

8-2010

DOUBLE-STRAND BREAK REPAIR PATHWAYS IN DNA STRUCTURE- INDUCED GENETIC INSTABILITY

Diem T. Kha

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the [Laboratory and Basic Science Research Commons](#)

Recommended Citation

Kha, Diem T., "DOUBLE-STRAND BREAK REPAIR PATHWAYS IN DNA STRUCTURE-INDUCED GENETIC INSTABILITY" (2010). *UT GSBS Dissertations and Theses (Open Access)*. Paper 76.

This Thesis (MS) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

**DOUBLE-STRAND BREAK REPAIR PATHWAYS IN DNA STRUCTURE-INDUCED
GENETIC INSTABILITY**

BY

Diem Thi Kha, Bachelor of Science

APPROVED:

Karen M. Vasquez, Ph.D., Supervisory

David G. Johnson, Ph.D.

David L. Mitchell, Ph.D.

Feng Wang-Johanning, M.D., Ph.D.

Richard D. Wood, Ph.D.

APPROVED:

Dean, The University of Texas

**DOUBLE-STRAND BREAK REPAIR PATHWAYS IN DNA STRUCTURE-INDUCED
GENETIC INSTABILITY**

**A
THESIS**

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

by

Diem Thi Kha, B.S.
Houston, Texas

August 2010

DEDICATION

To my beloved late grandmother, Nu Thi Lam, my “ba ngoai”.

ACKNOWLEDGEMENTS

I would like to first thank Jesus Christ, my Lord and Savior, for being God in my life and leading me into the wonderful lab of Dr. Karen Vasquez. I thank Dr. Karen Vasquez for adopting me into the Vasquez lab as her first Master's degree-seeking student. She has been an awesome mentor to me in guiding, advising, and providing support and wisdom. The former and current members of the Vasquez lab were also supportive to my research studies. Specifically, special thanks to Dr. Graham Wang for teaching and giving me the basis of my research, along with Nithya Natrajan and Dr. Lynn Harrison for being the frontrunners in contributing to the pieces of this work too.

In relation to my Master's degree pursuit, I would like to thank all of my participating supervisory committee members: Dr. David Johnson, Dr. Rodney Nairn, Dr. David L. Mitchell, Dr. Feng Wang-Johanning, Dr. Richard Wood and including my supervisor, Dr. Karen Vasquez. I appreciate their time, feedback and the effort they made in aligning all necessary requirements for finishing and completing my research here.

Besides the great working environment and location of Science Park Research Division in Smithville, TX, I enjoyed the company of other colleagues, classmates and faculty members that taught the classes and shared their knowledge of their own research. In addition, I am grateful for the works of the molecular biology core for their sequencing analysis; Sarah Smith (Henninger) in manuscript preparation and distributing GSBS forms to the necessary parties; Kevin Lin for statistical analysis help and Becky Brooks for her coordination in organizing, reminding and administering to all the students.

Lastly, I am blessed to have the love and encouragement of my family and friends.

DOUBLE-STRAND BREAK REPAIR PATHWAYS IN DNA STRUCTURE-INDUCED GENETIC INSTABILITY

Diem Thi Kha, M.S.

Supervisory professor: Karen M. Vasquez, Ph.D.

Genetic instability in mammalian cells can occur by many different mechanisms. In the absence of exogenous sources of DNA damage, the DNA structure itself has been implicated in genetic instability. When the canonical B-DNA helix is naturally altered to form a non-canonical DNA structure such as a Z-DNA or H-DNA, this can lead to genetic instability in the form of DNA double-strand breaks (DSBs) (1, 2). Our laboratory found that the stability of these non-B DNA structures was different in mammals versus *Escherichia coli* (*E.coli*) bacteria (1, 2). One explanation for the difference between these species may be a result of how DSBs are repaired within each species. Non-homologous end-joining (NHEJ) is primed to repair DSBs in mammalian cells, while bacteria that lack NHEJ (such as *E.coli*), utilize homologous recombination (HR) to repair DSBs. To investigate the role of the error-prone NHEJ repair pathway in DNA structure-induced genetic instability, *E.coli* cells were modified to express genes to allow for a functional NHEJ system under different HR backgrounds. The *Mycobacterium tuberculosis* NHEJ sufficient system is composed of Ku and Ligase D (LigD) (3). These inducible NHEJ components were expressed individually and together in *E.coli* cells, with or without functional HR (RecA/RecB), and the Z-DNA and H-DNA-induced mutations were characterized. The Z-DNA structure gave rise to higher mutation frequencies compared to the controls, regardless of the DSB repair pathway(s) available; however, the type of mutants produced after repair was greatly dictated on the available DSB repair system, indicated by the shift from 2% large-scale deletions in the total mutant population to 24% large-scale deletions when NHEJ was present (4). This suggests that NHEJ has a role in the large deletions induced by Z-DNA-forming sequences. H-DNA structure, however, did not exhibit an increase in mutagenesis in the newly engineered *E.coli* environment, suggesting the involvement of other factors in regulating H-DNA formation/stability in bacterial cells. Accurate repair by established DNA DSB repair pathways is essential to maintain the stability of eukaryotic and prokaryotic genomes and our results suggest that an error-prone NHEJ pathway was involved in non-B DNA structure-induced mutagenesis in both prokaryotes and eukaryotes.

TABLE OF CONTENTS

	PAGE
APPROVAL PAGE	i
TITLE PAGE	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	vii
LIST OF TABLES	ix
CHAPTER I: BACKGROUND AND INTRODUCTION	1
I.1 NON-B DNA BACKGROUND.....	2
<i>I.1.1 Z-DNA</i>	4
<i>I.1.2 H-DNA</i>	6
I.2 DNA DOUBLE-STRAND BREAK REPAIR	10
<i>I.2.1 Non-homologous End-Joining in Mammalian Cells and Bacterial Cells</i>	11
<i>I.2.2 Homologous Recombination in Mammalian Cells and Bacterial Cells</i>	13
I.3 HYPOTHESIS AND SPECIFIC AIMS.....	15
CHAPTER II: DOUBLE-STRAND BREAK REPAIR PATHWAYS IN Z-DNA AND H-DNA- INDUCED GENOMIC INSTABILITY	17
II.1 INTRODUCTION	18
II.2 MATERIALS AND METHONDS	20
II.3 RESULTS AND DISCUSSION.....	25
CHAPTER III: FUTURE DIRECTIONS	54
III.1 FUTURE DIRECTIONS.....	55
REFERENCES	59
VITA	69

LIST OF FIGURES

FIGURE 1.	THE B-DNA STRUCTURE AND SOME EXAMPLES OF NON-B DNA STRUCTURES	3
FIGURE 2.	THE MUTATION FREQUENCIES OF THE NON-B DNA STRUCTURES, Z-DNA AND H-DNA.....	7
FIGURE 3.	Z-DNA-INDUCED MUTATIONS FROM TWO CELL TYPES, BACTERIAL AND MAMMALIAN CELLS	8
FIGURE 4.	LM-PCR RESULTS OF BOTH, Z-DNA AND H-DNA.....	9
FIGURE 5.	MODEL OF THE NON-HOMOLOGOUS END-JOINING PATHWAY FOR MAMMALIAN CELLS AND PROKARYOTES	12
FIGURE 6.	MODEL OF THE HOMOLOGOUS RECOMBINATION PATHWAY FOR MAMMALIAN CELLS AND PROKARYOTES	14
FIGURE 7.	WESTERN ANALYSES TO VERIFY THE JOINT EXPRESSION OF THE MT-NHEJ PROTEINS, MT-KU AND MT-LIGD, FROM DIFFERENT BACTERIAL STRAINS USED IN THE STUDY	27
FIGURE 8.	WESTERN ANALYSES TO VERIFY INDIVIDUAL EXPRESSION OF THE MT-NHEJ PROTEINS, MT-KU OR MT-LIGD, IN THE MT-KU ONLY AND MT-LIGD ONLY STRAINS	28
FIGURE 9.	THE INDUCIBLE FACTOR, L-ARABINOSE, IS NOT MUTAGENIC IN TO THE EXPERIMENTAL MUTATION-REPORTER SYSTEM <i>PER SE</i>	29
FIGURE 10.	CELLS EXPRESSING THE FUNCTIONAL MT-NHEJ PROTEINS PROCESS THE DSBS IN A MUTAGENIC FASHION	32
FIGURE 11.	SPONTANEOUS AND Z-DNA-INDUCED MUTATION FREQUENCIES AND SPECTRA IN THE RECA- STRAIN	37

FIGURE 12.	SPONTANEOUS AND Z-DNA-INDUCED MUTATION FREQUENCIES AND SPECTRA IN THE RECB- STRAIN	38
FIGURE 13.	SPONTANEOUS AND Z-DNA-INDUCED MUTATION FREQUENCIES AND SPECTRA IN THE WT/NHEJ+ STRAIN	39
FIGURE 14.	SEQUENCED JUNCTIONS OF THE Z-DNA-INDUCED MUTANTS FROM THE RECB- STRAIN	40
FIGURE 15.	Z-DNA-INDUCED AND SPONTANEOUS MUTATION FREQUENCIES AND SPECTRA IN MT-KU ONLY AND MT-LIGD ONLY BACTERIAL STRAINS	42
FIGURE 16.	PROPOSED MODEL FOR THE ROLE OF DOUBLE-STRAND REPAIR PATHWAYS IN PROCESSING OF Z-DNA-INDUCED DSB	45
FIGURE 17.	THE INDUCIBLE FACTOR, L-ARABINOSE, DOES NOT AFFECT H-DNA-INDUCED MUTAGENESIS <i>PER SE</i>	48
FIGURE 18.	SPONTANEOUS AND H-DNA-INDUCED MUTATION FREQUENCIES FROM THE RECA- STRAIN	49
FIGURE 19.	SPONTANEOUS AND H-DNA-INDUCED MUTATION FREQUENCIES FROM THE RECB- STRAIN	50
FIGURE 20.	SPONTANEOUS AND H-DNA-INDUCED MUTATION FREQUENCIES FROM THE WT/NHEJ+ STRAIN	51

LIST OF TABLES

**TABLE 1. BACTERIAL STRAINS USED IN THIS STUDY WITH THEIR
PROFICIENCIES AND DEFICIENCIES IN HR (RECA/RECB)
AND MT-NHEJ (KU/LIGD)..... 24**

CHAPTER I: BACKGROUND AND INTRODUCTION

I.I NON-B DNA BACKGROUND

Research on DNA structure did not cease progression after the scientific revolutionary discovery by Watson and Crick in 1953 of the canonical B-form right-handed DNA double helix conformation (displayed on the top in Figure 1) (5). On the contrary, the field increased with many other findings in the study of the dynamics of DNA structure, specifically identifying and characterizing at least twelve different (non-B DNA) structures that differ from the traditional B-DNA conformation originally described by Watson and Crick. Examples of such structures include hairpins/cruciforms, left-handed Z-DNA, intramolecular triplex H-DNA, G-quadruplex (tetraplex) DNA, and slipped/sticky DNA (shown in Figure 1) (6). These secondary structures have been detected *in vitro* and also several have been verified *in vivo* by different methods such as using antibodies in fluorescence immunostaining (7). This growing list of non-B DNA structures is still open and through better *in vitro* and/or *in vivo* detection more structures are yet to be discovered and characterized (6).

The relevance of identifying additional types of non-B DNA structures and the continual desire for more knowledge about these structures lies in the strong correlation with their structural effects and regulation of biological processes such as DNA replication, transcription, recombination, and genome stability (2, 8-10). Many non-B DNA-forming sequences have been mapped to “hotspots” of chromosome breakage in diseases such as in myeloma, leukemia and lymphomas, neurodegenerative and genomic disorders (for review see (10, 11)). For example, in mammalian cells, the human oncogene *c-MYC* and proto-oncogene *BCL-2* contain non-B DNA-forming sequences that are often subjected to DNA double-stranded breaks (DSBs) and chromosomal translocation (1, 2, 12-15). These unusual DNA structures can be recognized, processed and treated differently than the canonical B-DNA in the cell. The rigorous genome maintenance from DNA repair machinery might encounter non-B DNA as ‘damaged’ and may begin to proactively remove or correct such ‘damage’ from the genome (16). Therefore, the role of DNA repair and its correspondence to these non-B DNA structures are of interest. We have discovered that some types of non-B DNA (e.g. Z-DNA and H-DNA) can cause DSBs and result in large deletions and rearrangements in mammalian cells (1, 2). Thus, the role of DNA DSB repair pathways, non-homologous end-joining (NHEJ) and homologous recombination (HR), in non-B DNA induced mutations are of interest and of importance.

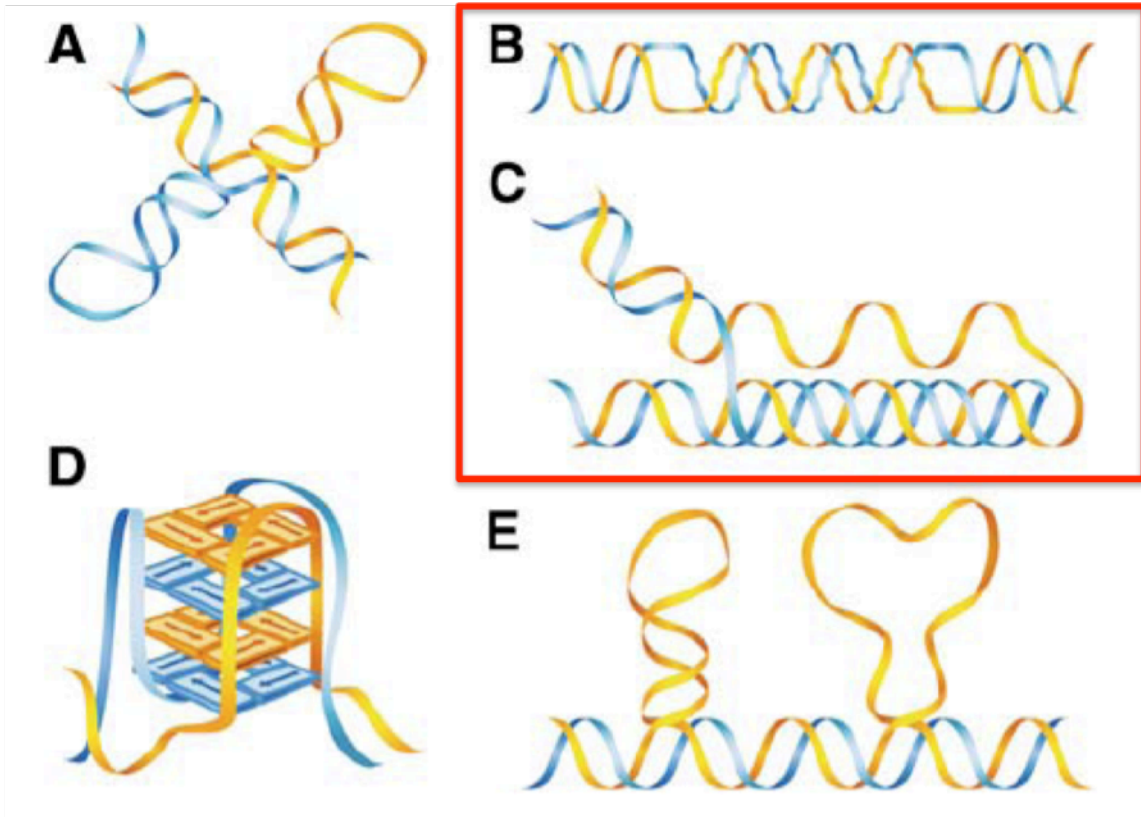


Figure 1. The B-DNA structure and some examples of non-B DNA structures. The canonical Watson and Crick B-DNA structure is placed above all the examples of non-B DNA structures. (A) Hairpin/cruciform DNA; (B) left-handed Z-DNA; (C) intramolecular triplex H-DNA; (D) G-quadruplex (tetraplex) DNA; and (E) slipped DNA. The two non-B DNA structures, Z-DNA and H-DNA, are the subject of my thesis project and are boxed in red. (Adapted from Zhao et al., 2009)

I.I.I Z-DNA

The canonical Watson and Crick B-DNA double helix can reverse its winding direction from a right-hand helix to a left-handed Z-DNA helix in negative supercoiled environments, which can occur, for example, during replication or transcription. Z-DNA structures can form at alternating purine-pyrimidine regions, such as (CG:CG)_n and (CA:TG)_n, leaving the DNA structure with a zigzag arrangement as seen in Figure 1B (17-19). In this Z-DNA secondary structural conformation, the helix becomes elongated; making a deeper narrow minor groove that also shifts the visibility of the major groove (17-19). At the intersecting B-Z junctions, two bases are extruded from the helix and might be susceptible potential sites for DNA modification (20). Many proteins have been found to preferentially bind to the Z-DNA structure, such as the ADAR1 protein, vaccinia virus E3L protein, DLM-1 protein and the RecA protein (16, 21, 22). The Z-DNA structure, and not the linear sequence *per se*, can have an operational function in regulating gene expression, deletion/translocation, and DNA recombination that includes initiating HR by alleviating DNA supercoiling (16, 22-27). There is an abundance of Z-DNA-forming sequences in the eukaryotic genome with an approximate estimate of one per three thousand base pair (bp) in the human genome (28, 29). Z-DNA-forming sequences co-localize in the genome with hotspots for chromosomal breakage in human diseases, including translocation-related cancers, leukemias and lymphomas (10, 11). In addition, Z-DNA-forming sequences co-localize to the breakpoint hotspots of the amyloid precursor protein (APP), Presenilin and ApoE that are connected to the Alzheimer's disease (6, 30).

In our laboratory, we have studied Z-DNA-induced genomic instability in bacterial *E.coli* cells, cultured mammalian cells, and transgenic mice; we found that Z-DNA stimulates mutations in all three systems (1, 31, 32). Figure 2A shows a specific example of the comparison of the Z-DNA-induced mutation frequencies between mammalian COS-7 cells versus *E.coli* DH5 α cells. Using a *lacZ'* mutation-reporter system via facile blue/white screening, individual mutation frequencies were calculated and graphed in Figure 2A for different sequences that have varying degrees of Z-DNA-forming propensity; CG(14) has the highest capability, decreasing in the hairpin-forming sequences RW1009/ UY1, and lastly, the least capable with the control scramble sequence, CON. Of all the sequences, the CG(14) sequence, the most capable in forming the Z-DNA structure, yielded the highest mutation frequencies in both the

mammalian and bacterial cells, compared to the hairpin-forming and control sequences. This suggests that it was the Z-DNA structure, rather than the hairpin structure that caused the mutagenesis in mammalian cells (1).

Upon the characterization of the mutants, using a restriction analysis with a double digestion by EagI and BssSI enzymes, and sequencing analysis, a more notable difference was seen in the spectrum of mutations caused by Z-DNA between the mammalian cells and bacterial cells. Shown on the right panel of Figure 3, the majority of the mammalian Z-DNA-induced mutants, approximately 95% of them, have a complete loss of the 877 bp restriction fragment containing the 28 bp Z-DNA-forming CG repeat sequence (1). Furthermore, through sequence analysis, more than 85% of the Z-DNA-induced large-scale deletions in mammalian cells were found to have 1-6 bp of microhomologies at their junctions, implicating a NHEJ-type mechanism in their processing (1). However, on the left panel of Figure 3, the 877 bp Z-DNA containing restriction fragment remained visibly intact for all Z-DNA-induced mutations generated in *E.coli*. The mutants generated in *E.coli* have small-scale deletions or expansions within the repeating CG units (1). Furthermore, we have evaluated these DSBs caused by Z-DNA using Ligation-mediated PCR (LM-PCR), as shown in Figure 4A, and the results showed that the DSBs were generated within and around Z-DNA-forming sequences in mammalian cells, supporting the speculation that Z-DNA-induced mutants were a result of DSB repair (1).

I.I.2 H-DNA

Homopurine:homopyrimidine regions in the genome that contain mirror repeat symmetry can transform a double-stranded B-DNA helix (duplex) into a three-stranded helix (triplex) with a complementary unpaired strand. The single strand from the duplex disassociates via energy provided by negative supercoiling and twists its backbone parallel inward to the adjacent strand of the underlying duplex (intra-), to form an intramolecular triplex known as H-DNA (33-35) (Figure 1C). The newly bound third strand in the triplex designates its classification by its direction, composition, and dependence of the pH. If the third strand is pyrimidine-rich and parallel to the duplex it will be classified as (Y*R:Y) or if the strand is purine-rich and anti-parallel to the duplex it will be classified as (R*R:Y).

Similar to the Z-DNA genome profile, there is an abundance of H-DNA-forming sequences in the mammalian genome, with the approximate estimate of one in every fifty-thousand bp in the human genome (36). Several disease-linked genes, such as the human oncogene *c-MYC*, have H-DNA-forming sequences located in promoter and/or exon/intron regions that are involved in regulating gene expression and co-mapped with breakage hotspots found in human diseases (37-39). Unlike Z-DNA, which was mutagenic in both mammalian cells and bacteria cells, we found that H-DNA-forming sequences were only mutagenic in mammalian cells but not in bacterial *E.coli* cells (1, 2) (Figure 2). Figure 2B shows the H-DNA-induced mutation frequencies from mammalian COS-7 cells. The different sequences used, include a sequence from the *c-MYC* promoter region that co-localizes with break hotspots (cMyc), its control derivative, MycAG, and three model H-DNA-forming sequences with the different degrees of ability to adopt H-DNA. GG32 is ranked as the highest in its ability to adopt an H-DNA structure, with GG32 having the greatest capacity to adopt H-DNA and GA32 having the lowest propensity to adopt H-DNA, and a control sequence CON. Each H-DNA-forming plasmid had a significant fold increase in its mutation frequency compared to the control plasmids. The characterization of the H-DNA-induced mutants generated in the mammalian cells indicated the same majority of large-scale deletions and microhomology footprint in the deletion junctions as the Z-DNA-induced mutants (1, 2). H-DNA also induces DSBs in mammalian cells, as demonstrated through LM-PCR analysis that showed the amplified regions between a primer and a DNA breakpoint in lanes 9-12 (Figure 4B shows the H-DNA LM-PCR results).

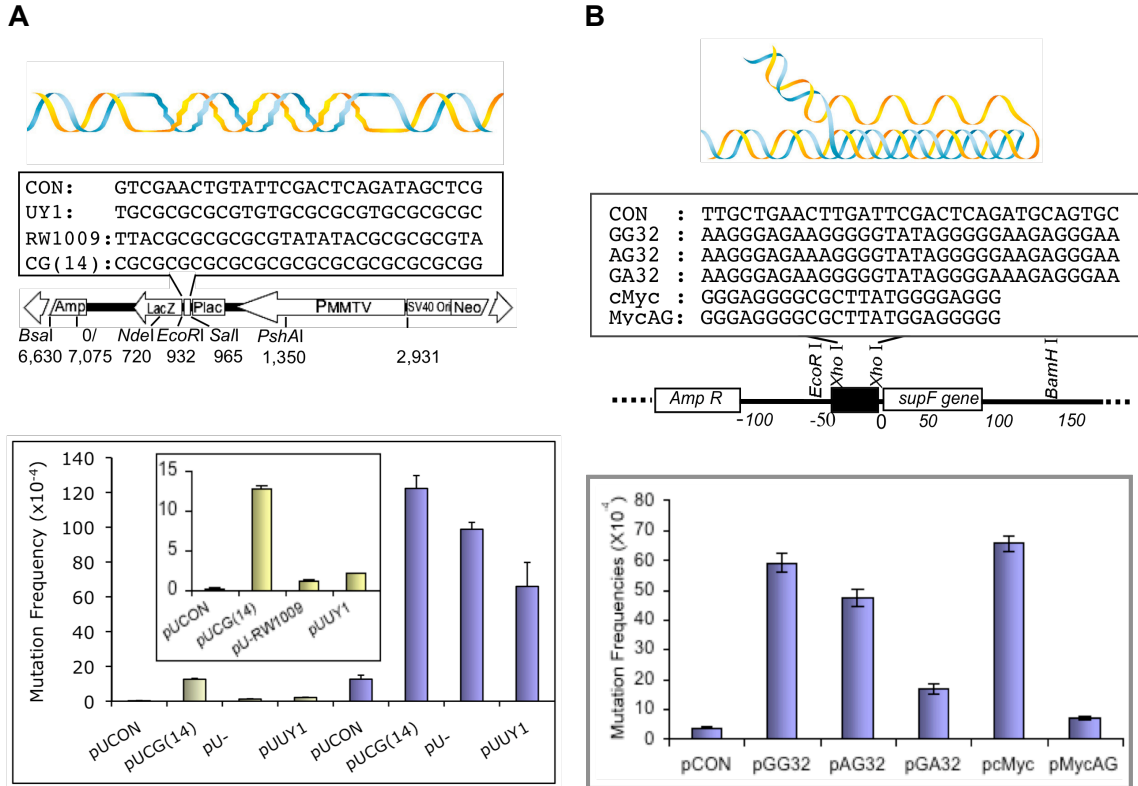


Figure 2. The mutation frequencies of the non-B DNA structures, Z-DNA and H-DNA. (A) The Z-DNA structure, and not hairpins or slipped structures, is responsible for the genetic instability in mammalian cells. The Z-DNA modeled structure, positions and sequences of the different inserts are shown above and were inserted in the *lacZ'* mutation reporter shuttle vector pUCNIM. The Z-DNA-induced *lacZ'* mutation frequencies from *E. coli* DH5 α cells (shown in the yellow columns) and from mammalian COS-7 cells (shown in the purple columns) are combined. The yellow columns of the bacterial mutation frequencies are magnified in the inserted panel. (B) The H-DNA-forming sequences that have different relative abilities to adopt the H-DNA structure are mutagenic in mammalian COS-7 cells. The H-DNA structure and sequences from the *c-MYC* gene, cMyc, its control derivative, MycAG, three model H-DNA-forming sequences with the ability to adopt H-DNA in the order of GG32>AG32>GA32, and the control sequence CON are shown above and were inserted in the *supF* mutation reporter gene in the shuttle vector pSP189. Mutation frequencies of these plasmids were calculated and graphed after their transfection into mammalian COS-7 cells. Error bars show the standard errors of the mean. (Adapted from Wang et al., 2006; and Wang and Vasquez, 2004)

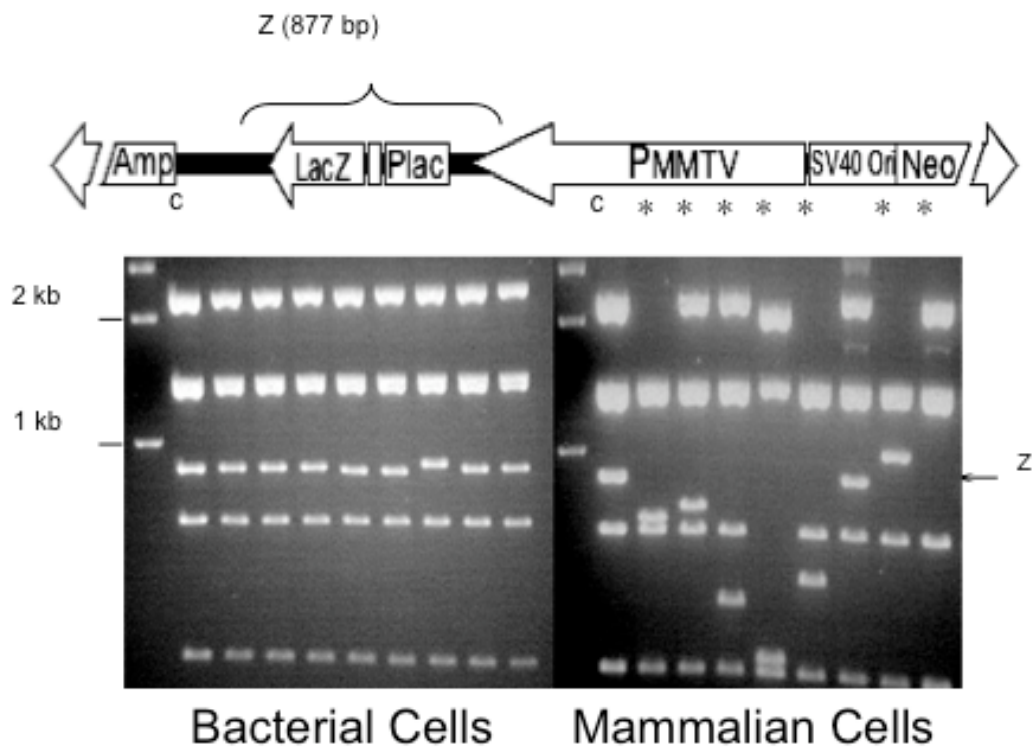


Figure 3. Z-DNA-induced mutations from two cell types, bacterial and mammalian cells. Different mutation types shown in the spectra of the pUCG(14)-induced *lacZ'* mutants from *E.coli* DH5 α cells (left panel) and mammalian COS-7 cells (right panel). The mutants were digested with EagI and BssSI and the “Z” marks the 877 bp Z-DNA fragment. The “*” marks the mutants with large-scale deletions (>50 bp), and the “c” labels the control plasmid. (Adapted from Wang et al., 2006)

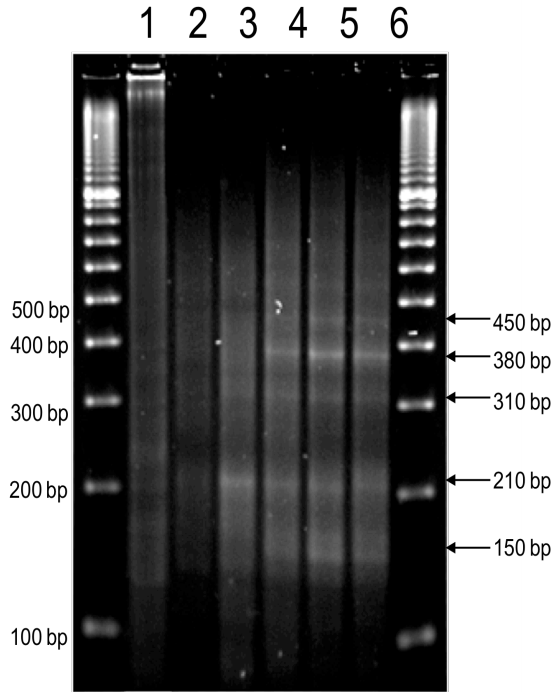
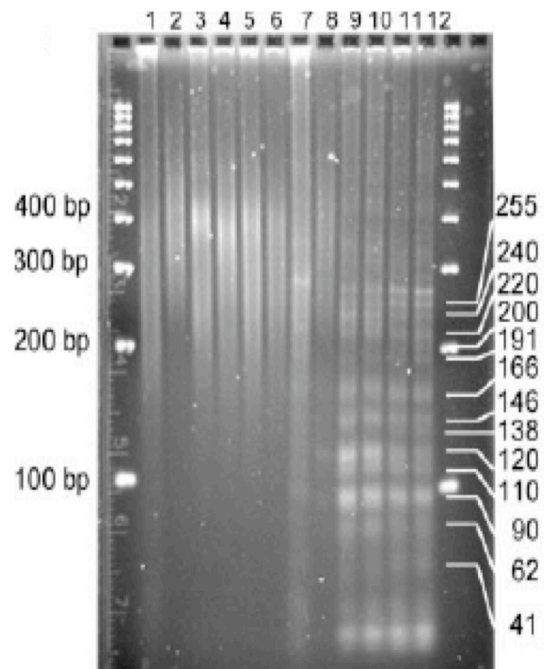
A**B**

Figure 4. LM-PCR results of both, Z-DNA and H-DNA. The lanes represent different time points, different cell types and different plasmids used. (A) The agarose gel electrophoresis shows the LM-PCR analysis of Z-DNA; with lane 1 and 2 coordinating to the results of the Z-DNA and CON plasmids recovered from incubation in *E.coli* cells and mammalian cells, respectively. In lanes 3 through 6 are the LM-PCR results of the Z-DNA plasmid incubated at increasing increments of time in mammalian COS-7 cells from 4 hours (hrs) to 48 hrs. (B) The agarose gel electrophoresis shows the LM-PCR analysis of H-DNA. Lanes 1-6 used a primer on the linker that amplified all the plasmid pieces, while lanes 7-12 used the specific and linker primer to amplify the specific breakpoints. The control plasmid CON was replicated in mammalian COS-7 cells for 48 hrs, shown in lanes 1 and 7, while the H-DNA-forming sequence, pcMYC, replicated for 0, 4, 8, 24 and 48 hrs in the other sequential lanes following the control lanes. (Adapted from Wang et al., 2006; and Wang and Vasquez, 2004)

I.2 DNA DOUBLE-STRAND BREAK REPAIR

Of all the potential types of DNA damage present in a cell, DSBs are one of the most dangerous types of lesions because of the detrimental effects the breakage can leave if the strands are left unrepaired. The cells can potentially face cell death and/or unfavorable genomic rearrangements that can initiate carcinogenesis (40, 41). On the upside, however, there are some benefits that pertain to having programmed DSBs. For instance, the creation of genetic diversification, the segregation of chromosomes in the course of meiosis for all eukaryotes, and the production of the immune diversification from V(D)J and class-switch recombination in vertebrates (42, 43). The two main characterized repair systems that process DSBs are the NHEJ pathway and the HR pathway.

In the NHEJ pathway, which predominates in G_0/G_1 cells where there is no homologous template, uses short (1-6 bp) microhomologies to repair DSBs and this usually results in loss of the sequence information, making the process error-prone (44). Thus, NHEJ is known as the more error-prone pathway, between the two DSB repair pathways, because of potential generation of small insertions and deletions, particularly in microhomologous regions (45, 46).

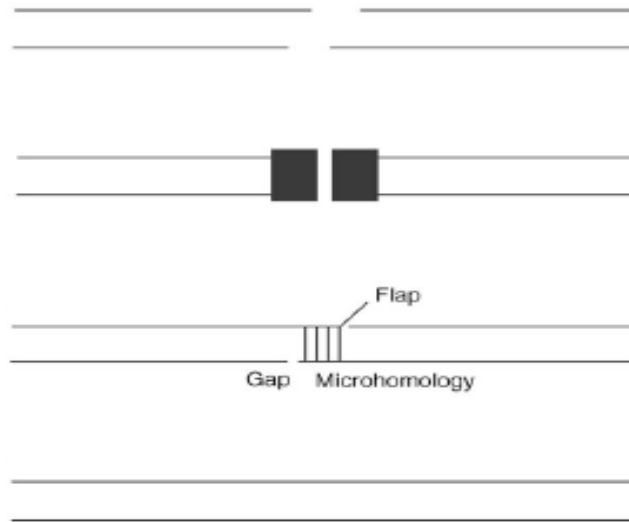
In the HR pathway, homologous templates are used in the repair of DSBs, requiring the assistance from an undamaged sister chromatid that has regions of DNA homology for the production of accurate or inaccurate repair products (47). These repaired products may or may not have crossover regions in its exchange with the homologous template; if a crossover does occur, the crossover may cause undesired genome rearrangements or a loss in heterozygosity (48). Generally, HR is considered as the more error-free pathway. However a loss of heterozygosity or genome rearrangements of tumor suppressor genes and proto-oncogenes can lead to their inactivation or activation, respectively, which can further lead to tumorigenesis (49). In addition, the occurrence of small misalignments within repetitive sequences during the repair process makes HR less than completely “error-free”.

Both DSB repair pathways are conserved from yeast to vertebrates. The details of both NHEJ and HR mechanisms will be discussed in greater detail, in reference to in mammalian and bacterial cells.

1.2.1 Non-homologous End-Joining in Mammalian Cells and Bacterial Cells

Non-homologous end-joining is the primary pathway for repairing DSBs in mammalian cells and is active throughout the cell cycle, especially in the cell cycle phase when a homologous sister chromatid is not available (e.g. G₁ and G₀) (45, 50-54). The NHEJ pathway is not present in all prokaryotes, such as *Escherichia coli*; however, a few prokaryotic species, such as *Mycobacterium tuberculosis*, *Mycobacterium smegmatis*, *Bacillus subtilis* and others, do contain a functional NHEJ mechanism (55).

In mammalian cells, after a DSB occurs, the Ku heterodimer (Ku70 and Ku80) binds to the DNA DSB ends and aligns the ends in a sequence- and overhang-independent manner with a high affinity (56-58). Once Ku is bound to the DNA ends, it recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and activates its kinase function (59). Together, the Ku and DNA-PKcs form the DNA-PK complex that autophosphorylates itself and other proteins, in order to recruit more repair factors such as the Artemis protein and DNA X family polymerases (e.g. Pol μ , λ , and terminal transferase (TdT)) that assist in the processing of the ends (60, 61). Eventually, XRCC4 (X-ray repair complementing defective repair in Chinese hamster cells 4) protein and Cernunnos-XLF (XRCC4-like factor) are recruited to the ends, which will stimulate DNA ligase IV and form an X4-L4 complex that completes repair with ligation of the DNA ends (62-65) (Figure 5). A simple NHEJ system can be seen in some prokaryotes, in which a two-component complex, such as the one in *Mycobacterium tuberculosis* with its Ku homodimer and DNA ligase D (LigD), can be sufficient for repair (3). The NHEJ system is unavailable (or yet to be identified) in *E.coli*, which uses HR instead of NHEJ to process DSBs. The prokaryotic Ku is a homodimeric quaternary structure (66) that lacks the eukaryotic von Wille brand factor A (vWA) domain of Ku70/Ku80 and also the SAP domain in the carboxy-terminal of the Ku70 heterodimer (67). LigD, a single polypeptide, has an ATP-dependent ligase (LIG) domain, a polymerase (POL) domain and a phosphoesterase (PE) domain that can perform multiple catalytic functions. It acts as a DNA-dependent DNA polymerase, a DNA-dependent RNA primase, a 3'-5' single stranded DNA exonuclease, a terminal transferase, and a DNA ligase (3, 68-71). Figure 5 shows a general schematic, where Ku would bind to the DSB ends, like in eukaryotes, and recruits LigD, which works to catalyze the processing of the ends and ligation of the break (3, 72).



<i>M. tuberculosis</i>	<i>S. cerevisiae</i>	<i>Homo sapiens</i>	Properties/function
Mt-Ku	Hdf1/Hdf2	Ku70/Ku80	Non-specific dsDNA end binding
Mt-Lig		DNA-PK _{cs}	Protein kinase
	Dnl4/Lif1	DNA ligase IV/XRCC4	Multifunctional protein with primase, nuclease, and ligase domains
	Rad50/Mre11/Xrs2	Rad50/Mre11/Nbs1	ATP-dependent DNA ligase
	Pol4	Pol μ	Mre11:3' \rightarrow 5' exonuclease, structure-specific endonuclease Rad50: ATP binding
	Rad27	Pol λ	DNA polymerase
		FEN1	DNA polymerase
			5' Flap endonuclease

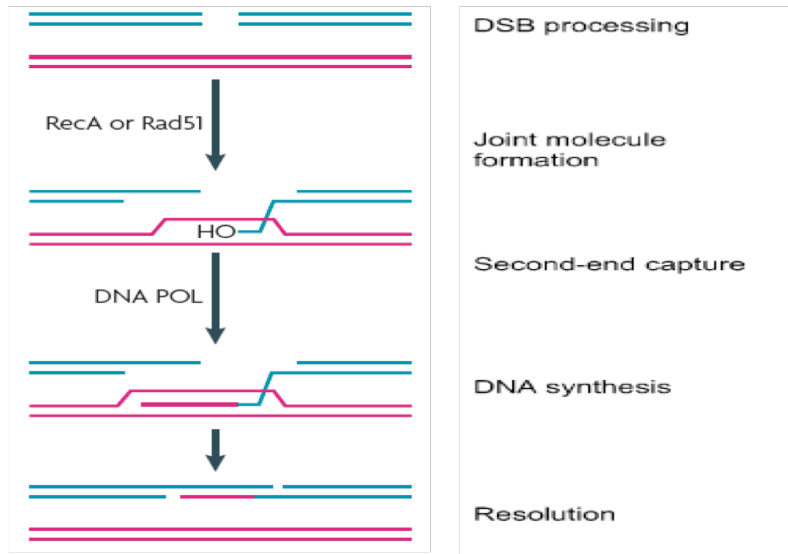
Figure 5. Model of the non-homologous end-joining pathway for mammalian cells and prokaryotes. In mammalian cells, the initiation of the NHEJ pathway begins with the heterodimer Ku70/80 binding to the DNA ends of a DSB. Ku then recruits DNA-PKcs to the DNA termini and the two DSB ends are brought together. Ku and DNA-PKcs form the DNA-PK complex that autophosphorylates itself, subsequently also triggering a phosphorylation cascade of other proteins for the recruitment of the Artemis protein and other end-processing factors, such as the DNA polymerase X, that will produce proper DNA ends required for the final resolution of repair. After being processed, the DSB ends are ligated by the X4-L4 complex with the XLF protein. In prokaryotes that have the NHEJ homologues, the initiation step starts when the homodimer Ku binds to the broken DNA ends. Then the recruitment of LigD occurs via its polymerase domain (PolDom) and this protein is sufficient for processing of the ends, ligation, and resolution of repair. (Adapted from Hefferin and Tomkinson, 2005)

I.2.2 Homologous Recombination in Mammalian Cells and Bacterial Cells

Homologous recombination processes and repairs DSBs differently than the NHEJ pathway, starting with the use of an homologous template and followed by a set of different components in mammalian cells versus bacterial cells. HR usually occurs in the S and G₂ phase of the mammalian cell cycle, when a sister chromatid is available (45, 53, 54). The general schematics of the HR mechanism in mammalian cells and bacterial cells are similar, as shown in Figure 6. However, there are different factors or multiple factors involved in each step that will be described below.

In the HR pathway, two broken DSB ends are not simply rejoined, as it could be done in the NHEJ pathway. Therefore, in mammalian cells, the Mre11-Rad50-Nbs1 (MRN) complex, with the help of the CtIP-BRCA1-BARD1 complex, resects the 5' ends of the DSB DNA ends to yield 3'-OH single strand DNA (ssDNA) overhangs (73, 74). The RPA protein coats the ssDNA. DNA-damage checkpoint responses and several other recombination mediators are initiated, including the Rad52 group (Rad50, Rad52, Rad54 and the Rad51 paralogs) and BRCA2, to assist Rad51 nucleoprotein filament assembly (75-79). The Rad51 nucleoprotein filament and its interacting partners catalyze the homology search and strand invasion reaction (47). The homologous template is used for DNA synthesis and the missing break region is copied by DNA polymerases. A three-way junction, known as a D-loop, and holiday junctions are formed during this process. These intermediates are resolved with resolvases that produce crossover or non-crossover products (80-83). To finish, DNA ligase ligates the strands for resolution of repair (73, 74).

In *E.coli*, the RecBCD complex binds to the broken DNA ends to serve as the helicase and nuclease that produces the resected 3'-OH ssDNA ends (74, 84-86). Following the resection process, RecBCD is responsible for loading the RecA protein onto the DNA, which is a homolog to the mammalian Rad51 protein (74, 84-86). In the alternate RecFOR pathway, RecQ and RecJ factors can be interchangeable with the RecBCD complex (87-90). Once the RecA protein is loaded onto the 3'-OH single-stranded DNA, it will catalyze the homology search and strand invasion reactions (91). Branch migration is facilitated by the RuvAB complex and DNA synthesis occurs with the formation of intermediates structures (87). Resolutions of these junctions are carried out by the addition of the RuvC protein to the RuvAB complex and the final step, the ligation reaction, is done by Ligase A (47).



Processes mediated	Proteins	
	<i>E. coli</i>	Eukaryotes ^a
Endprocessing	RecBCD, RecQ, RecJ	Mre11 complex
Negotiating single-stranded DNA	SSB	RPA
Recombinase loading	RecBCD, RecFOR	Rad52, Rad51 paralogs
Recombinase nucleo-protein filament stabilization		Rad54
Joint molecule formation by strand invasion	RecA	Rad51, Rad54
Branch migration	RuvAB	Rad54
Resolution of crossed DNA strands	RuvC	Mus81/Eme1, Rad51C/Xrcc3

Figure 6. Model of the homologous recombination pathway for mammalian cells and prokaryotes. In mammalian cells, the initiation of the HR pathway begins with resection of the DNA ends at DSBs by the MRN complex with the help of the CtIP-BRCA1-BARD1 complex. After the generation of 3'-OH single-stranded DNA, RPA coats those strands and then Rad52 or BRCA2 loads Rad51 to catalyze the strand invasion process into the sister chromatid that serves as the intact homologous template for resynthesis. DNA polymerase (POL) copies the sister chromatid over the break. Different recombination intermediates, such as Holiday junctions, are then resolved with resolvases that will produce crossover or non-crossover products. DNA ligase seals these products for completion of repair. Some of these proteins may have multiple roles in HR. Proteins with similar functions in *E.coli* as in eukaryotes are listed in the table above. LigaseA seals final nicks and completes repair, resulting in crossover or non-crossover products. (Adapted from Shuman and Glickman, 2007; and Wyman et al., 2004)

I.3 HYPOTHESIS AND SPECIFIC AIMS

Many types of DNA damaging agents from both exogenous and endogenous sources can lead to genomic instability. These lesions are generally accounted and compensated for by a range of available DNA repair systems in cells. Intriguingly, in the absence of an exogenous DNA damaging source, studies have shown that naturally occurring non-B DNA structures can induce genomic instability in mammalian cells and bacterial cells (1, 2, 8). Two of these non-B DNA structures studied in my thesis work are Z-DNA and H-DNA (see Figure 1) (10).

Our laboratory has previously found that Z-DNA and H-DNA-forming sequences induce higher frequencies of mutagenesis than control B-DNA-forming sequences. The majority of the mutations induced by H-DNA and Z-DNA were large-scale deletions with microhomologies at their junctions in mammalian cells (1, 2). However, in bacterial *E.coli* cells, the Z-DNA forming CG repeat-induced mutants had small-scale deletions within the repetitive sequence and the H-DNA structure was found to be quite stable in bacteria (1, 2). Both non-B DNA structures caused DSBs in mammalian cells, while the situation in bacterial cells was quite different. These data suggested that the non-B DNA-forming sequences may be processed differently in mammalian cells than in bacterial cells, but the specific mechanism(s) for this difference is not clear. I hypothesize that the error-prone NHEJ repair of DSBs generated surrounding Z- and H-DNA structures is responsible for the large-scale deletions and rearrangements in mammalian cells. Further, I speculate that the difference between mammalian cells and bacterial cells, in relation to the Z-DNA and H-DNA-induced genomic instability, could be due to the different DSB repair systems available. Mammalian cells use the error-prone NHEJ as the predominant DSB repair system and bacterial *E.coli* cells predominantly use the less mutagenic HR pathway available for repairing its DSBs, since NHEJ is not available (92-95). The following are my specific aims for my research studies:

Specific Aim 1: To study the roles of NHEJ and HR on Z-DNA-induced genetic instability by determining the Z-DNA-induced mutation frequencies and spectra in modified *E.coli* strains with proficiencies and/or deficiencies in NHEJ (Ku/LigD) and HR.

Specific Aim 2: To study the roles of NHEJ and HR on H-DNA-induced genetic instability by determining the H-DNA-induced mutation frequencies and spectra in modified *E.coli* strains with proficiencies and/or deficiencies in NHEJ (Ku/LigD) and HR.

CHAPTER II: DOUBLE-STRAND BREAK REPAIR PATHWAYS IN Z-DNA AND H-DNA-INDUCED GENOMIC INSTABILITY

II.I INTRODUCTION

After the introduction of the canonical B-DNA double helix by Watson and Crick (5), the scientific door remained opened for the discoveries of many other types of DNA structures, otherwise known as non-B DNA structures. In the absence of an exogenous or endogenous source of DNA damage, these non-B DNA structures can induce genomic instability *in vitro* and *in vivo* (10, 31). Moreover, non-B DNA structures can be implicated as causative factors for mutagenesis and human diseases (11). Two non-B DNA structures of interest in the study of non-B DNA-induced genomic instability are Z-DNA and H-DNA conformations because of their frequency in the mammalian genome and association with human disease.

Sequences that can adopt Z-DNA or a H-DNA structures consist of repeats of alternating purine-pyrimidine or symmetrical homopurine:homopyrimidine mirror repeats, respectively. Shown in prior studies from our laboratory, Z-DNA and H-DNA-forming sequences cause a defined division in the mutation frequencies and spectra between mammalian cells versus bacterial cells (1, 2). Z-DNA induced a *lacZ'* mutation frequency of $\sim 120 \times 10^{-4}$ in mammalian cells, whereas in bacterial cells, a much lower mutation frequency of $\sim 12 \times 10^{-4}$ was found for the same sequence (1). In parallel, the mutation frequencies of some H-DNA-forming sequences from mammalian cells were much higher than the nearly undetectable bacterial mutation frequencies (2). The difference widened between the mammalian and bacterial cells when large-scale deletions were found in the mutants generated in mammalian COS-7 cells and mouse chromosomes, while in bacterial *E.coli* cells, H-DNA-forming sequences induced no significant mutants and the Z-DNA-forming CG(14) sequence only induced small expansions or contractions within the repetitive units of the sequence (1, 2, 31).

There may be a number of differences between mammalian cells and bacterial cells that could account for these interesting differences seen in the non-B DNA-induced mutation frequencies and spectra, such as different non-B DNA secondary structure formation between species. Additionally, the non-B DNA structures may be recognized and/or processed differently in mammalian cells than in bacterial cells. There is also a difference in the chromatin organization within both kinds of cell types, and the different types of DNA binding proteins may have a role in the species-specific outcomes seen in our previous results. For instance, the bacterial HR RecA protein binds preferably to Z-DNA structures than to B-DNA structures (21), while the homolog

of RecA, the mammalian Rad51, has a lower preference for Z-DNA (96). Specifically, the affinity of Rad51 for binding to ssDNA, being an ATPase and fulfilling the strand exchange reaction *in vitro*, is lower than that of RecA (96). Nonetheless, the most probable cause for the difference in the non-B DNA-induced mutagenesis in mammalian cells versus bacterial cells is in how the non-B DNA-induced DSBs are processed and repaired via their available DSB repair systems.

Throughout the mammalian cell cycle, except for the S and G2-phases, the dominating DSB repair system is the NHEJ repair pathway, which is often error-prone in losing the regions between microhomologous sequences (97, 98). While in the S and G2-phases, the less error-prone HR system is competitively utilized for repair of DSBs (45, 53, 92, 99, 100). In contrast, prokaryotes like *E.coli* that are lacking NHEJ use the relatively less mutagenic RecA-dependent HR pathway as the primary mechanism for repairing DSBs. However, during the strand invasion process in HR, when the ssDNA binds with the sister homologous template, there can be small misalignments that occurs in repetitive sequences (101). Considering the available DSB repair pathways within mammalian cell versus *E.coli* or rather a lack of multiple pathways in the case of *E.coli* cells, the differences in frequencies and types of Z-DNA and H-DNA-induced mutations between mammalian cells and *E.coli* may be because of their different available DSB repair pathway(s).

In order to study the roles of the DSB repair pathways, NHEJ and HR, in Z-DNA and H-DNA structure-induced genetic instability and to test my hypothesis, *E.coli* cells were modified to express the *Mycobacterium tuberculosis* (Mt) NHEJ system, which consisted of the Mt-Ku and Mt-LigD proteins (3). The developed engineered strains were either proficient or deficient in HR (RecA/RecB) and NHEJ (Mt-Ku/Mt-LigD), as listed in Table 1. The stability of the Z-DNA-forming CG(14) repeat and H-DNA-forming sequences (oriented in two directions, U and Y) were determined in the newly contrived inducible Mt-NHEJ setting, which mimicked the DSB repair setting in mammalian cells. Using a reporter system described in Wang et al. (1) with a blue/white mutation screening method, where the white colonies represent mutants of the *lacZ'* gene and the blue colonies represent the wild-type *lacZ'* function and production of complete β -galactosidase, the mutation frequencies were calculated and graphed, followed by characterization of selected mutants and sequencing analyses.

II.2 MATERIALS AND METHODS

Plasmid substrates

The Z-DNA-forming sequence made up of CG(14) repeats: CGCGCGCGCGCGCGCGCGCGCGCGCGCGCG, the H-DNA-forming sequence from the promoter region of the human *c-MYC* gene in the U direction, where there are purines in the coding strand: CGAGCTCCCTCCCCATAAGCGCCCCTCCC, the human *c-MYC* gene in the Y direction, where there are pyrimidines in the coding strand: GCGGGGAGGGGCGCTTATGGGGAGGGTTG, and a control scrambled sequence CON: CGAGCTATCTGAGTCGAATACAGTTCGAC, were used in this study and in previous studies by Wang et al. (1, 2). Each plasmid was constructed based on a 7,075 bp shuttle vector, pUCNIM, which has many essential components, including two antibiotic resistance genes: neomycin and ampicillin, an SV40 replication origin, a bacterial replication origin, and the insert sequence (non-B DNA-forming sequence or control sequence) placed within the *lacZ'* gene that expresses the amino-terminal fragment of the mutation-reporter gene, β -galactosidase (1). Oligonucleotides were purchased from the Midland Certified Reagent Co (Midland, TX), to allow for cloning of the inserts of interest into the pUCNIM shuttle vector plasmid at the EcoRI-Sall cassette, in the region between the promoter and the *lacZ'* coding region. The plasmids were named accordingly: pUCG14, pUMYCu, pUMYCy and pUCON. Linearized pUCON and pUCG14 digested with EcoRI were used to observe the DSB repair efficiency after expression of the Mt-NHEJ vectors.

Bacterial strains

The different bacterial *E.coli* strains used in this study are listed in Table 1, including each strain's genetic background with their corresponding proficiencies and/or deficiencies in HR (RecA/RecB) and NHEJ (Ku/LigD) after L-arabinose induction (Ara+); highlighted in red are the deficiencies. All bacterial strains were obtained from Dr. Lynn Harrison's lab (Louisiana Health Sciences Center) and they were constructed from the parental wild-type (WT) *E.coli* (Hfr KL16 (PO-45) thi-1 relA1 spoT1e14 λ -) strain. *M.tuberculosis* NHEJ repair proteins, Ku (Mt-Ku) and DNA Ligase D (Mt-LigD), expression vectors were constructed and integrated into the bacterial strains as

described by Malyarchuk et al. (55). The Mt-Ku and Mt-LigD expression vectors were tagged at the N-terminus with a His-tag that did not distort their activity and allowed for western analysis to identify protein expression (55). The expression of the Mt-NHEJ proteins, Mt-Ku and Mt-LigD, are induced with L-arabinose (i.e. functional NHEJ). Every strain contains one or both of the Mt-NHEJ components, except for the WT strain that has neither. Two isogenic strains have deficiencies in HR: the RecA- and RecB- strains, with mutations in *recA* (KL16 *recA56 srlC300::Tn10*) and *recB* (*recB268::Tn10*), respectively. The WT/NHEJ+ strain has both the wild-type *E.coli* HR repair system and the Mt-NHEJ repair system, after L-arabinose induction. Mt-Ku or Mt-LigD proteins were also expressed individually in wild-type *E.coli* cells and were labeled as the Mt-Ku only and Mt-LigD only strains, accordingly.

Preparing NHEJ proficient or deficient competent *E.coli* cells

After overnight growth of the desired bacterial strains in 150 ml LB culture with appropriate antibiotics at 37°C and shaking at 250 rpm, 1.5 ml of the culture were transferred and split into two fresh 150 ml LB for a 100-fold dilution. One flask was labeled as Ara+ and the other as Ara-, no antibiotics were added and diluted cultures were placed in 37°C incubation and 250 rpm shaking for one hour. After one hour of incubation, 3 ml of 10% filtered L-arabinose was added to the Ara+ labeled culture, bringing the L-arabinose concentration to a final concentration of 0.2% to induce the expression of the Mt-NHEJ genes, *Mt-Ku* and/or *Mt-LigD*. Another additional 2 hours of continual growth occurred before the bacteria reached the exponential phase with an OD600 of approximately 0.6, and were harvested by centrifugation in 30 ml tubes at 4500 rpm and at 4°C for 20 minutes. Two sequential 20 minutes washes and centrifugation were done with 25 ml of detergent free ddH₂O, after removing the supernatant each time. In the final step, the pellets were transferred to 1.5 ml microcentrifuge tubes and re-suspended in 10% glycerol for a final centrifugation at 4000 rpm at 4°C for 20 minutes. After removal of the supernatant, the cells were aliquoted into 1.5 ml tubes for storage at -80°C, until proper use. These competent cells were placed on ice when ready for western analysis or the electroporation step in the mutagenesis assay.

Western Analysis

Western analyses were performed with modification as described in Malyarchuk et al. to verify the expression of Mt-Ku and Mt-LigD proteins (55). After placing the competent cells on ice, 20 μ l of bacteria were re-suspended in 500 μ l of 50 mM Tris-Cl (pH 7.4) and sonicated prior to centrifugation at 4°C at 12,000 rpm for 5 minutes. After the removal of the supernatant, pellets were re-suspended by vortexing in 25 μ l H₂O and 25 μ l of 2 x SDS with dithiothreitol (DTT) added to room temperature 3 x SDS stock. The samples were heated in a boiling 100°C water bath for 5 minutes and centrifuged for 10 minutes, 12,000 rpm at room temperature. Fifty μ l of the samples were then loaded onto a ready-made 4–20% gradient SDS polyacrylamide gel for fractionation via electrophoresis at 100 Volts (V) for 2 hrs and transferred onto a 0.2 μ m nitrocellulose membrane by electroblotting at 650 milliamps (mA) for 1.5-2 hrs (Bio-Rad Laboratories, Inc, Hercules, CA). The blots were washed in tris-buffered saline (TBS) with 0.1% Tween 20 (TBS/T) for one minute, prior to being blocked in 5% nonfat milk (TBS/T) for one hour. The membrane was probed with a His-Tag (27E8) mouse monoclonal antibody in the 5 % nonfat milk (TBS/T) solution (diluted 1:1000; Cell Signaling Technology, Inc, Boston, MA) overnight with gentle shaking at 4°C. This antibody recognizes six consecutive histidines on the tagged proteins. On the next day, these blots were washed three times for 10 minutes each in TBS/T. Bound His-Tag antibody was detected using chemiluminescent ECL detection reagents with the incubation of a secondary horseradish peroxidase anti-mouse antibody (diluted 1:3000; Amersham, Milano, Italy).

Mutagenesis assay

Two hours prior to the cells becoming competent and one hour after transformation, all of the Mt-NHEJ expressing bacterial strains were exposed to a final concentration of 0.2% L-arabinose added to the medium to induce and sustain expression of Mt-Ku and/or Mt-LigD proteins. Fifty ng of all the plasmids (pUCG14, pUMYCu, pUMYCy and pUCON) were transfected into 40 μ l of the modified or wild-type bacterial *E.coli* cells by electroporation with the use of the Bio-Rad Gene Pulser II (Bio-Rad, Hercules, CA) at 1.7 kV per reaction and were then resuspended in 1 ml LB. After a one-hour recovery period, 200 μ l was transferred to fresh 5 ml LB. The bacterial strains that were designated for Mt-NHEJ expression, either for expression of both proteins or individual components, were labeled as Ara+ with 100 μ l of 10% L-arabinose

added, to bring to a final concentration of 0.2% L-arabinose, for the induction of the gene(s) expression in these strains. The cultures were grown overnight in the LB mixture with ampicillin (Amp; 100 μ g/ml) and kanamycin (Kan; 50 μ g/ml), at 37°C incubation and 250 rpm shaking. After approximately 16 hours of overnight incubation, the Qiagen QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA) was used according to the manufacturer's recommendations for the isolation of the plasmids. A second transformation was done with 10 ng of the recovered plasmid DNA (at concentrations of 10 ng/ μ l) and 20 μ l of commercial DH5 α -derived cells, NEB 5-alpha Electrocompetent *E.coli* (New England Biolabs, Ipswich, MA) diluted 1:1 in 10% glycerol. The bacteria were grown on LB-ampicillin-kanamycin (same concentrations as described above) plates with 5-bromo-4-chloro-3-indolyl β -D-galactoside (X-gal) (50 mg/ml) and isopropyl β -D-thiogalactoside (IPTG) (160 mg/ml) for overnight in a 37°C incubation oven. This allows for a blue/white screening of the *lacZ'* mutants generated in the Ara- or Ara+ bacterial strains after the first transformation. The number of white (mutant) colonies were counted and divided over the number of total (blue and white) colonies to determine the mutation frequencies. A pairwise t-test was performed as the statistical analysis test on the collected data.

For further analysis, restriction analysis was performed, where selected mutants were individually picked, streaked on another LB/Amp/Kan/X-Gal/IPTG agarose plate to verify mutagenesis, and then grown overnight in 2 ml LB-ampicillin-kanamycin liquid medium. As previously described, the Qiagen QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA) was used to recover the plasmid DNA of the mutants. A double digestion was performed with EagI and BssSI enzymes on the spontaneous or non-B DNA-induced mutant plasmid DNA and the stock control plasmid DNA were separated via electrophoresis on 1.4% agarose gels, which will result in seven fragments from the wild-type plasmid and the restriction fragments with 877 bp for the non-B DNA fragment (Figs. 9B, 10B, 10D, 11B, 12B, 13B, 15B and 15D). Mutant DNA was sent to the MB Core facility (UTMDACC at Science Park) for sequencing to verify mutant sequences and to analyze deletion junctions. The total number of large-scale deletions from all Mt-NHEJ expressing strains was calculated into a percentage of large deletions to total deletions for all the bacteria expressing Mt-NHEJ (Ara+) versus when Mt-NHEJ proteins are not expressed (Ara-). The difference between the two groups was statistically analyzed with a Fisher exact test.

	HR RecA/RecB	Mt-NHEJ Ku/LigD (Ara+)
WT	+/+	-/-
RecA-	-/+	+/+
RecB-	+/-	+/+
WT/NHEJ+	+/+	+/+
Mt-Ku only	+/+	+/-
Mt-LigD only	+/+	-/+

Table 1. Bacterial strains used in this study with their proficiencies and deficiencies in HR (RecA/RecB) and Mt-NHEJ (Ku/LigD). The “+” signifies that the corresponding protein was present, and likewise, “-“ signifies that the corresponding protein was not present. (Adapted from Kha et al., 2010)

II.3 RESULTS AND DISCUSSION

Mt-NHEJ genes expression and mutagenic repair of DSBs in *E.coli*

Integration of the L-arabinose inducible Mt-Ku and Mt-LigD expression vectors into *E.coli* chromosomes was performed in Dr. Lynn Harrison's laboratory and the modified *E.coli* cells with Mt-NHEJ function were tested for the rejoining capabilities on transformed linearized plasmids with various overhangs (55). Table 1 lists the generated *E.coli* strains used with their corresponding HR and NHEJ capabilities noted in the "+" or "-" symbols. The wild-type (WT) strain is categorized as fully HR functional with unaltered RecA and RecB proteins, while this strain is naturally lacking in NHEJ function. The wild-type strain engineered to contain the Mt-NHEJ expression vectors, Mt-Ku and Mt-LigD, is listed as the WT/NHEJ+ strain. Individual expression of Mt-Ku or Mt-LigD into wild-type *E.coli* cells are referred to as the Mt-Ku only and Mt-LigD only strains. The two strains that are deficient in wild-type HR function, but have NHEJ capabilities, are referred to as the RecA- or RecB- strains; with alterations in their *recA* or *recB* genes, respectively.

Western analyses were conducted to verify the levels of Mt-Ku and Mt-LigD protein after their induction with L-arabinose (Ara+) as shown in Figure 7A. 'BW35' does not contain Mt-Ku nor Mt-LigD expression vectors, which is the same as the WT strain in Table 1. 'BWKu' and 'BWLig,' contain individual expression vectors with only Mt-Ku or only Mt-LigD being expressed when L-arabinose was supplemented into the LB culture. The anti-his tag monoclonal antibody recognized the Mt-Ku protein, which is of ~35 kDa, while Mt-LigD is detected at ~90 kDa (55). Lastly, 'BWKu/Lig#2' represents the strain that has both Mt-Ku and Mt-LigD expression vectors integrated into the *E.coli* chromosomes, as seen in the Ara+ column with the two bands at the appropriate sizes. Similar to the previous published western analyses, two bands at the ~35 kDa and ~90 kDa mark would also be reflected in our western blots for the RecA-, RecB- and WT/NHEJ+ strains because they contain both of the Mt-NHEJ proteins after L-arabinose is added (Ara+ columns in Figure 7B-D). The WT strain would show no bands, with or without L-arabinose (boxed in red on Figure 7B), and the individual Mt-Ku only and Mt-LigD only strains would only contain one band after induction of their individual protein (Ara+ columns in Figure 8B). Thus, in Figure 7B-D and Figure 8B, these western blots verify the correct expression of the NHEJ genes, according to the

strains listed in Table 1 with their matching labeled Mt-NHEJ descriptions after L-arabinose induction (Ara+). And with western analyses verifying Mt-NHEJ induction, studies on Z-DNA and H-DNA-forming sequences can be conducted in all the strains to determine the effects of NHEJ and HR on non-B DNA genomic instability.

The addition of L-arabinose to the LB culture for induction of Mt-NHEJ gene expression does not interfere *per se* with the stability of the *lacZ'* reporter gene on neither the control plasmid nor on the non-B DNA-forming plasmids in the WT *E.coli* strain that does not have integration of the Mt-NHEJ expression vectors (Figure 9A). The high mutation frequencies of the Z-DNA-forming sequence, before and after addition of L-arabinose, further confirm previous published results about the mutagenic potential of Z-DNA in wild-type bacterial cells (1). Z-DNA is still measurably mutagenic in the WT cells with mutation frequencies between 14 to 17 $\times 10^{-4}$, which is ~20-fold higher than the mutation frequencies of the control plasmid pUCON that are around $\sim 1 \times 10^{-4}$ with or without L-arabinose (Figure 9A). As expected, regardless of the presence of L-arabinose, the restriction analyses of the Z-DNA-induced mutants revealed that the major event was small-scale deletions within the CG repeats, as indicated by the minor size change of the restriction fragment that contains the Z-DNA-forming CG(14) sequence, and further confirmed by sequence analysis (3 large-scale deletion mutants/20 total mutants; Figure 9B).

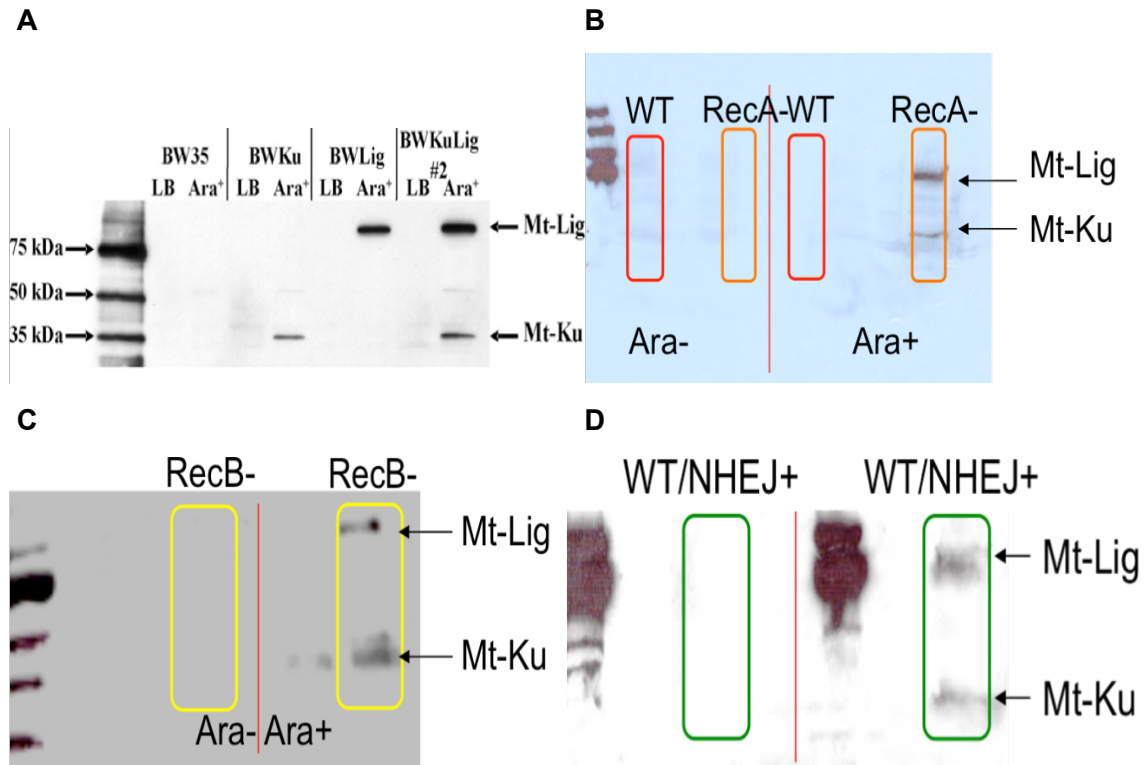


Figure 7. Western analyses to verify the joint expression of the Mt-NHEJ proteins, Mt-Ku and Mt-LigD, from different bacterial strains used in the study. The cultures were grown in LB or LB with a final concentration of 0.2% L-arabinose (Ara+) and a His-Tag monoclonal antibody was used on the 0.2 μ m nitrocellulose membranes to probe and recognize proteins containing five consecutive histidines that are tagged on the Mt-NHEJ proteins. (A) The wild-type (BW35) strain, the strain with only the Mt-Ku expression vector (BWKu), the strain with only the Mt-LigD expression vector (BWLig) and the strain that contains both Mt-NHEJ protein expression vectors (BWKuLig#2) were examined for protein(s) expression. (Adapted from Malyarchuk et al., 2007) (B) There was no expression of the Mt-NHEJ proteins detected in the wild-type (WT) cells that do not have the expression vectors, while in the RecA- cells that contain the Mt-NHEJ expression vector, Mt-NHEJ was present after Ara+ induction versus no expression in the Ara- cells. (C) Mt-NHEJ proteins are expressed in RecB- cells that contain the Mt-NHEJ expression vector, after Ara+ induction versus no expression in the Ara- cells. (D) In the WT/NHEJ+ cells, which are the WT cells with the Mt-NHEJ expression vectors, there was expression of both the Mt-NHEJ proteins after Ara+ induction versus no expression in the Ara- cells.

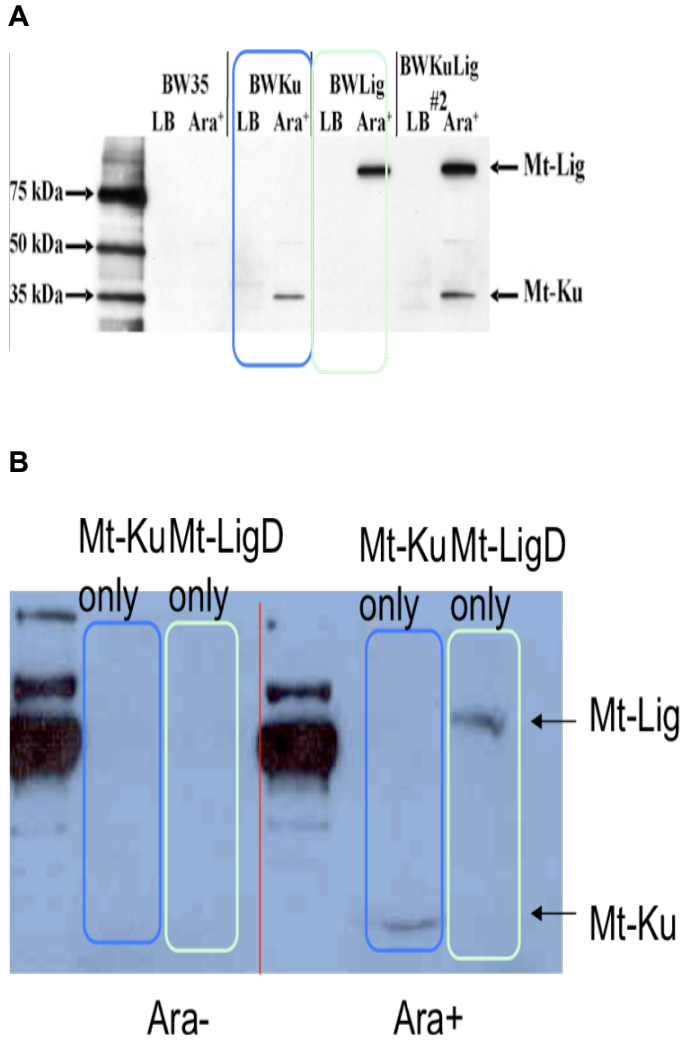
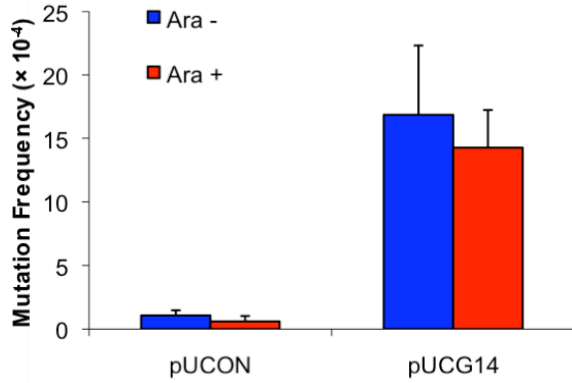


Figure 8. Western analyses to verify individual expression of the Mt-NHEJ proteins, Mt-Ku or Mt-LigD, in the Mt-Ku only and Mt-LigD only strains. The cultures were grown in LB or LB with a final concentration of 0.2% L-arabinose (Ara⁺) and a His-Tag monoclonal antibody was used on the 0.2- μ m nitrocellulose membranes to probe and recognize proteins containing five consecutive histidines that are tagged on the Mt-NHEJ proteins. (A) The protein expression examined in the strain with only the Mt-Ku expression vector (BWKu), boxed in blue, and the strain with only the Mt-LigD expression vector (BWLig) boxed in green. (Adapted from Malyarchuk et al., 2007) (B) There was Mt-Ku protein expression in Mt-Ku only cells (blue box) and Mt-LigD protein expression in Mt-LigD only cells (green box) after Ara⁺ induction versus no induction in the Ara⁻ cells.

A



B

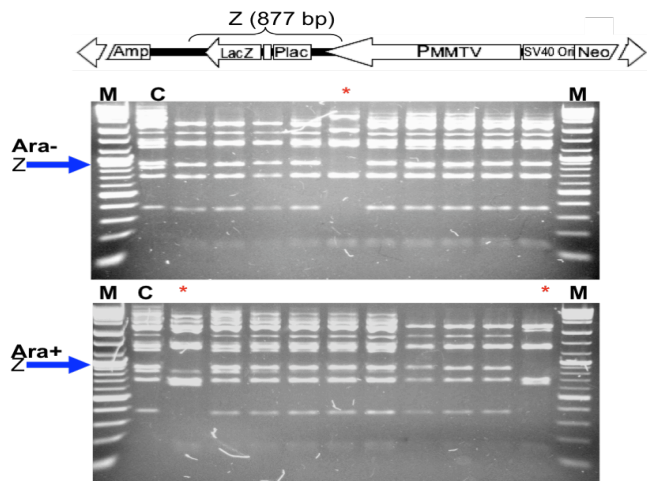


Figure 9. The inducible factor, L-

arabinose, is not mutagenic in the experimental mutation-reporter system *per se*.

(A) L-arabinose (Ara+), the inducible factor of the Mt-NHEJ proteins, did not affect the mutation frequencies of the control plasmid and Z-DNA plasmid in wild-type (WT) *E.coli* cells that do not have the Mt-NHEJ expressing system. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments. (B) The mutation spectra is shown on an agarose gel, a representative gel for the 20 total Z-DNA-induced mutants analyzed from WT *E.coli* cells without the Mt-NHEJ proteins expression vectors, before (Ara-, top panel) and after L-arabinose supplementation (Ara+, bottom panel). The arrows refer to the Z-DNA-forming fragments; “*” marks the large-scale deletions (≥50 bp); “C” labels the control plasmid; and “M” labels the size standard marker. (Adapted from Kha et al., 2010)

Previously, an end-joining assay was used to demonstrate that the expression of Mt-Ku and Mt-LigD in *E.coli* can effectively rejoin linearized plasmids that have 2 bp overhangs in a RecA and RecB-independent manner (55). Our blue/white mutation screening can also be applied to see if functional NHEJ is actually exhibited in these modified *E.coli* cells. After EcoRI digestion, linearized pUCON and pUCG14 plasmids that have “sticky-ends” were transformed into the RecA- and RecB- strains that contain an inducible Mt-NHEJ system to detect the frequency of inaccurate NHEJ repaired plasmids (the white-colored mutant colonies) over total repaired plasmids (mutant white-colored and wild-type blue-colored colonies) recovered from the strains when there is NHEJ (NHEJ+) or when the Mt-NHEJ genes are not expressed (NHEJ-; without L-arabinose induction; Figure 10A and 10C).

When there was expression of the Mt-NHEJ proteins, there were approximately 20-fold more colonies observed than in the un-induced cells. This indicated that the HR repair mechanism was substantially reduced in the RecA- and RecB-deficient strains and the transfected linearized plasmids were more receptive to repair when the NHEJ mechanism was available, upon L-arabinose induction. Figure 10A and 10C shows the very low mutation frequencies generated from the linearized plasmids in the NHEJ-groups, illustrating that there might be low amounts of natural rejoining of the linearized plasmids or other repair mechanisms that differ from NHEJ and traditional HR repair that requires both RecA and RecB. When Mt-NHEJ proteins were present in both RecA- and RecB- strains, there were ~1,000-fold increases in the mutation frequencies of both pUCON and pUCG14-linearized plasmids, confirming that the expression of Mt-Ku and Mt-LigD was sufficient for processing DSBs in a mutagenic fashion. This is consistent with the predominately inaccurate end-joining products previously seen, and concordantly, to the repair products of the NHEJ mechanism (55).

Moreover, the mutants from linearized plasmids recovered from the NHEJ+ cells had a majority of large-scale deletions as assessed by restriction enzyme and gel electrophoresis analyses (14/22; Figure 10B and 10D). The RecB- strain was also subjected to transformation with “blunt-ended” linearized plasmids that contain a DSB placed four bp from the *lacZ'* reporter gene after an EcoICRI digestion, and the repair of this kind of DSB was found to be more efficient in NHEJ+ cells, as indicated by ~10 to 30-fold more colonies than in the NHEJ- cells and the vast majority of the repair products were mutants (data not shown). These results verify that the two-component

Mt-NHEJ proteins were effectively expressed and functional in repairing DSBs *in vivo* in a mutagenic fashion in our system.

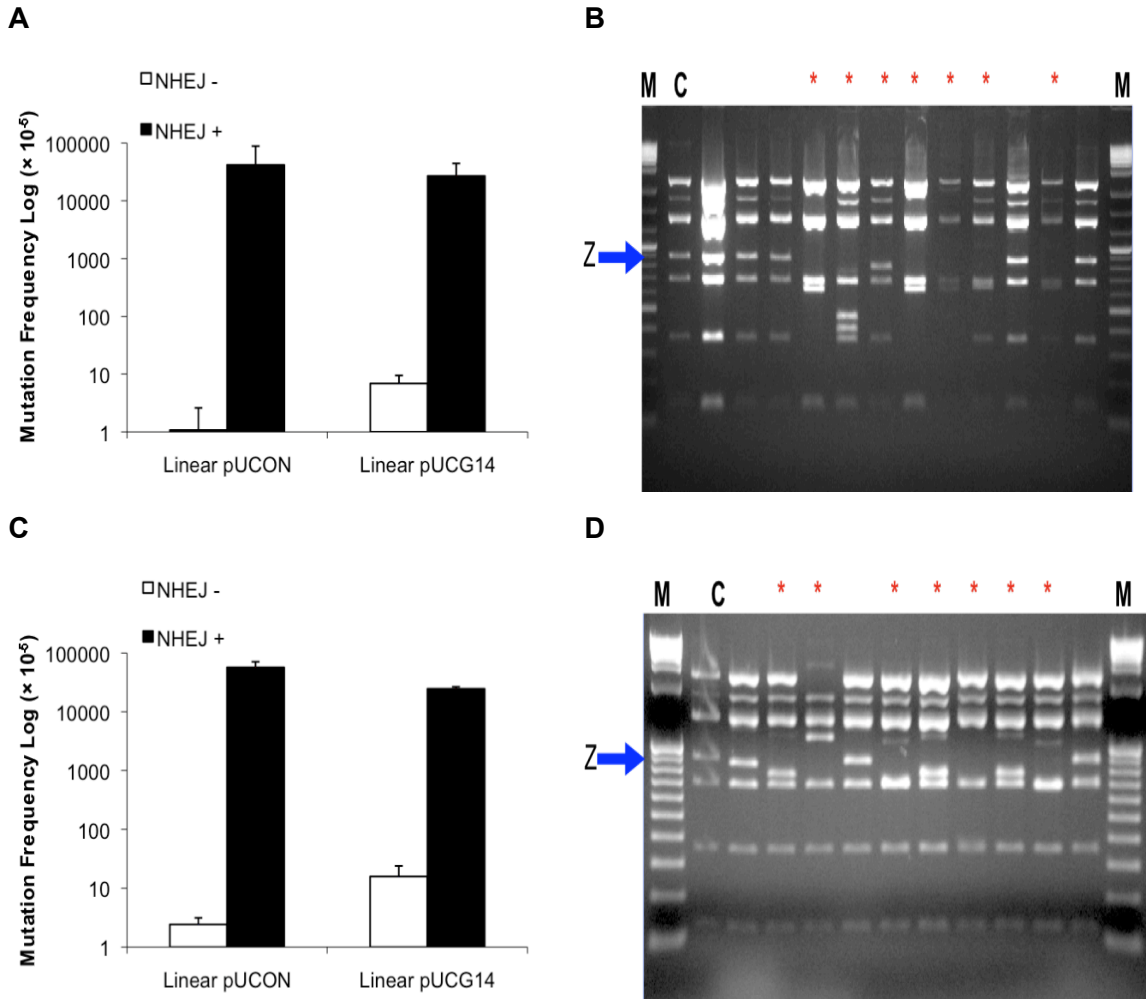


Figure 10. Cells expressing the functional Mt-NHEJ proteins process the DSBs in a mutagenic fashion. (A) The repair of EcoRI-generated DSBs resulted in mutation frequencies in the presence (NHEJ+) or absence (NHEJ-) of functional NHEJ in RecA-cells. (B) The NHEJ+ mutant spectrum after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis. (C) The repair of EcoRI-generated DSBs caused mutation frequencies in the presence (NHEJ+) or absence (NHEJ-) of functional NHEJ in RecB- cells. (D) The NHEJ+ mutant spectrum after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis. Linearization of pUCG14 and pUCON was done by EcoRI digestion. The error bars indicate the standard deviation of two separate experiments. The arrows refer to the Z-DNA-forming fragments; “*” marks the large-scale deletions (≥ 50 bp); “C” labels the control plasmid; and “M” labels the size standard marker. (Adapted from Kha et al., 2010)

NHEJ effects Z-DNA-induced mutagenesis in modified *E.coli*

The processing of DSBs begins with Ku binding to the DSB ends in NHEJ repair and RecB binding to the DSB ends in HR repair (47, 67). However, the exhibited divergence between the two DSB repair pathways lies in the subsequent steps of HR, when RecA directs the newly resected 3'-OH ssDNA into a homologous sister chromatid for the strand invasion reaction (91). To eliminate the interactions of RecA with DSB ends and to determine the role of NHEJ on non-B DNA-induced mutagenesis, we used a RecA⁻ strain that is deficient in RecA, but has the Mt-NHEJ expression vectors integrated.

In the RecA⁻ strain, the presence of Mt-NHEJ proteins does not significantly affect the Z-DNA-induced mutation frequency (Figure 11A). The Z-DNA-forming plasmid remains highly mutagenic compared to the control plasmid pUCON, regardless of Mt-NHEJ induction, with mutation frequencies of $\sim 15 \times 10^{-4}$ versus $< 1 \times 10^{-4}$, respectively. Although there was not a distinct difference in the Z-DNA-induced mutation frequencies with or without NHEJ, the mutation spectra differ in the presence or absence of functional NHEJ (Figure 11B). In Figure 11B, the top panel shows the types of Z-DNA-induced mutants from the RecA⁻ strain when Mt-NHEJ was not present (NHEJ⁻); the fragment containing the Z-DNA-forming sequence can be seen, similar to previously published work from our lab (1). The majority of the mutations were small expansions or deletions within the repeats. Confirmed by direct DNA sequencing analysis of mutant clones, the overwhelming majority of mutations found were small-scale deletions when NHEJ was absent (no large deletions (>50 bp) were detected in 22 total mutants; top panel in Figure 11B). In contrast, when Mt-NHEJ was introduced into the RecA⁻ cells, some portion of the Z-DNA-induced mutants were found to have undergone large-scale deletion events, as seen in the bottom panel of Figure 11B (3 large-scale deletion mutants/22 total mutants).

Since the binding and processing activities of RecB occur before RecA activities in the HR pathway, the non-B DNA-induced mutagenesis generated in the RecA⁻ strain may not fully reflect the exclusive work of NHEJ or a complete shut down of the HR pathway. DSB ends sequestered by RecB may prevent Mt-Ku and Mt-LigD from performing NHEJ on the broken DNA ends. Therefore, in an attempt to reduce HR efficiency by a different mechanism, we studied a RecB⁻ strain that is deficient in this

critical factor that processes the DSB ends and facilitates RecA loading in the HR mechanism (102).

When Mt-NHEJ proteins were present in the RecB⁻ strain, the spontaneous and non-B DNA-induced DSBs were directed into the more mutagenic NHEJ repair pathway, as indicated by an overall increase in the mutation frequencies for all the non-B DNA plasmids. Prior to Mt-NHEJ induction, the Z-DNA-induced mutation frequency was $\sim 13 \times 10^{-4}$ (Figure 12A). This mutation frequency modestly increased to $\sim 17 \times 10^{-4}$ when in the presence of NHEJ (p-value < 0.01; Figure 12A). Even though the statistical analysis of the Z-DNA-induced mutation frequencies showed a significant increase after Mt-NHEJ induction, the control plasmid pUCON (or spontaneous) mutation frequency had a significant increase as well, with its frequency rising from $\sim 1.5 \times 10^{-4}$ to $\sim 7 \times 10^{-4}$ when NHEJ was present (p-value < 0.01; Figure 12A). Therefore, the mutations generated in the RecB-deficient cells showed that NHEJ could increase mutagenesis in a sequence-independent manner, since DSBs can be generated on plasmid DNA spontaneously. However, contrary to the mutagenesis frequency results, there was a striking difference seen in the types of Z-DNA-induced mutations compared to the spontaneous mutations generated in the RecB⁻ strain (Figure 12B). The top panel of Figure 12B shows mutants generated in the absence of NHEJ (NHEJ⁻), and the bottom panel shows mutants generated after Mt-NHEJ induction (NHEJ⁺).

In contrast to the NHEJ-mediated mutations seen in the RecB⁻ strain, in the WT/NHEJ⁺ strain (containing both RecA and RecB), the expression of the Mt-NHEJ genes in wild-type *E.coli* cells did not show a significant impact in altering the spontaneous or the non-B DNA-induced mutation frequencies (Figure 13A). Notably, the Z-DNA-induced mutation frequency of $\sim 16 \times 10^{-4}$ before Mt-NHEJ induction (Figure 13A) was not significantly higher than $\sim 11 \times 10^{-4}$ in Mt-NHEJ⁺ cells based on our statistical analysis. These results from the WT/NHEJ⁺ strain are surprising, since there was an expectation for the NHEJ and HR pathways to be competitive in repairing the non-B DNA-induced DSBs, when both pathways were made available. In addition, we predicted that in the presence of the more 'error-prone' mutagenic NHEJ repair pathway, there would be more mutations generated.

This prediction was based on the idea that if there was an absence of one DSB repair pathway, then there should be a shunt in the DSB repair towards the other available repair pathway. HR typically provides an accurate and 'error-free' repair of

DSBs with the assistance of a homologous sister chromatid as a template. During the ssDNA invasion reaction and homologous pairing event in HR, misalignments can occur, which can lead to small expansions or deletions at simple repeat sequences, like the CG(14) repeats in the Z-DNA-forming sequence used in this study. These events are often restricted to the repetitive area and do not affect the neighboring sequences. In NHEJ repair, as opposed to HR repair, there is the possibility of generating larger deletions of the sequences between the two homologous regions. It is possible that when both NHEJ and HR pathways are available in the Ara⁺ cells, the HR pathway is still the major repair activity of the DSBs, and thus, the effects of NHEJ can be overshadowed by HR repair. Notably, similar misalignment can also occur during DNA replication, leading to small deletions and insertions within the repeats.

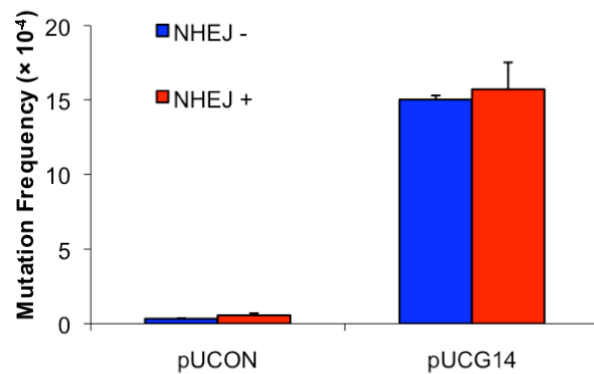
We were able to determine if the presence of NHEJ in *E.coli* alters the mutation frequencies and types of mutants induced by non-B DNA structures, with the observed mutation frequencies after Mt-NHEJ induction and characterization of the mutants by restriction digestion and sequencing analyses. DNA plasmids obtained from the mutants generated in each strain, with or without Mt-NHEJ induction, were digested with EagI and BssSI and then electrophoresed on a 1.4% agarose gel. And overall, the presence of NHEJ system did substantially shift the mutation spectra with more proportions of large-scale deletions after Mt-NHEJ was induced. The three strains that contain Mt-NHEJ expression vectors, when induced with L-arabinose, came with an assortment of genetic backgrounds, where one strain was wild-type in HR and the other two strains had a deficiency in the HR pathway, either in RecA-deficiency or RecB-deficiency.

Collectively, the percentage of large-scale deletions in the Z-DNA-induced mutants was calculated as shown in Figure 12C for the NHEJ-deficient (NHEJ⁻) and the NHEJ-proficient (NHEJ⁺) groups. The ratio of large-scale deletions of Z-DNA-induced mutants to total number of mutants from the RecA⁻, RecB⁻ and WT/NHEJ⁺ strains shifted from 2% (1 large-scale deletion mutant/56 total mutants, without Mt-NHEJ induction) to 24% (12 large-scale deletion mutants/50 total mutants, with Mt-NHEJ induction), implicating the NHEJ repair pathway in the Z-DNA-induced large-scale deletions. Sequencing analysis confirmed that when the cells were lacking NHEJ, there was a majority of small-scale deletions or insertions with 2-24 bp alteration within the Z-DNA-forming CG(14) sequence. There were, however, a significant number of large-

scale deletion (≥ 50 bps) mutants that had complete loss of the Z-DNA-forming sequence and some adjoining sequences when functional NHEJ was present. The junctions in these mutants were composed of 2-4 bp of microhomologies, similar to what was found in mammalian cells (Figure 14) (1). The Z-DNA results demonstrate that the mutations produced in the NHEJ+ *E.coli* cells were similar to the types of mutations generated in mammalian cells, which contain an endogenous NHEJ mechanism.

In support of this notion, using Ligation-mediated PCR (LMPCR) experiments as previously described in Wang et al. (1) with a modification procedure to isolate the plasmid DNA from bacterial cells, the Z-DNA CG(14) sequence did indeed lead to DSBs in *E.coli*. The plasmid DNA extracted from MBM7070 bacterial cells following the modified method described by Zhang et al. (103) allowed for collection of small DNA, rather than collecting circled DNA as in the alkaline lysis methods used previously in Wang et al. (1). The Z-DNA CG(14) sequence was found to cause a DSB hotspot in bacteria, while there was no breakpoint found near the control sequence in plasmid pUCON (data not shown and produced by Dr. Graham Wang in the Vasquez lab). This finding further demonstrated that DSBs were caused by Z-DNA in bacteria, which is similar to what was seen in mammalian cells.

A



B

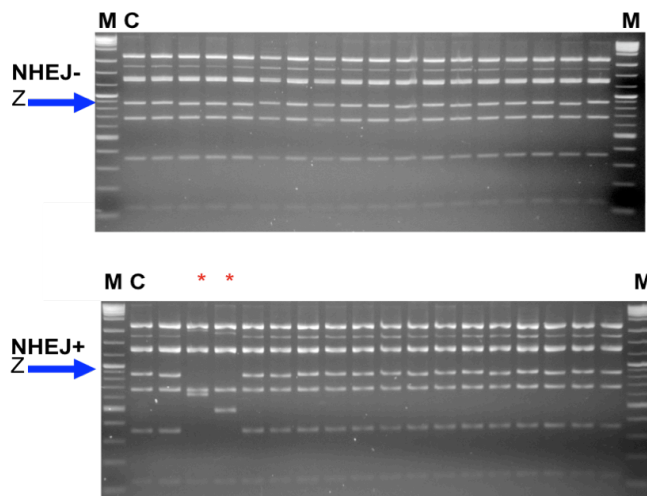


Figure 11. Spontaneous and Z-DNA-induced mutation frequencies and spectra in the RecA⁻ strain. (A) Spontaneous and Z-DNA-induced mutation frequencies in the RecA⁻ strain in the presence (NHEJ⁺) or absence (NHEJ⁻) of the Mt-NHEJ proteins. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments. (B) The mutants induced by Z-DNA are shown on a representative gel for the 44 Z-DNA-induced mutants analyzed from the RecA⁻ strain after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis from NHEJ⁻ (top panel) and from NHEJ⁺ (bottom panel) strains. The arrows refer to the Z-DNA-forming fragments; “*” marks the large-scale deletions (≥ 50 bp); “C” labels the control plasmid; and “M” labels the size standard marker.

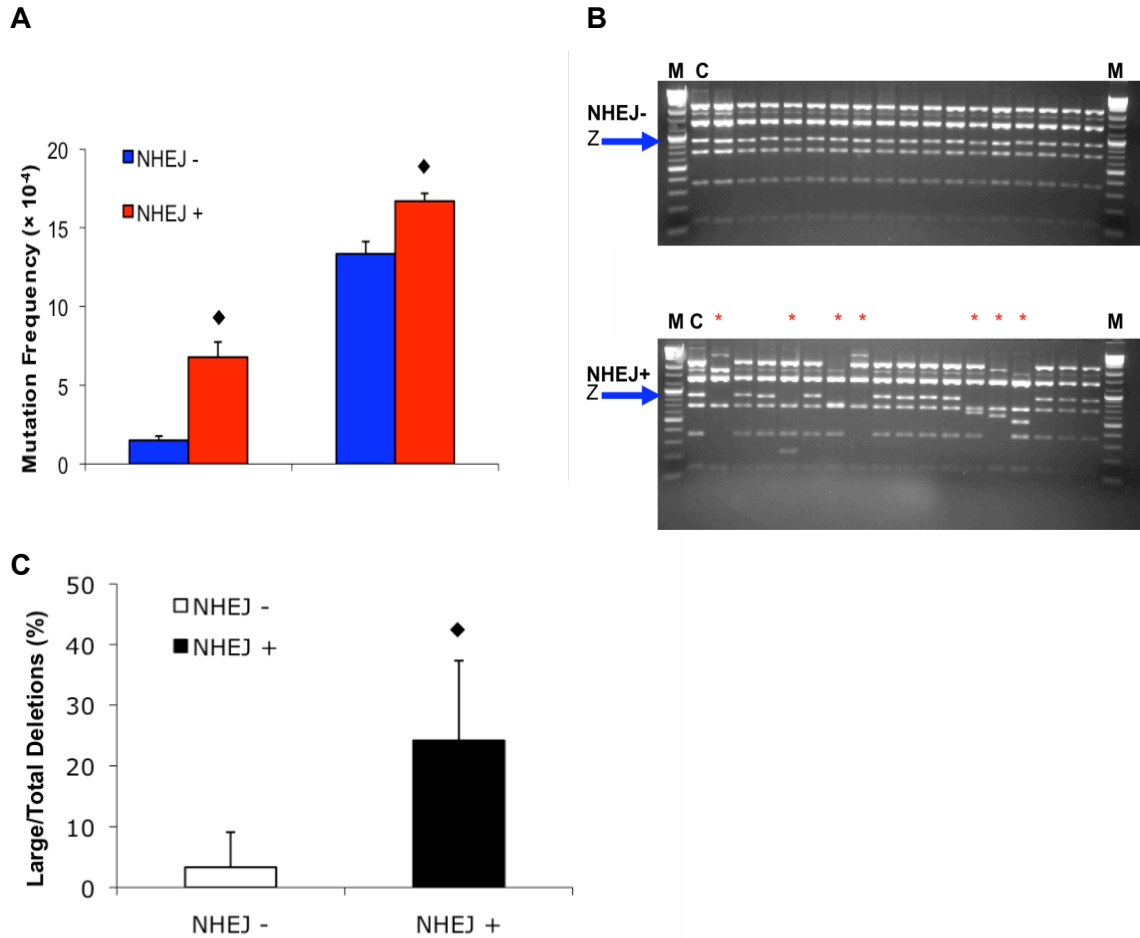
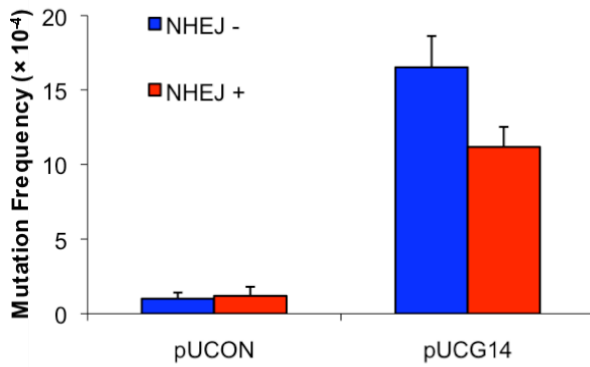


Figure 12. Spontaneous and Z-DNA-induced mutation frequencies and spectra in the RecB- strain. (A) Spontaneous and Z-DNA-induced mutation frequencies in RecB-cells in the presence (NHEJ+) or absence (NHEJ-) of the Mt-NHEJ proteins. (B) The Z-DNA induced mutation spectrum is shown on a representative gel for the 42 Z-DNA-induced mutants analyzed. There are >100,000 colonies in each group and >100 mutants screened; the error bars indicate the standard deviation of three separate experiments and bacterial groups. The arrows refer to the Z-DNA-forming fragments; “*” marks the large-scale deletions (≥50 bp); “C” labels the control plasmid; and “M” labels the size standard marker. The Z-DNA-induced mutants were digested with EagI and BssSI, as seen in the separation of the seven fragments by agarose gel electrophoresis. (Adapted from Kha et al., 2010). (C) Total number of large-scale deletions to total number of deletions from the combined NHEJ-deficient (NHEJ-) and NHEJ-proficient (NHEJ+) bacterial strains. The “♦” symbol marks the significant difference (p-value < 0.01) between the NHEJ- and NHEJ+ induction.

A



B

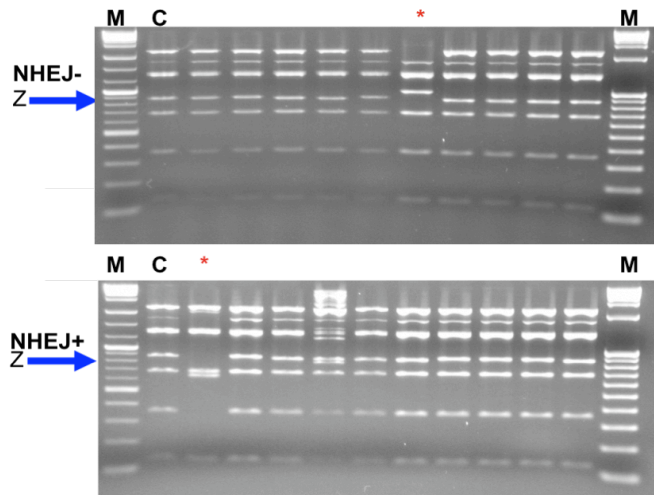


Figure 13. Spontaneous and Z-

DNA-induced mutation frequencies and spectra in the WT/NHEJ+ strain. (A) Spontaneous and Z-DNA-induced mutation frequencies in the WT/NHEJ+ strain in the presence (NHEJ+) or absence (NHEJ-) of the Mt-NHEJ proteins. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments. (B) The Z-DNA-induced mutations are shown on a representative gel for the 20 Z-DNA-induced mutants analyzed from the WT/NHEJ+ strain after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis; NHEJ- (top panel) and from NHEJ+ (bottom panel). The arrows refer to the Z-DNA-forming fragments; “*” marks the large-scale deletions (≥ 50 bp); “C” labels the control plasmid; and “M” labels the size standard marker.

The functional NHEJ system, and not Ku or LigD individual activity, is involved in the Z-DNA-induced large-scale deletions

The multi-domain Mt-LigD protein is an essential partner in the “two-component” Mt-NHEJ system. On the C-terminus of Mt-LigD there is a ligase domain, while the polymerase domain resides at the N-terminus, and in between there is a nuclease domain that has 3' to 5' exonuclease activity (3, 68-71). The multiple activities of Mt-LigD raises the question of whether there is a single protein activity that could, rather than NHEJ activity as a whole, cause the large-scale deletions seen in the Z-DNA-induced mutants in Ara⁺ cells. Also in question is the Mt-Ku homodimer, which binds to DNA breaks and can also alter the type of DSB repair implemented in processing the DSBs depending on other repair factors available (LigD or LigC, etc) (3, 55, 68, 104, 105). Therefore, to sort out whether the NHEJ pathway or the individual components, Mt-Ku or Mt-LigD proteins, are responsible for the large-scale deletions in the Z-DNA-induced mutants, the Z-DNA-forming plasmid was transformed into the Mt-Ku only and Mt-LigD only strains to calculate and compare the Z-DNA-induced mutation frequencies and spectra to those found in cells containing both Mt-Ku and Mt-LigD. These particular strains only express the individual components, whether Mt-Ku or Mt-LigD, which makes them insufficient in the full Mt-NHEJ activities seen in the RecA-, RecB- and WT/NHEJ+ strains after L-arabinose induction.

In the absence of Mt-Ku in the wild-type *E.coli* cells, the Z-DNA-induced mutation frequency was $\sim 17 \times 10^{-4}$, which was the same Z-DNA-induced mutation frequency for the WT strain before L-arabinose was supplemented into the LB (Figure 15A and 9A). The introduction of Mt-Ku into the wild-type *E.coli* cells produced a Z-DNA-induced mutation frequency of $\sim 32 \times 10^{-4}$ (Figure 15A). The difference in mutation frequencies in the Mt-Ku only strain was not significantly different from the uninduced cells (p value = 0.16). Similarly, in the cells that only expressed Mt-LigD, the Z-DNA-induced mutation frequency was not significantly altered from $\sim 21 \times 10^{-4}$, without Mt-LigD induction, to $\sim 17 \times 10^{-4}$, with Mt-LigD induction (Figure 15C). Our data is in agreement with previous results that showed that the Mt-LigD nuclease activity is not involved in processing DNA breaks independently, and that Mt-Ku works with Mt-LigD as the two-component NHEJ repair system (3, 106).

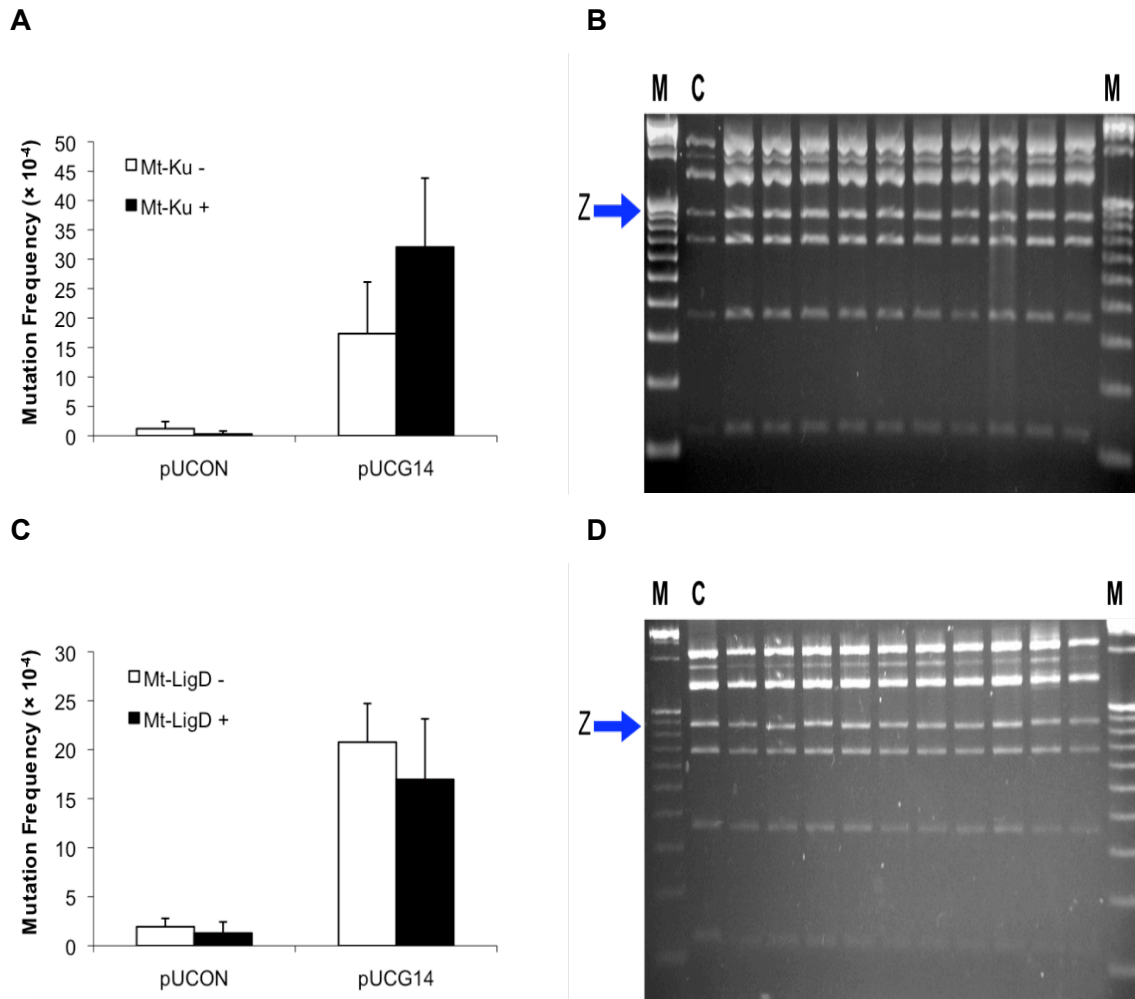


Figure 15. Z-DNA-induced and spontaneous mutation frequencies and spectra in Mt-Ku only and Mt-LigD only bacterial strains. (A) Z-DNA-induced and pUCON-induced (i.e. spontaneous) mutation frequencies in Mt-Ku only cells in the presence (Mt-Ku+) or absence (Mt-Ku-) of the Mt-Ku protein. (B) The pUCG14 mutant spectrum after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis from Mt-Ku+ cells. (C) Z-DNA-induced and pUCON-induced mutation frequencies in Mt-LigD only cells in the presence (Mt-LigD+) or absence (Mt-LigD-) of the Mt-LigD protein. (D) The pUCG14 mutation spectrum after EagI and BssSI digestion and separation of the seven fragments by agarose gel electrophoresis from Mt-LigD+ cells. There are >70,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments. The arrows refer to the Z-DNA-forming fragments; “C” labels the control plasmid; and “M” labels the size standard marker. (Adapted from Kha et al., 2010)

Moreover, the types of mutants induced by Z-DNA in both of the Mt-Ku only and Mt-LigD only strains were consistent with the inability of the individual proteins to process the DSBs. The mutations were small-deletions within the CG(14) repeat (20 small-scale deletion mutants/20 total mutants; Figure 15B and 15D), as also seen prior to Mt-Ku or Mt-LigD expression in the wild-type *E.coli* cells (20 small-scale deletion mutants/20 total mutants; data not shown). Regardless of the Mt-Ku or Mt-LigD expression status, the Z-DNA-induced mutation frequencies and spectra were unaltered. It is indeed the entire NHEJ system, rather than either Mt-Ku or Mt-LigD alone that is responsible for generating the large-scale deletions on the Z-DNA-forming plasmids in the modified bacterial *E.coli* cells.

There are other potential factors in addition to the traditional HR and NHEJ components, which may be involved in non-B DNA-induced genomic instability in our study. *In vivo*, there are other DSB repair mechanisms that are not directly categorized under HR or NHEJ, such as the alternative RecFOR homologous recombination pathway in prokaryotes that is also available and independent of the RecBCD function in HR (89). As noted from our assays, there highly accurate DNA end rejoining of “sticky-ended” DSBs generated from EcoRI digestion, suggesting a direct ligation of the DSBs could also be involved. Interestingly, repair of the “blunt-ended” DSBs resulting from EcoI/CRI digestion occurred in a more mutagenic fashion (more than 60% of the recovered mutants had a mutation on the *lacZ* gene 4-bp from the targeted DSB; data not shown). There are also recent new findings of an end-joining repair mechanism in *E.coli* that has not been characterized prior to this study (107). This mechanism is different than canonical NHEJ, and is dependent on ligase-A activity. Nonetheless, the efficiency levels of these processes (identified or as yet unidentified) may not be as high as in the HR or NHEJ pathway and they contribute a very minor role for processing the DSBs, as is evident by the reduced amount of colonies after the linearized plasmids were transformed into the RecA-/RecB- strains without Mt-NHEJ induction compared to the cells that are proficient in RecBCD and RecA or the cells that had the Mt-NHEJ proteins present.

When NHEJ and HR are both available, there can be competition for the DSB intermediates, as previously reported (108). Even so, this is not seen in the processing of the Z-DNA-induced DSBs when both DSB repair systems were made available, as in the modified *E.coli* cells that contained the Mt-NHEJ proteins. If NHEJ was not present,

such as in the wild-type *E.coli* cells, then the repair of the Z-DNA-induced DSBs may be shunted to the HR pathway or to other available pathways (Figure 16), resulting in accurate ligation of small deletion/expansions in the repetitive sequences. And although the novel expression of Mt-Ku and Mt-LigD in *E.coli* cells can be sufficient for the initiation of processing broken DNA ends and the ligation of the DSB strands in a NHEJ-like fashion, different repair outcomes can be produced when these proteins are expressed in various genetic backgrounds; indicating that the host *E.coli* cell proteins can assist with the repair processes (3, 55, 106). Additionally, independent of DSBs and/or their repair processing, there can also be contractions or expansions within the CG(14) repeats via slippage events during DNA replication (1), which can overshadow the small-scale deletions or expansions resulting from DSB repair. Thus, the contribution of NHEJ proteins in Z-DNA-induced mutagenesis could be underestimated in our system.

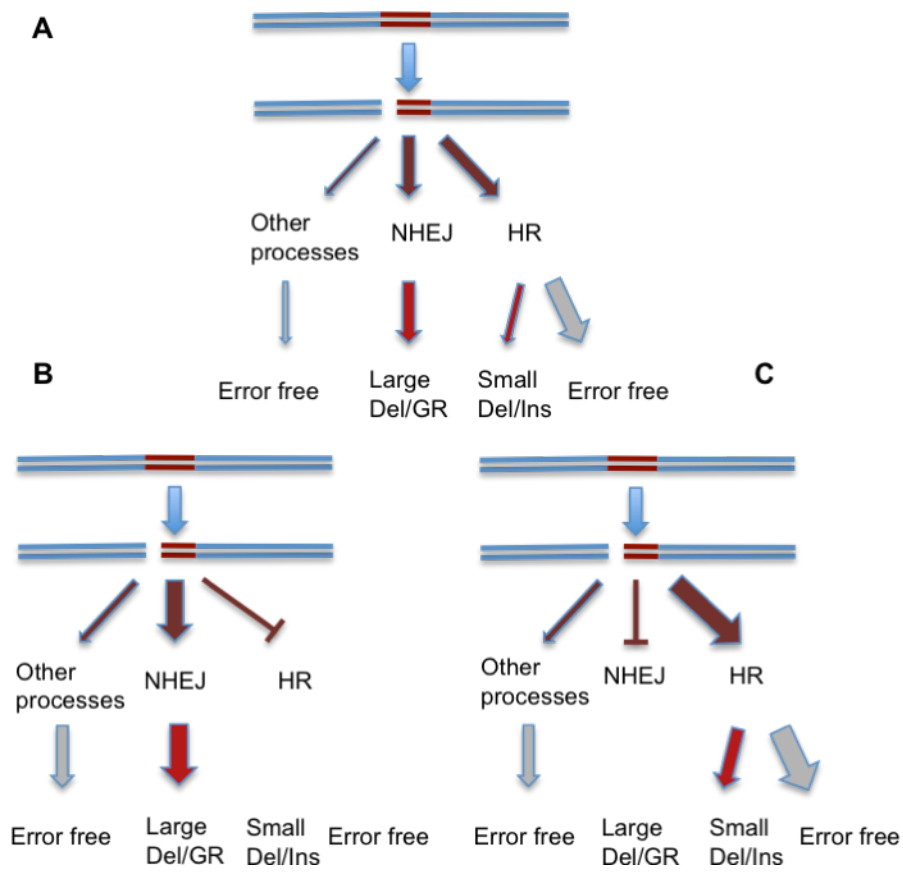


Figure 16. Proposed model for the role of double-strand repair pathways in processing of Z-DNA-induced DSB. (A) HR, NHEJ, and other processes, such as direct ligation of DNA breaks, may participate in the processing of Z-DNA-induced DSBs. (B) In mammalian cells, where NHEJ is available, large-scale deletions are prevalent after the Z-DNA-induced DSBs are processed by that mechanism. (C) However, in *E.coli*, where NHEJ is absent, the HR pathway largely repairs the DSBs and this may produce small expansions/contractions within the repeat because of misalignment events. (Adapted from Kha, et al., 2010)

NHEJ does not affect H-DNA-induced mutagenesis in modified *E.coli*

Previous results in our laboratory demonstrated that Z-DNA is mutagenic in both mammalian cells and bacterial cells, while H-DNA is only mutagenic in mammalian cells and not in wild-type *E.coli* (1, 2). This quite stable, non-mutagenesis characteristic of H-DNA in bacteria may possibly be due to the lack of an error prone NHEJ pathway in *E.coli* cells. In addition, since the H-DNA-forming sequence tested in this study was not a simple repetitive sequence, there is a greater possibility for an accurate alignment of the invading strand into the homologous template during HR, which would not result in small expansions or deletions, allowing for error-free repair of H-DNA-induced DSBs. Therefore, the H-DNA-induced mutagenesis was re-evaluated in modified *E.coli* that express the Mt-NHEJ proteins to mimic a similar DSB repair situation in bacteria as in mammalian cells.

As previously found for our Z-DNA studies described above, the addition of the inducible factor, L-arabinose, to the LB culture for induction of Mt-NHEJ protein expression did not influence the spontaneous or H-DNA-induced mutation frequencies in WT *E.coli* cells (Figure 17). The low mutation frequencies of the H-DNA-forming sequences, which were all below 1×10^{-4} , validates previous observations that H-DNA is not mutagenic in wild-type *E.coli* cells (Figure 17) (2). And similar to the Z-DNA-induced genetic instability study, the H-DNA-induced genetic instability was evaluated in most of the same strains listed in Table 1, except for the Mt-Ku only and Mt-LigD only strains. In the RecA⁻ and WT/NHEJ⁺ strains that have the Mt-NHEJ expression vectors, very similar mutation frequencies were found between the H-DNA and control plasmids. In both strains, there were consistently low mutation frequencies that had no significant differences when Mt-NHEJ is expressed (Figure 18 and 20). In the RecA⁻ cells, the mutation frequencies remained low, before and after Mt-NHEJ induction, with all frequencies below $\sim 1 \times 10^{-4}$ (for both U and Y-direction; Figure 18). The combined average mutation frequencies of the two H-DNA-forming sequences (U and Y-direction) were at $\sim 2 \times 10^{-4}$, before Mt-NHEJ induction, and $\sim 1.8 \times 10^{-4}$, after Mt-NHEJ induction in the WT/NHEJ⁺ strain (Figure 20). This lack of an effect of NHEJ on H-DNA-induced genetic instability is further confirmed through characterization of the mutants. The restriction and sequencing analyses of a few available spontaneous control plasmid pUCON-induced mutants and H-DNA-induced mutants showed that the types of mutations consisted of a mixture of small-scale and large-scale deletions when Mt-

NHEJ was not induced and when Mt-NHEJ was induced, for both plasmids (data not shown).

In the RecB⁻ cells, there was a significant increase in both the H-DNA and pUCON-induced (i.e. spontaneous) mutation frequencies when Mt-NHEJ was expressed, which is similar to the Z-DNA-induced genetic instability results from the RecB⁻ strain (Figure 19). The mutation frequencies for the H-DNA-forming sequences, prior to Mt-NHEJ induction, were averaged to $\sim 1 \times 10^{-4}$ (for both U and Y-directions; Figure 19). Following Mt-NHEJ induction, the H-DNA-induced mutation frequencies had a ~ 5 and a half-fold increase (p -value < 0.01 ; Figure 19). This ~ 5 and a half fold increase after NHEJ expression, however, is not H-DNA specific, because the pUCON-induced mutagenesis is also increased by the same fold after Mt-NHEJ induction (p -value < 0.01 ; Figure 19). Clearly, the data for the H-DNA-forming sequences strongly indicates that H-DNA is not mutagenic in bacteria cells, regardless of the expression of NHEJ.

There are several possible factors that may come into play for this lack of H-DNA-induced mutagenesis in the modified *E.coli*. Unlike the Z-DNA-induced mutagenesis results, the types of mutants found and mutation frequencies of the H-DNA-forming sequences were nearly identical to the control plasmid pUCON-induced mutation frequencies and spectra (Figure 18-20). In the RecB-deficient strain, the mutation frequencies for both the control plasmid pUCON and H-DNA-forming plasmid increased when Mt-NHEJ was induced (NHEJ⁺; Figure 19). With this increase in the number of mutants, there was also an increase in the ratio of large-scale deletions in the spontaneous and H-DNA-induced mutants when NHEJ was available (data not shown), which we believe is due to the error-prone NHEJ repair of spontaneous DSBs generated in bacteria. The characterization of the spontaneous and H-DNA-induced mutants in the other strains, with and without Mt-NHEJ, showed a mix of point mutations and large-scale deletions for all the strains studied (data not shown), suggesting that the sequences that can form H-DNA structures were stable in bacterial cells, and the spontaneous mutations included a low level of large-scale deletions, probably as a result of nuclease digestion and re-ligation.

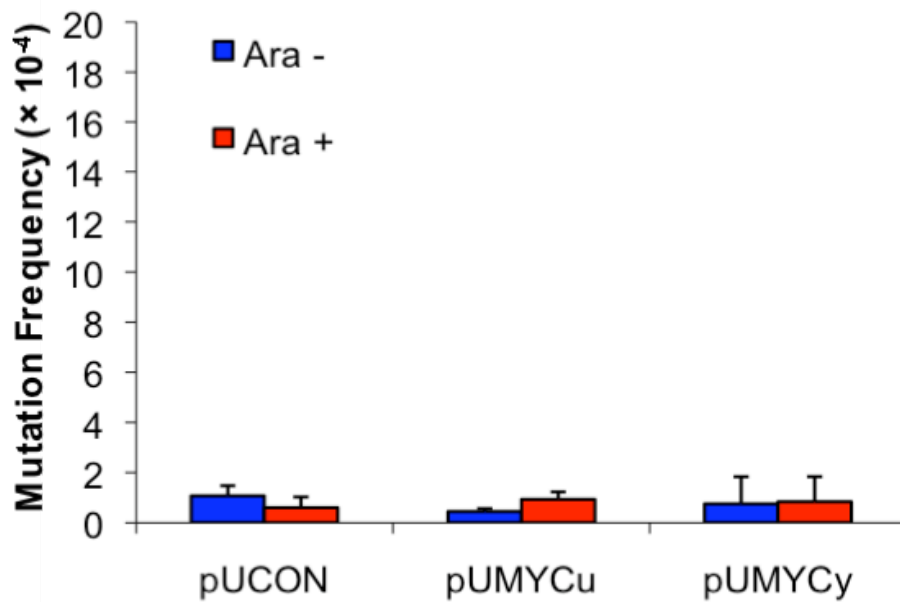


Figure 17. The inducible factor, L-arabinose, does not affect H-DNA-induced mutagenesis *per se*. L-arabinose (Ara+), the inducible factor of the Mt-NHEJ proteins, did not effect the mutation frequencies of the control plasmid and H-DNA plasmid in wild-type (WT) *E.coli* cells that do not have the Mt-NHEJ expressing system. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments.

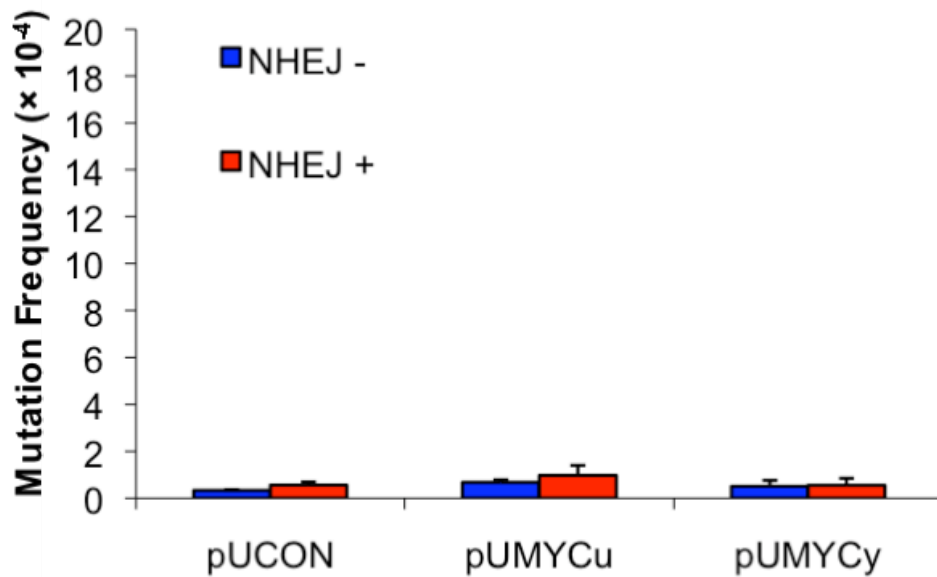


Figure 18. Spontaneous and H-DNA-induced mutation frequencies from the RecA-strain. This graph shows the spontaneous and H-DNA-induced mutation frequencies in the RecA- strain in the presence (NHEJ+) or absence (NHEJ-) of the Mt-NHEJ proteins. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments.

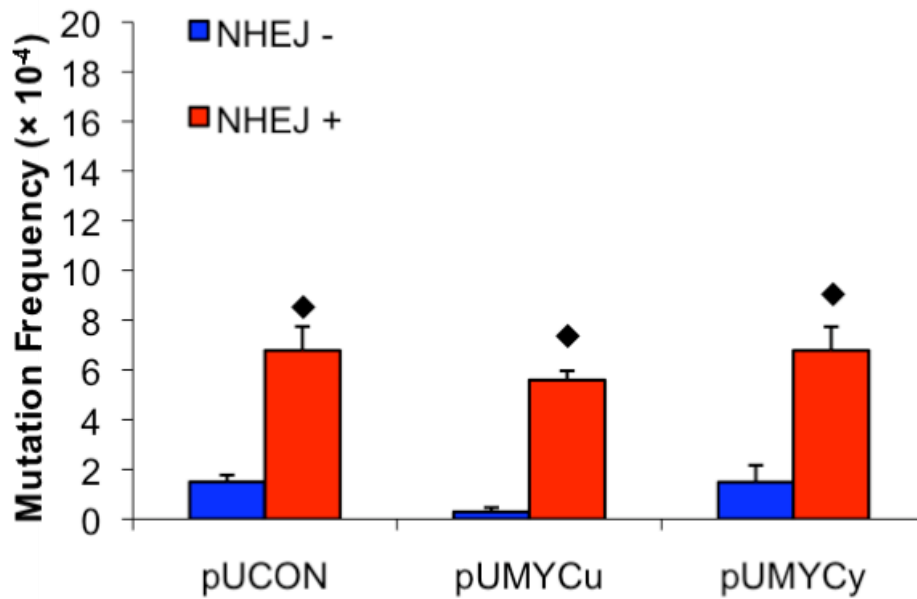


Figure 19. Spontaneous and H-DNA-induced mutation frequencies from the RecB-strain. This graph shows the spontaneous and H-DNA-induced mutation frequencies in the RecB- strain in the presence (NHEJ+) or absence (NHEJ-) of the Mt-NHEJ proteins. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments.

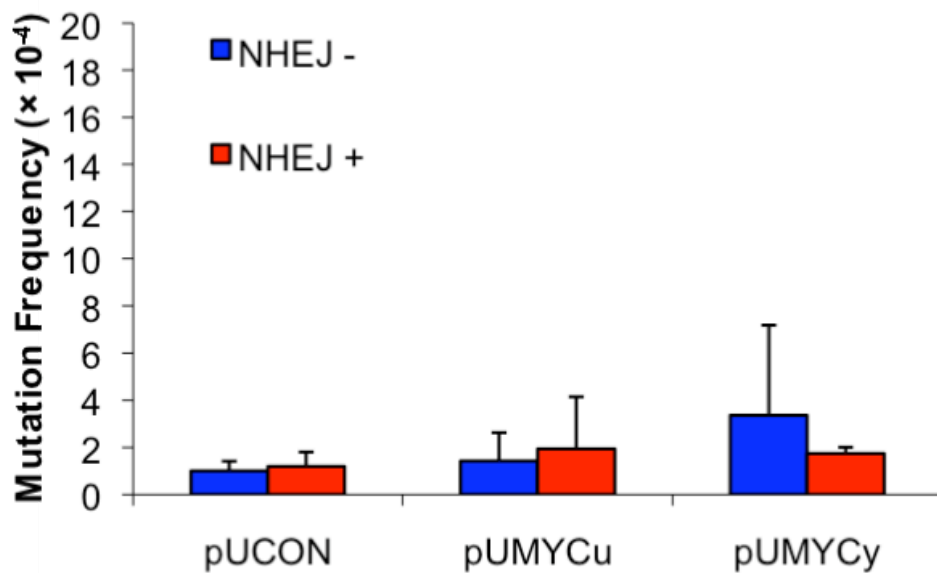


Figure 20. Spontaneous and H-DNA-induced mutation frequencies from the WT/NHEJ+ strains. This graph shows the spontaneous and H-DNA-induced mutation frequencies in the WT/NHEJ+ strain in the presence (NHEJ+) or absence (NHEJ-) of the Mt-NHEJ proteins. There are >100,000 colonies in each group and the error bars indicate the standard deviation of three separate experiments.

Summary

We studied the roles of NHEJ and HR on non-B DNA-induced genetic instability by determining the Z-DNA and H-DNA-induced mutation frequencies and mutant characterization in modified *E.coli* strains with proficiencies and/or deficiencies in NHEJ (Ku/LigD) and HR (RecA/RecB). To summarize, in exploring for possible explanation(s) for the different non-B DNA-induced mutagenesis patterns between mammalian cells versus bacterial cells, our results suggested that, if made available, NHEJ can repair the spontaneous and non-B DNA-induced DSBs in a mutagenic manner.

Previous Z-DNA and H-DNA-induced mutagenesis results revealed that the non-B DNA structures, which induced DSBs, had higher mutation frequencies and larger deletions in the mutants generated in mammalian cells versus those generated in *E.coli* cells (1, 2). NHEJ is presumably the more error-prone pathway compared to the HR pathway. The generation of inaccurate repair products via NHEJ is specifically reflected in the types of Z-DNA-induced mutants found when both the *M.tuberculosis* NHEJ proteins, Mt-Ku and Mt-LigD, were supplemented together into the modified *E.coli* cells that had varying proficiencies of HR (RecA/RecB). The statistically significant increase in large-scale deletions in the Z-DNA-induced mutants was visually apparent in the mutation spectra and was confirmed in the sequencing analyses of the Z-DNA-induced mutants recovered from the modified *E.coli* strains that contained functional NHEJ (Figure 11-13). Moreover, when the Mt-Ku and Mt-LigD proteins were expressed individually in the Mt-Ku only and the Mt-LigD only strains, respectively, there was a lack of large-scale deletions in the Z-DNA-induced mutants (Figure 15). The type of Z-DNA-induced mutants found in those strains was more similar to the types of Z-DNA-induced mutants found in wild-type *E.coli* that does not contain a NHEJ mechanism (Figure 9) or the modified *E.coli* cells that had no NHEJ induction (NHEJ-; Figure 11-13). Therefore, without complete NHEJ function or availability, the types of Z-DNA-induced mutants that were detected predominantly consisted of small-scale deletions. And when NHEJ was induced in the modified *E.coli* cells, the types of Z-DNA-induced mutants shifted from small-scale deletions to large-scale deletions, similar to those of Z-DNA-induced mutants in mammalian cells that contain endogenous NHEJ (1, 4).

The types of Z-DNA-induced mutants were dependent on the available DSB repair pathway, going from 2% large-scale deletions/rearrangements in total mutants to 24% large-scale deletions when NHEJ became available. The Z-DNA-induced mutation

frequencies, however, generally remained the same with or without induction of the NHEJ pathway (4). The Z-DNA-forming sequence consistently induced high levels of mutations in different *E.coli* strains, either with or without NHEJ and/or HR pathways, and the presence of NHEJ did not significantly change the Z-DNA-induced mutation frequencies in the bacterial cells (Figure 9, 11-13, 15). Only when NHEJ became available in *E.coli*, did the detection of large-scale deletions and rearrangements on the Z-DNA plasmid show a significant increase, suggesting that NHEJ repair is involved in the large-scale deletion and rearrangements caused by this non-B DNA structure, which increases our mechanistic understanding of non-B DNA-induced genetic instability in various species.

The H-DNA-induced mutagenesis frequencies remained stable and were not altered in the presence or absence of NHEJ or HR in the newly modified *E.coli* cells (Figure 17-20). The H-DNA mutagenesis induced in bacterial cells versus mammalian cells continued to be different, even after the effort to gap the bridge between the two cell-types, with the addition of Mt-NHEJ into the bacterial *E.coli* cells. The H-DNA mutation frequencies were consistently at the same level as spontaneous mutations in the modified *E.coli* cells, before and after NHEJ was induced (Figure 17-20). Since the H-DNA-induced mutation frequencies and spectra followed that of the control plasmid pUCON, more studies of other repair pathways and other factors involved with this non-B DNA structure are warranted. In general, extended studies to further clarify the mechanisms of DNA structure-induced genetic instability are needed to broaden our understanding of how DNA structure influences human diseases, genetic instability, and evolution.

CHAPTER III: FUTURE DIRECTIONS

III.I FUTURE DIRECTIONS

Studying the relationship between DSB repair pathways and non-B DNA-induced genetic instability revealed important qualitative roles of NHEJ and HR on the repair of non-B DNA-induced DSBs, specifically finding the mutagenic role of NHEJ in producing the Z-DNA-induced large-scale deletions as seen in mammalian cells. Since the discovery of the B-DNA double helical confirmation by Watson and Crick over 50 years ago (5), several other types of DNA structure have been identified and characterized. These alternative DNA structures seem to play functional roles in the cell, as we have demonstrated here. The work described in this thesis may provide a peephole into further studies on the functional role of DNA structure.

As stated before, there are other factors that could help further explain the difference in the non-B DNA-induced mutagenesis found in mammalian cells versus bacterial cells that was not elucidated in this study. For example, there are other DNA binding proteins, in addition to RecA, that can contribute to the differences in non-B DNA-induced mutagenesis between mammalian cells and bacterial cells. Some of these DNA binding proteins are involved in other DNA repair mechanisms that are different from the DSB repair pathways, NHEJ and HR, such as the nucleotide excision repair (NER) and the mismatch repair (MMR) mechanisms. Specifically, like RecA, which binds preferentially to the Z-DNA structure over the B-DNA structure (21), unpublished chromatin immunoprecipitation results from our laboratory showed that antibodies against the XPA and MSH2 proteins, which are involved in NER and MMR, respectively, were also enriched at Z-DNA-forming sequences compared to the B-DNA control sequence (6). Thus, other DNA repair mechanisms, such as NER and MMR, and not solely NHEJ and HR, could also have roles in non-B DNA-induced genetic instability. Indeed, studies from our laboratory of non-B DNA-induced mutagenesis in the absence of functional NER and MMR have revealed a role for XPA and MSH2 in the process (6). It would also be interesting to determine potential crosstalk involved between the repair pathway components, in relation to the non-B DNA-induced genetic instability.

To further assess the roles of DSB repair pathways on DNA structure-induced genetic instability, there are further in depth studies that would allow the expansion of this particular study and will be proposed as followed. To further support to the idea that the NHEJ pathway did indeed cause the types of large-scale deletions in the Z-DNA-

induced mutants found from mammalian cells, we could examine mammalian cells with deficiencies in NHEJ. Results from these kinds of studies may be relevant to genetic instability and human disease.

The use of *E.coli* cells allowed for an easy first approach to characterize differences in some of the DSB repair components between mammals and bacteria and allowed us to determine that NHEJ is important in DNA structure-induced mutagenesis. However, direct comparisons between species cannot be made due to differences in the NHEJ and HR systems in various organisms. As stated, the RecA protein in the prokaryotic HR pathway has some differences in its activity compared to the mammalian homolog Rad51 (21, 96). Both of the DSB repair pathways in mammalian cells are indeed more complex compared to the prokaryotic DSB repair pathways. A better observation on the role of DSB repair pathways in non-B DNA-induced genetic instability could be done in mammalian cells that are deficient in NHEJ repair.

We could introduce non-B DNA plasmids that are able to replicate within these NHEJ-deficient cells and study their non-B DNA-induced mutagenic potential. A good collaborative candidate would be Dr. Chengming (Ben) Zhu, whose laboratory is in the Department of Immunology at the University of Texas M.D. Anderson Cancer Center, who studies mouse models with NHEJ-deficiencies in ligase IV (Lig 4^{-/-}) and a hypomorphic mutation in p53, p53R172P, that is not embryonic lethal and does not develop lymphomagenesis (109, 110). We have obtained the mouse embryonic fibroblasts (MEFs) from Dr. Zhu and future work with these cells is planned. In addition, these mouse models that have knockouts in NHEJ components could be crossed with our mouse models that carry the non-B DNA sequences in their chromosomes (31). The NHEJ-deficient mouse models could help solidify the association of NHEJ with the large-scale deletions in the Z-DNA-induced mutants from mammalian cells and mice. We speculate that without the components of NHEJ, there may be a shift from large-scale deletions to small-scale deletions in the Z-DNA (or other non-B DNA)-induced mutants, corroborating our results from the modified *E.coli* cells with no NHEJ induction or the incomplete NHEJ induction (Ku-only and LigD only strains).

The NHEJ-knockout mouse model study would also re-evaluate the cause for the high rate of Z-DNA and H-DNA-induced mutation frequencies found in mammalian cells (1, 2). It will be interesting to see if the high levels of Z-DNA and H-DNA-induced

mutagenesis would remain when in a repair environment that is more similar to bacterial *E.coli* cells, which do not contain NHEJ.

In addition to the use of mouse models to study DSB repair pathways in non-B DNA-induced genetic instability, we could also study non-B DNA-induced genetic instability in human cells with small interfering RNA (siRNA) that knockdown NHEJ factors (or HR factors) as well. In a recent study, Fattah et al. used recombinant adeno-associated viral knockout vectors (rAAV) to produce isogenic human somatic cell lines that were deficient in the NHEJ components (Ku, DNA-PKcs, XLF, and LIGIV) (104). These cells could prove useful to our laboratory for further studies. In addition to observing the effects of the NHEJ knockouts in their study, they also observed the role of 'alternative' NHEJ pathway (A-NHEJ) in DSB repair.

With the preliminary identification of an alternative back-up NHEJ (A-NHEJ or B-NHEJ) in higher eukaryotes that uses DNA ligase III, poly(ADP-ribose) polymerase-1 (PARP-1) and histone H1 (111), we could determine if the non-B DNA-induced DSBs could be shunted into this sub-pathway, which would allow an update of our current model in Figure 16, and further categorized or branch out the model into the 'classical' NHEJ (C-NHEJ) and A-NHEJ/B-NHEJ pathways. Similar to the possibility of other repair processes (NER, MMR) in the mammalian cells that could affect non-B DNA-induced genetic instability, a possible next step for this study could be to elucidate the contribution of the NER and MMR pathways in conjunction with A-NHEJ/B-NHEJ pathway, and/or competition with the classical NHEJ and HR pathways in non-B DNA-induced genetic instability in mammalian cells. Although recent discoveries of an end-joining (A-EJ) repair mechanism in *E.coli* cells may be different from the A-NHEJ found in eukaryotes, preliminary studies can be done in the bacterial cells as well for comparisons with our results in bacterial systems (107).

Building from this study, with the use of MEFs, NHEJ-deficient mice models, and siRNA or rAAV knockout of NHEJ in human cells to extend our findings, we could shed a greater light into the peephole of studying the role of the DSB repair pathways in non-B DNA-induced genetic instability. This will help further our understanding of DNA structure-induced genetic instability, evolution and human diseases; with the long-term goal of improving or developing targeted treatments for diseases that are linked to DNA structure-induced genetic instabilities. In fact, G.M. Zaunbrecher et al. have initiated attempts for improving gene-targeting by trying to affect the ratio of HR to NHEJ for the

enhancement of extra chromosomal recombination in somatic cells (112). Reaching new frontiers may mean crossing these two areas of research, which could give birth to creative and beneficial genetic tools.

REFERENCES

1. Wang, G., L. A. Christensen, and K. M. Vasquez. 2006. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc Natl Acad Sci U S A* 103:2677-2682.
2. Wang, G., and K. M. Vasquez. 2004. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A* 101:13448-13453.
3. Della, M., P. L. Palmboos, H. M. Tseng, L. M. Tonkin, J. M. Daley, L. M. Topper, R. S. Pitcher, A. E. Tomkinson, T. E. Wilson, and A. J. Doherty. 2004. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* 306:683-685.
4. Kha, D. T., G. Wang, N. Natrajan, L. Harrison, and K. M. Vasquez. Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. *J Mol Biol* 398:471-480.
5. Watson, J. D., and F. H. Crick. 1974. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in *Nature*, number 4356 April 25, 1953. *Nature* 248:765.
6. Zhao, J., A. Bacolla, G. Wang, and K. M. Vasquez. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci* 67:43-62.
7. Wang, G., J. Zhao, and K. M. Vasquez. 2009. Methods to determine DNA structural alterations and genetic instability. *Methods* 48:54-62.
8. Bacolla, A., A. Jaworski, J. E. Larson, J. P. Jakupciak, N. Chuzhanova, S. S. Abeysinghe, C. D. O'Connell, D. N. Cooper, and R. D. Wells. 2004. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* 101:14162-14167.
9. Bacolla, A., and R. D. Wells. 2004. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 279:47411-47414.
10. Wang, G., and K. M. Vasquez. 2006. Non-B DNA structure-induced genetic instability. *Mutat Res* 598:103-119.
11. Bacolla, A., and R. D. Wells. 2009. Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog* 48:273-285.

12. Adachi, M., and Y. Tsujimoto. 1990. Potential Z-DNA elements surround the breakpoints of chromosome translocation within the 5' flanking region of *bcl-2* gene. *Oncogene* 5:1653-1657.
13. Raghavan, S. C., and M. R. Lieber. 2004. Chromosomal translocations and non-B DNA structures in the human genome. *Cell Cycle* 3:762-768.
14. Raghavan, S. C., P. Chastain, J. S. Lee, B. G. Hegde, S. Houston, R. Langen, C. L. Hsieh, I. S. Haworth, and M. R. Lieber. 2005. Evidence for a triplex DNA conformation at the *bcl-2* major breakpoint region of the t(14;18) translocation. *J Biol Chem* 280:22749-22760.
15. Raghavan, S. C., and M. R. Lieber. 2006. DNA structures at chromosomal translocation sites. *Bioessays* 28:480-494.
16. Wang, G., and K. M. Vasquez. 2007. Z-DNA, an active element in the genome. *Front Biosci* 12:4424-4438.
17. Singleton, C. K., J. Klysik, S. M. Stirdivant, and R. D. Wells. 1982. Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* 299:312-316.
18. Peck, L. J., A. Nordheim, A. Rich, and J. C. Wang. 1982. Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. *Proc Natl Acad Sci U S A* 79:4560-4564.
19. Wang, A. H., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich. 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* 282:680-686.
20. Ha, S. C., K. Lowenhaupt, A. Rich, Y. G. Kim, and K. K. Kim. 2005. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* 437:1183-1186.
21. Kim, J. I., J. Heuser, and M. M. Cox. 1989. Enhanced *recA* protein binding to Z DNA represents a kinetic perturbation of a general duplex DNA binding pathway. *J Biol Chem* 264:21848-21856.
22. Herbert, A., and A. Rich. 1999. Left-handed Z-DNA: structure and function. *Genetica* 106:37-47.
23. Wahls, W. P., L. J. Wallace, and P. D. Moore. 1990. The Z-DNA motif d(TG)₃₀ promotes reception of information during gene conversion events while

- stimulating homologous recombination in human cells in culture. *Mol Cell Biol* 10:785-793.
24. Blaho, J. A., and R. D. Wells. 1989. Left-handed Z-DNA and genetic recombination. *Prog Nucleic Acid Res Mol Biol* 37:107-126.
 25. Weinreb, A., D. A. Collier, B. K. Birshtein, and R. D. Wells. 1990. Left-handed Z-DNA and intramolecular triplex formation at the site of an unequal sister chromatid exchange. *J Biol Chem* 265:1352-1359.
 26. Wahls, W. P., and P. D. Moore. 1990. Homologous recombination enhancement conferred by the Z-DNA motif d(TG)₃₀ is abrogated by simian virus 40 T antigen binding to adjacent DNA sequences. *Mol Cell Biol* 10:794-800.
 27. Juranic, Z., M. Kidric, R. Tomin, I. Juranic, I. Spuzic, and J. Petrovic. 1991. The importance of the specific Z-DNA structure and polyamines in carcinogenesis: fact or fiction. *Med Hypotheses* 35:353-357.
 28. Hamada, H., and T. Kakunaga. 1982. Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* 298:396-398.
 29. Schroth, G. P., P. J. Chou, and P. S. Ho. 1992. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J Biol Chem* 267:11846-11855.
 30. Vasudevaraju, P., Bharathi, R. M. Garruto, K. Sambamurti, and K. S. Rao. 2008. Role of DNA dynamics in Alzheimer's disease. *Brain Res Rev* 58:136-148.
 31. Wang, G., S. Carbajal, J. Vijg, J. DiGiovanni, and K. M. Vasquez. 2008. DNA structure-induced genomic instability in vivo. *J Natl Cancer Inst* 100:1815-1817.
 32. Freund, A. M., M. Bichara, and R. P. Fuchs. 1989. Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc Natl Acad Sci U S A* 86:7465-7469.
 33. Htun, H., and J. E. Dahlberg. 1988. Single strands, triple strands, and kinks in H-DNA. *Science* 241:1791-1796.
 34. Wells, R. D. 1988. Unusual DNA structures. *J Biol Chem* 263:1095-1098.
 35. Jain, A., G. Wang, and K. M. Vasquez. 2008. DNA triple helices: biological consequences and therapeutic potential. *Biochimie* 90:1117-1130.
 36. Schroth, G. P., and P. S. Ho. 1995. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res* 23:1977-1983.
 37. Kinniburgh, A. J. 1989. A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. *Nucleic Acids Res* 17:7771-7778.

38. Pestov, D. G., A. Dayn, E. Siyanova, D. L. George, and S. M. Mirkin. 1991. H-DNA and Z-DNA in the mouse c-Ki-ras promoter. *Nucleic Acids Res* 19:6527-6532.
39. Mirkin, S. M., V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, and M. D. Frank-Kamenetskii. 1987. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* 330:495-497.
40. Rich, T., R. L. Allen, and A. H. Wyllie. 2000. Defying death after DNA damage. *Nature* 407:777-783.
41. Hoeijmakers, J. H. 2001. Genome maintenance mechanisms for preventing cancer. *Nature* 411:366-374.
42. Keeney, S., and M. J. Neale. 2006. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem Soc Trans* 34:523-525.
43. Soulas-Sprauel, P., P. Rivera-Munoz, L. Malivert, G. Le Guyader, V. Abramowski, P. Revy, and J. P. de Villartay. 2007. V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. *Oncogene* 26:7780-7791.
44. O'Driscoll, M., and P. A. Jeggo. 2006. The role of double-strand break repair - insights from human genetics. *Nat Rev Genet* 7:45-54.
45. Moore, J. K., and J. E. Haber. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16:2164-2173.
46. Wilson, T. E., U. Grawunder, and M. R. Lieber. 1997. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* 388:495-498.
47. Wyman, C., D. Ristic, and R. Kanaar. 2004. Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst)* 3:827-833.
48. Pardo, B., B. Gomez-Gonzalez, and A. Aguilera. 2009. DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci* 66:1039-1056.
49. Lengauer, C., K. W. Kinzler, and B. Vogelstein. 1998. Genetic instabilities in human cancers. *Nature* 396:643-649.
50. Critchlow, S. E., and S. P. Jackson. 1998. DNA end-joining: from yeast to man. *Trends Biochem Sci* 23:394-398.

51. Lieber, M. R. 1999. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. *Genes Cells* 4:77-85.
52. Pastink, A., J. C. Eeken, and P. H. Lohman. 2001. Genomic integrity and the repair of double-strand DNA breaks. *Mutat Res* 480-481:37-50.
53. Takata, M., M. S. Sasaki, E. Sonoda, C. Morrison, M. Hashimoto, H. Utsumi, Y. Yamaguchi-Iwai, A. Shinohara, and S. Takeda. 1998. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J* 17:5497-5508.
54. Rothkamm, K., I. Kruger, L. H. Thompson, and M. Lobrich. 2003. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol* 23:5706-5715.
55. Malyarchuk, S., D. Wright, R. Castore, E. Klepper, B. Weiss, A. J. Doherty, and L. Harrison. 2007. Expression of *Mycobacterium tuberculosis* Ku and Ligase D in *Escherichia coli* results in RecA and RecB-independent DNA end-joining at regions of microhomology. *DNA Repair (Amst)* 6:1413-1424.
56. Cary, R. B., S. R. Peterson, J. Wang, D. G. Bear, E. M. Bradbury, and D. J. Chen. 1997. DNA looping by Ku and the DNA-dependent protein kinase. *Proc Natl Acad Sci U S A* 94:4267-4272.
57. Bliss, T. M., and D. P. Lane. 1997. Ku selectively transfers between DNA molecules with homologous ends. *J Biol Chem* 272:5765-5773.
58. Walker, J. R., R. A. Corpina, and J. Goldberg. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412:607-614.
59. Yaneva, M., T. Kowalewski, and M. R. Lieber. 1997. Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *EMBO J* 16:5098-5112.
60. Ma, Y., U. Pannicke, K. Schwarz, and M. R. Lieber. 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108:781-794.

61. Garcia-Diaz, M., K. Bebenek, J. M. Krahn, L. Blanco, T. A. Kunkel, and L. C. Pedersen. 2004. A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol Cell* 13:561-572.
62. Grawunder, U., M. Wilm, X. Wu, P. Kulesza, T. E. Wilson, M. Mann, and M. R. Lieber. 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388:492-495.
63. Li, Z., T. Otevrel, Y. Gao, H. L. Cheng, B. Seed, T. D. Stamato, G. E. Taccioli, and F. W. Alt. 1995. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* 83:1079-1089.
64. Modesti, M., J. E. Hesse, and M. Gellert. 1999. DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *EMBO J* 18:2008-2018.
65. Wu, P. Y., P. Frit, L. Malivert, P. Revy, D. Biard, B. Salles, and P. Calsou. 2007. Interplay between Cernunnos-XLF and nonhomologous end-joining proteins at DNA ends in the cell. *J Biol Chem* 282:31937-31943.
66. Weller, G. R., B. Kysela, R. Roy, L. M. Tonkin, E. Scanlan, M. Della, S. K. Devine, J. P. Day, A. Wilkinson, F. d'Adda di Fagagna, K. M. Devine, R. P. Bowater, P. A. Jeggo, S. P. Jackson, and A. J. Doherty. 2002. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297:1686-1689.
67. Pitcher, R. S., N. C. Brissett, and A. J. Doherty. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* 61:259-282.
68. Gong, C., P. Bongiorno, A. Martins, N. C. Stephanou, H. Zhu, S. Shuman, and M. S. Glickman. 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol* 12:304-312.
69. Zhu, H., and S. Shuman. 2005. A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). *J Biol Chem* 280:418-427.
70. Zhu, H., and S. Shuman. 2005. Novel 3'-ribonuclease and 3'-phosphatase activities of the bacterial non-homologous end-joining protein, DNA ligase D. *J Biol Chem* 280:25973-25981.
71. Zhu, H., and S. Shuman. 2007. Characterization of *Agrobacterium tumefaciens* DNA ligases C and D. *Nucleic Acids Res* 35:3631-3645.

72. Pitcher, R. S., N. C. Brissett, A. J. Picher, P. Andrade, R. Juarez, D. Thompson, G. C. Fox, L. Blanco, and A. J. Doherty. 2007. Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J Mol Biol* 366:391-405.
73. Hartlerode, A. J., and R. Scully. 2009. Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem J* 423:157-168.
74. Bernstein, K. A., and R. Rothstein. 2009. At loose ends: resecting a double-strand break. *Cell* 137:807-810.
75. Sung, P., L. Krejci, S. Van Komen, and M. G. Sehorn. 2003. Rad51 recombinase and recombination mediators. *J Biol Chem* 278:42729-42732.
76. Van Dyck, E., A. Z. Stasiak, A. Stasiak, and S. C. West. 1999. Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* 398:728-731.
77. McIlwraith, M. J., E. Van Dyck, J. Y. Masson, A. Z. Stasiak, A. Stasiak, and S. C. West. 2000. Reconstitution of the strand invasion step of double-strand break repair using human Rad51 Rad52 and RPA proteins. *J Mol Biol* 304:151-164.
78. Song, B., and P. Sung. 2000. Functional interactions among yeast Rad51 recombinase, Rad52 mediator, and replication protein A in DNA strand exchange. *J Biol Chem* 275:15895-15904.
79. van Gent, D. C., J. H. Hoeijmakers, and R. Kanaar. 2001. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2:196-206.
80. Johnson, R. D., and M. Jasin. 2000. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* 19:3398-3407.
81. Nagaraju, G., S. Odate, A. Xie, and R. Scully. 2006. Differential regulation of short- and long-tract gene conversion between sister chromatids by Rad51C. *Mol Cell Biol* 26:8075-8086.
82. Richardson, C., M. E. Moynahan, and M. Jasin. 1998. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev* 12:3831-3842.
83. Ip, S. C., U. Rass, M. G. Blanco, H. R. Flynn, J. M. Skehel, and S. C. West. 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature* 456:357-361.

84. Anderson, D. G., and S. C. Kowalczykowski. 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* 90:77-86.
85. Churchill, J. J., and S. C. Kowalczykowski. 2000. Identification of the RecA protein-loading domain of RecBCD enzyme. *J Mol Biol* 297:537-542.
86. Churchill, J. J., D. G. Anderson, and S. C. Kowalczykowski. 1999. The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in constitutive recombination activation. *Genes Dev* 13:901-911.
87. Amundsen, S. K., and G. R. Smith. 2003. Interchangeable parts of the *Escherichia coli* recombination machinery. *Cell* 112:741-744.
88. Sakai, A., and M. M. Cox. 2009. RecFOR and RecOR as distinct RecA loading pathways. *J Biol Chem* 284:3264-3272.
89. Ivancic-Bace, I., P. Peharec, S. Moslavac, N. Skrobot, E. Salaj-Smic, and K. Brcic-Kostic. 2003. RecFOR function is required for DNA repair and recombination in a RecA loading-deficient *recB* mutant of *Escherichia coli*. *Genetics* 163:485-494.
90. Morimatsu, K., and S. C. Kowalczykowski. 2003. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* 11:1337-1347.
91. Roman, L. J., and S. C. Kowalczykowski. 1989. Formation of heteroduplex DNA promoted by the combined activities of *Escherichia coli* *recA* and *recBCD* proteins. *J Biol Chem* 264:18340-18348.
92. Mao, Z., M. Bozzella, A. Seluanov, and V. Gorbunova. 2008. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* 7:1765-1771.
93. Kim, J. S., T. B. Krasieva, H. Kurumizaka, D. J. Chen, A. M. Taylor, and K. Yokomori. 2005. Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells. *J Cell Biol* 170:341-347.
94. Mao, Z., M. Bozzella, A. Seluanov, and V. Gorbunova. 2008. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle* 7:2902-2906.
95. Mansour, W. Y., S. Schumacher, R. Roskopf, T. Rhein, F. Schmidt-Petersen, F. Gatzemeier, F. Haag, K. Borgmann, H. Willers, and J. Dahm-Daphi. 2008.

- Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res* 36:4088-4098.
96. Shin, D. S., C. Chahwan, J. L. Huffman, and J. A. Tainer. 2004. Structure and function of the double-strand break repair machinery. *DNA Repair (Amst)* 3:863-873.
 97. Meek, K., V. Dang, and S. P. Lees-Miller. 2008. DNA-PK: the means to justify the ends? *Adv Immunol* 99:33-58.
 98. Wyman, C., and R. Kanaar. 2006. DNA double-strand break repair: all's well that ends well. *Annu Rev Genet* 40:363-383.
 99. Allen, C., A. Kurimasa, M. A. Brenneman, D. J. Chen, and J. A. Nickoloff. 2002. DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. *Proc Natl Acad Sci U S A* 99:3758-3763.
 100. Delacote, F., M. Han, T. D. Stamato, M. Jasin, and B. S. Lopez. 2002. An *xrcc4* defect or Wortmannin stimulates homologous recombination specifically induced by double-strand breaks in mammalian cells. *Nucleic Acids Res* 30:3454-3463.
 101. Li, X., and W. D. Heyer. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* 18:99-113.
 102. Arnold, D. A., and S. C. Kowalczykowski. 2000. Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. *J Biol Chem* 275:12261-12265.
 103. Zhang, S., and R. J. Meyer. 1995. Localized denaturation of *oriT* DNA within relaxosomes of the broad-host-range plasmid R1162. *Mol Microbiol* 17:727-735.
 104. Fattah, F., E. H. Lee, N. Weisensel, Y. Wang, N. Lichter, and E. A. Hendrickson. Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. *PLoS Genet* 6:e1000855.
 105. Pitcher, R. S., L. M. Tonkin, A. J. Green, and A. J. Doherty. 2005. Domain structure of a NHEJ DNA repair ligase from *Mycobacterium tuberculosis*. *J Mol Biol* 351:531-544.
 106. Aniuoku, J., M. S. Glickman, and S. Shuman. 2008. The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev* 22:512-527.

107. Chayot, R., B. Montagne, D. Mazel, and M. Ricchetti. An end-joining repair mechanism in *Escherichia coli*. *Proc Natl Acad Sci U S A* 107:2141-2146.
108. Rapp, A., and K. O. Greulich. 2004. After double-strand break induction by UV-A, homologous recombination and nonhomologous end joining cooperate at the same DSB if both systems are available. *J Cell Sci* 117:4935-4945.
109. Van Nguyen, T., N. Puebla-Osorio, H. Pang, M. E. Dujka, and C. Zhu. 2007. DNA damage-induced cellular senescence is sufficient to suppress tumorigenesis: a mouse model. *J Exp Med* 204:1453-1461.
110. Tavana, O., N. Puebla-Osorio, M. Sang, and C. Zhu. Absence of p53-dependent apoptosis combined with nonhomologous end-joining deficiency leads to a severe diabetic phenotype in mice. *Diabetes* 59:135-142.
111. Iliakis, G. 2009. Backup pathways of NHEJ in cells of higher eukaryotes: cell cycle dependence. *Radiother Oncol* 92:310-315.
112. Zaunbrecher, G. M., P. W. Dunne, B. Mir, M. Breen, and J. A. Piedrahita. 2008. Enhancement of extra chromosomal recombination in somatic cells by affecting the ratio of homologous recombination (HR) to non-homologous end joining (NHEJ). *Anim Biotechnol* 19:6-21.

VITA

Diem Thi Kha was born in Vietnam on October 26, 1983, and the daughter of Phuong Duong. After completing her work at Alief Hastings High School, Houston, Texas in 2001, she entered The University of Texas at Austin, Texas. She received the degree of Bachelor of Science with a major in biochemistry from UT in May 2005. For the next two years, she worked as a research technician in the Department of Medical Genetics Laboratories at Baylor College of Medicine. In August of 2007 she entered The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences.

Permanent address:
13215 Benford Dr.
Houston, Texas 77083