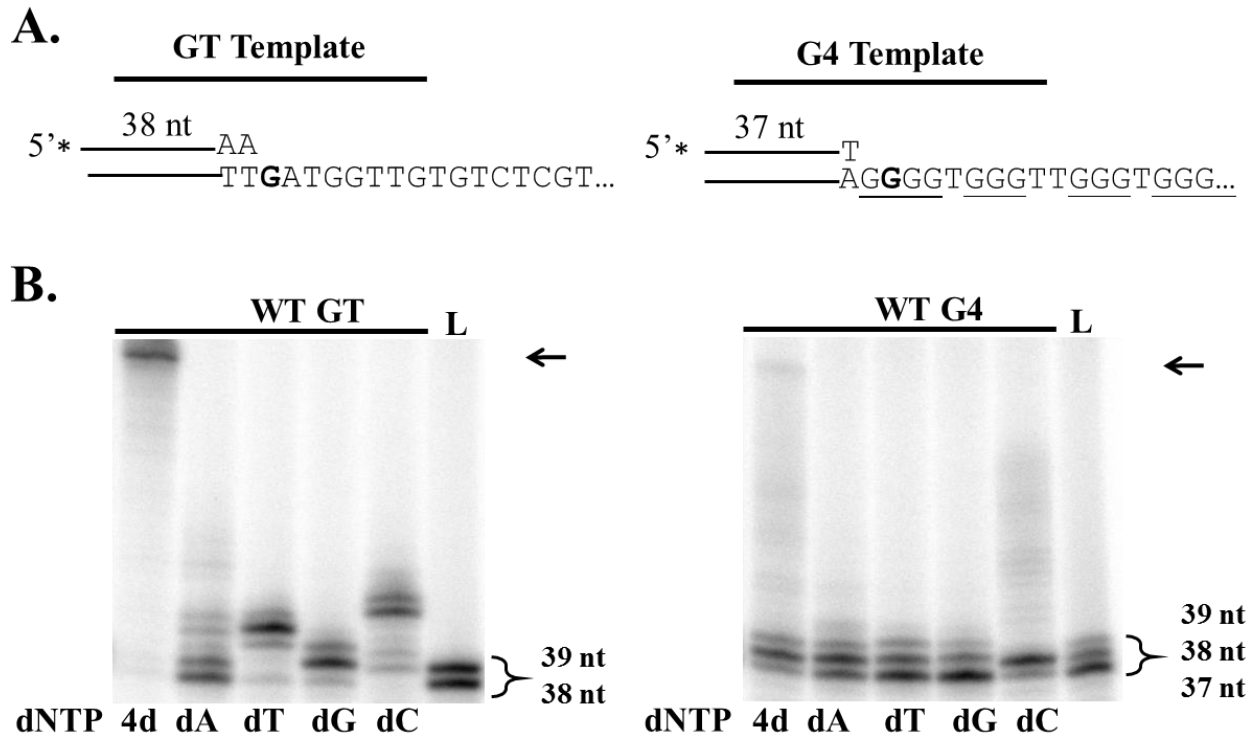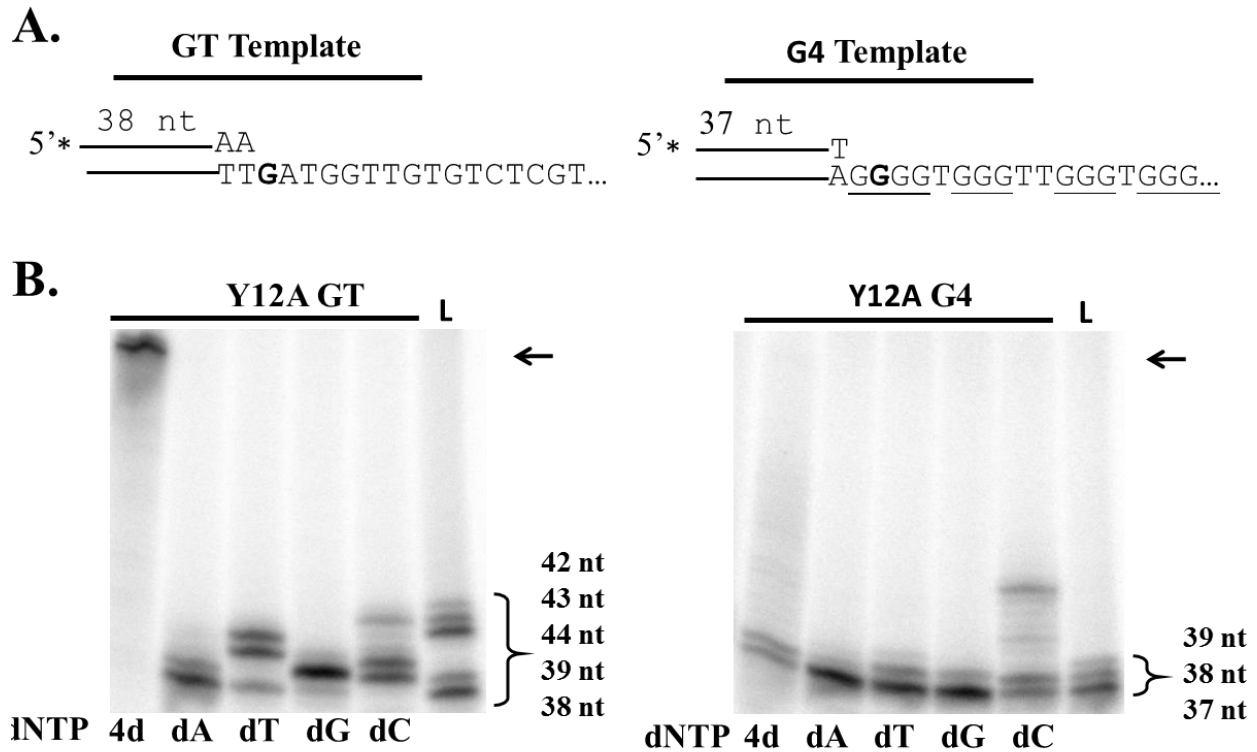   ————————AG**G**GGTGGGTTGGGTGGG...

B.

Figure 16. Y12A DPO4 Fidelity at G4 DNA. Representative images of extension reactions are shown with the template annotated along the top and the running start primer length. Complete extension products are marked with an arrow. Nucleotide contents for each reaction/lane are annotated along the bottom. The last lane contains a ladder. (A) Single nucleotide extension reactions of Y12A DPO4 on unstructured control template. Primer was 38 nucleotides long and ladder contains 38, 39, 42, 43, and 44 oligonucleotide lengths. (B) Single nucleotides extension reactions of Y12A DPO4 on structured experimental template. Primer was 37 nucleotides long and ladder contained 37, 38, and 39 oligonucleotide lengths.

CHAPTER V: CONCLUSIONS

Guanine-rich areas in the genome support the formation of a thermostable four-stranded structure called G-quadruplex DNA (G4 DNA). Guanine-rich regions in the genome are preserved in promoter regions, ribosomal DNA, and immunoglobulin heavy-chain switch regions and are highly conserved within the human population and between different species suggesting a role in DNA metabolism (Huppert and Balasubramanian, 2005; Todd, et al., 2005; Capra, et al., 2010; Maizels, 2006). Even so, areas capable of G4 DNA formation are prone to higher mutagenesis rates and have increased genomic instability, though causes are not fully understood (Maizels, 2006; Maizels, 2015; Bochman, et al., 2012; Rhodes and Lipps, 2015; Vasquez and Wang, 2013). In this investigation, I assayed various repair capabilities on oligonucleotides containing a single G4 DNA structure, providing molecular rationale for the increases in mutagenesis observed at guanine-rich loci.

In Chapter II, I investigated the ability of base excision repair to remove uracil around G4 DNA using two major uracil DNA glycosylases; *E. coli* UDG and human SMUG1 (Visnes, et al., 2009; Kavli, et al., 2002; Pettersen, et al., 2007). Uracil is one of the most common form of lesions in the genome and mutagenesis would quickly accumulate if repair mechanisms were impaired (Krokan, et al., 2002). These data suggest inhibited repair of uracil by UDG adjacent to G4 DNA and in between two G4 DNAs in close proximity of each other (Figure 8). Inhibited repair of uracil by UDG suggests a possible accumulation of point substitutions during replication, where UDG is most active. Data here also suggests large inhibition of hSMUG1 in between two G4 DNAs in close proximity of each other as well as three bases away 3' and over five bases away 5' to G4 DNA (Figure 10). Inhibited repair of uracil by hSMUG1 suggests a possible accumulation of point substitutions in highly transcribed regions of non-proliferation

cells, where hSMUG1 is most active. Together these data can possibly account for a fraction of the mutagenesis and genomic instability observed at G4-capable sequences.

In Chapter III, I investigated the ability of *E. coli* exonuclease III (ExoIII) and exonuclease I (ExoI) to excise DNA containing a G4 structure. ExoIII is able to facilitate repair of damaged nucleotide bases due to radiation and oxidation and is the major abasic site endonuclease for BER in *E. coli* (Lovett, 2011; Weiss, 1976; Weiss, 1981). ExoI is part of a large network of proteins involved in replication and DNA repair, specifically mismatch repair, and is shown to be an important component in recombination (Butland, et al., 2005; Friedman-Ohana and Cohen, 1998; Thoms, et al., 2008; Thoms and Wackernagel, 1998; Viswanathan, et al., 2001; Viswanathan and Lovett, 1998; Burdett, et al., 2001; Cooper, et al., 1993). These data indicate ExoIII and ExoI are unable to excise past G4 DNA even after extended periods of time (Figure 13 and Figure 14 respectively). This suggests G4 DNA is a block to normal exonuclease activity, and given their roles in DNA repair pathways, could possible account for a fraction of the mutagenesis and genomic instability observed at G4-capable sites.

Finally in Chapter IV, I investigated the bypass ability and fidelity requirements of thermostable translesion polymerase DPO4 past G4 DNA (Ling, et al., 2001; Sale, et al., 2012; Boudsocq, et al., 2001). Misincorporation ability was inhibited and complete extension on structured templates was observed in small quantities compared to unstructured control templates indicating DPO4 is able to bypass G4 DNA and fidelity is increased across structures (Figure 16). Sugar selectivity requirements were investigated using Y12A DPO4 mutant known to have a mutation that decreases sugar selectivity (Sherrer, et al., 2010). When compared to WT DPO4, extensions with Y12A DPO4 on structured DNA resulted in no complete extension and reduced ability to extend in cytosine indicating sugar selectivity capabilities are important in G4 DNA

68

bypass (Figure 17). Together these data indicated fidelity is an important component to G4 DNA bypass

In conclusion, G4 DNA is an impediment to normal DNA repair activities, including base excision repair, mismatch repair, and recombination. Increased mutagenesis at G4-capable sites could in large part be due to inhibited repair activities investigated in this work. Investigations on how the cell is able to accommodate an impediment to DNA metabolic activities are still underway though here I elucidated the involvement of Y-family translesion polymerases. Further work on G-quadruplex structures is required for a complete understanding of the consequences it presents on the genome as well as how the cell is able to maintain genome integrity in spite of it.

References

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews Genetics*, *13*(11), 770-780.

Boudsocq, F., Iwai, S., Hanaoka, F., & Woodgate, R. (2001). Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polη. *Nucleic Acids Research*, 29(22), 4607-4616.

Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S. T., & Modrich, P. (2001). In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proceedings of the National Academy of Sciences*, 98(12), 6765-6770.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., & Emili, J. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, *433*(7025), 531-537.

Capra, J. A., Paeschke, K., Singh, M., & Zakian, V. A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS computational biology*, *6*(7), e1000861.

Cooper, D. L., Lahue, R. S., & Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. *Journal of Biological Chemistry*, 268(16), 11823-11829.

Friedman-Ohana, R., & Cohen, A. (1998). Heteroduplex joint formation in Escherichia coli recombination is initiated by pairing of a 3′-ending strand. *Proceedings of the National Academy of Sciences*, *95*(12), 6909-6914.

Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, *33*(9), 2908-2916.

Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L.,

  Krokan, H. E., & Slupphaug, G. (2002). hUNG2 is the major repair enzyme for removal

  of uracil from U: A matches, U: G mismatches, and U in single-stranded DNA, with

  hSMUG1 as a broad specificity backup. *Journal of Biological Chemistry*, *277*(42),

  39926-39936.

Krokan, H. E., Drabløs, F., & Slupphaug, G. (2002). Uracil in DNA-occurrence, consequences

  and repair. *Oncogene*, *21*(58), 8935.

Ling, H., Boudsocq, F., Woodgate, R., & Yang, W. (2001). Crystal structure of a Y-family DNA

  polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*,

  107(1), 91-102

Lovett, S. T. (2011). The DNA exonucleases of escherichia coli. *EcoSal Plus*, *4*(2).

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nature*

  structural & molecular biology, *13*(12), 1055.

Maizels, N. (2015). G4-associated human diseases. *EMBO reports*, e201540607.

Pettersen, H. S., Sundheim, O., Gilljam, K. M., Slupphaug, G., Krokan, H. E., & Kavli, B.

  (2007). Uracil–DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base

  excision repair by distinct mechanisms. *Nucleic acids research*, *35*(12), 3879-3892.

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic*

  acids research, *43*(18), 8627-8637.

Sale, J. E., Lehmann, A. R., & Woodgate, R. (2012). Y-family DNA polymerases and their role

  in tolerance of cellular DNA damage. *Nature reviews. Molecular cell biology*, 13(3), 141.

Sherrer, S. M., Beyer, D. C., Xia, C. X., Fowler, J. D., & Suo, Z. (2010). Kinetic basis of sugar

    selection by a Y-family DNA polymerase from Sulfolobus solfataricus P2. *Biochemistry*,

    49(47), 10179-10186.

Thoms, B., Borchers, I., & Wackernagel, W. (2008). Effects of single-strand DNases ExoI, RecJ,

    ExoVII, and SbcCD on homologous recombination of recBCD+ strains of Escherichia

    coli and roles of SbcB15 and XonA2 ExoI mutant enzymes. *Journal of*

    bacteriology, *190*(1), 179-192.

Thoms, B., & Wackernagel, W. (1998). Interaction of RecBCD enzyme with DNA at double-

    strand breaks produced in UV-irradiated Escherichia coli: requirement for DNA end

    processing. *Journal of bacteriology*, *180*(21), 5639-5645.

Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence

    motifs in human DNA. *Nucleic acids research*, *33*(9), 2901-2907.

Vasquez, K. M., & Wang, G. (2013). The yin and yang of repair mechanisms in DNA structure-

    induced genetic instability. *Mutation Research/Fundamental and Molecular Mechanisms*

    of Mutagenesis, *743*, 118-131.

Visnes, T., Doseth, B., Pettersen, H. S., Hagen, L., Sousa, M. M., Akbari, M., Otterlei, M., Kavli,

    B., Slupphaug, G., & Krokan, H. E. (2009). Uracil in DNA and its processing by different

    DNA glycosylases. *Philosophical Transactions of the Royal Society of London B:*

    Biological Sciences, *364*(1517), 563-568.

Viswanathan, M., & Lovett, S. T. (1998). Single-strand DNA-specific exonucleases in

    Escherichia coli: roles in repair and mutation avoidance. *Genetics*, 149(1), 7-16.

Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P., & Lovett, S. T. (2001). Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair. *Journal of Biological Chemistry*, *276*(33), 31053-31058.

Weiss, B. (1976). Endonuclease II of Escherichia coli is exonuclease III. *Journal of Biological Chemistry*, *251*(7), 1896-1901.

Weiss, B. (1981). 12 Exodeoxyribonucleases of Escherichia coli. *The enzymes*, *14*, 203-231.